BDM Design Elements Pre-Analysis Plan

Horlick, Palma, Drichoutis

February 2026

1 General Information

1. Trial Title: BDM mechanism design choice impact on valuation elicitation
2. Location: United States of America
3. Primary Investigator: Dr. Marco Palma
4. Other Primary Investigators: Dr. Andreas Drichoutis & Benjamin Horlick
5. Keyword(s): Behavior
6. Additional Keyword(s): Underbidding, Overbidding, Game form recognition, Misbidding, BDM mechanism, induced value
7. JEL Code(s): D81, C90
8. Abstract: The incentive compatibility of the Becker-DeGroot-Marschak (BDM) mechanism implies valuations should not depend on the market offer range or where the induced value falls within that range. In our first experiment (AEARCTR-0016635), varying the support set's upper bound while holding the induced value fixed produces monotonic increases in mean willingness to accept. Risk preferences are not correlated with the size of the shift as hypothesized. This second experiment varies the induced value within a fixed range to test whether responses shift with value location. A similar shift in this environment implies a mechanical adjustment to the location of the induced value with respect to the range. Taken together, the experiments evaluate whether valuations are sensitive to response scale width and value placement, highlighting context dependence in a theoretically incentive-compatible mechanism.

2 Dates

1. Trial Start Date: February 23, 2026
2. Intervention Start Date: February 23, 2026
3. Intervention End Date: February 25, 2026
4. Trial End Date: February 25, 2026

3 Sponsors & Partners

1. Sponsor(s): N/A
2. Partner(s): N/A

4 Experimental Details

1. Intervention (Public): We implement the BDM mechanism in three repeated rounds to elicit willingness-to-accept for an induced value item. The value of the item is exogenously changed each round to $3, $6, and $9 in random order. The distribution of possible offers is fixed across rounds and ranges from $0 to $12.

2. Intervention (Hidden): N/A

3. Primary Outcome (End Point): Individual willingness-to-accept values in each treatment

4. Primary Outcomes (Explanation): We use the elicited values to test whether moving the relative placement of the induced value within the support set results in changes to the distribution of willingness-to-accept values. Please see the Analysis Plan in Section 6 for more details.

5. Secondary Outcomes (End Points): N/A

6. Secondary Outcomes (Explanation): Our survey maintains the same risk preference elicitation tasks from the initial experiment to minimize changes to the procedure; however, we omit the measures as secondary outcomes in this follow-up because our Analysis Plan does not include the corresponding heterogeneity analyses.

7. Experimental Design (Public): Our experiment is divided in two parts. In the first part, subjects are asked to submit offers to sell a card to the experimenter in three repeated rounds that vary within-subjects the value of the card at $3, $6 or $9 in random order. We fix the support set with lower and upper bounds of $0 and $12, respectively. At the end of the experiment, one of the three rounds is selected for realization. Following the BDM mechanism task, we ask respondents about their strategy selection process and give them the opportunity to explain a rationale in an open response format.

   The second part of the experiment elicits subjects' risk preferences. We implement the static version of the BRET (Crosetto and Filippin, 2013). Subjects complete the task once with no practice rounds. The task presents participants with 100 boxes arranged in a 10x10 grid. One randomly selected box holds a bomb while the other 99 contain a reward of $0.10. Each subject then chooses how many boxes to collect. If the bomb is among the selected quantity of boxes, the participant receives no additional earnings from the task; otherwise, we add the amount of money inside the collected boxes to the subject's total payoff.

   As a secondary measure of risk preferences, we also employ the self-reported risk question from the German Socio-Economic Panel survey (Dohmen et al., 2011).

8. Experimental Design (Hidden): N/A

9. Randomization Method: Computer; all randomizations are performed within Qualtrics.

10. Randomization Unit: Treatment order is randomized at the individual level.

11. Was the treatment clustered? No.

12. Sample Size

    (a) Planned Number of Clusters: 75 individuals

    (b) Planned Number of Observations: 225 observations (75 individuals and 3 rounds)

    (c) Sample size (or number of clusters) by treatment arms: N/A

    (d) Power calculation: We utilize the data from similarly parameterized treatments in Drichoutis et al. (2025) to estimate the inputs to our power calculations. With an induced value of $3.00, the average offer shifted from $2.701 with a support set upper bound of $4.00 to $3.158 with a support set upper bound of $6.00. Each subject participated in both treatments. The pooled standard deviation of the two offer distributions was 0.895. In our power analysis presented below, we conservatively assume a between-treatment correlation of 0.3, compared to the actual observed correlation of 0.501. The effect size, d_z, found by Drichoutis et al. (2025) is then 0.429, given by the formula:

$$d_z = \frac{\mu_1 - \mu_2}{\sigma_{diff}} \tag{1}$$

where μ_1 and μ_2 represent the mean offer value by treatment, and σ_{diff} is the correlation-adjusted pooled standard deviation:

$$\sigma_{diff} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2} \tag{2}$$

where σ_1 and σ_2 are the standard deviations by treatment and r is the coefficient of correlation (Cohen, 2013; Lakens, 2013).

We employ the asymptotic relative efficiency (ARE) method in our power calculation which estimates the sample size required under a parametric t-test at a given power level and converts the result to the sample size required by the nonparametric Wilcoxon signed-rank test that we use to test our primary hypothesis (Faul et al., 2009).

From Rosner et al. (2006), Cohen (2013), and Lakens (2013), the estimated sample size n to achieve the maximum significance level (probability of Type I error) under the Holm-Bonferroni step-down procedure with three joint hypotheses of $\alpha = \frac{0.05}{3} \approx 0.0167$ and a power level (probability of Type II error) of $\beta = 0.80$ under a one-sided t-test is given by:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{d_z^2} \tag{3}$$

where $z_{1-\alpha}$ and $z_{1-\beta}$ take the values of 2.13 and 0.84, respectively (Holm, 1979). The sample size required under a t-test is then 51. The ARE of the Wilcoxon test is $\frac{3}{\pi} \approx 0.955$, implying a sample size of 53 (Faul et al., 2009). Figure 1 presents the sensitivity of the sample size calculation to the assumed between-treatment correlation. Utilizing the actual observed correlation value of 0.509 yields a sample size of 39.

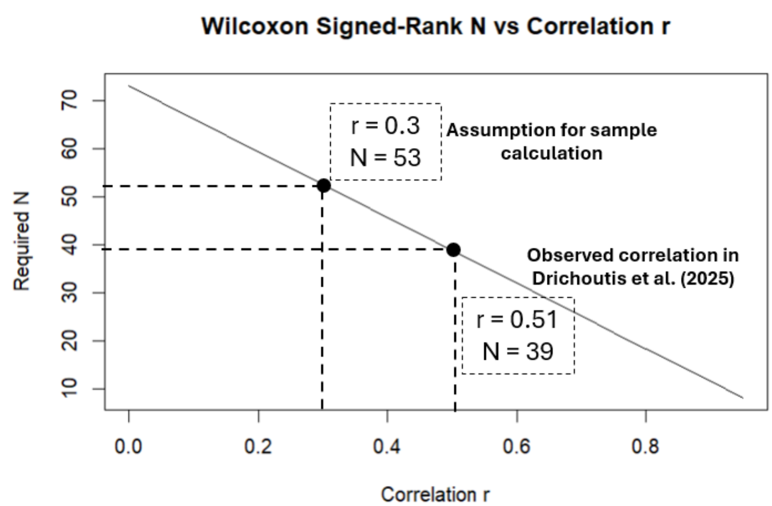

Figure 1: Since a higher correlation reduces σ_{diff}, our assumption of $r = 0.3$ means our sample size estimate is conservative.

As additional reference points, we conduct the same analysis using the results from our first experiment (AEARCTR-0016635). With an induced value of \$3, mean willingness to accept increased from \$3.413 in the \$6 support range treatment to \$5.451 with a pooled standard deviation of

2.069 and a between-treatment correlation of 0.42, yielding an effect size of 0.823 and sample size requirement of 15. Comparing the $6 treatment against the $4 treatment, we observe a mean shift of $0.737, a pooled standard deviation of 0.917, and correlation of 0.52, resulting in an effect size of 0.792 and sample size requirement of 16. Reducing the correlation to a hypothetical 0.25 increases the minimum sample size range to 17-22 subjects.

In the interest of conservatism, we adopt the larger threshold of 53 subjects determined for the prior study. Accordingly, we target a sample size of 75 participants, representing the minimum threshold grossed up for incomplete or otherwise unusable responses.

References for Experimental Details

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Crosetto, P., & Filippin, A. (2013). The 'bomb' risk elicitation task. *Journal of Risk and Uncertainty*, *47*(1).

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3).

Drichoutis, A. C., Palma, M. A., & Feldman, P. (2025). *Incentives and payment mechanisms in preference elicitation* [unpublished].

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149–1160.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. *Frontiers in psychology*, *4*, 863.

Rosner, B. A., et al. (2006). *Fundamentals of biostatistics* (Vol. 6). Thomson-Brooks/Cole Belmont, CA.

5 Institutional Review Board

1. IRB Name: Texas A&M Institutional Review Board
2. IRB Approval Date: February 16, 2026
3. IRB Approval Number: STUDY2025-0840: MOD00003539

6 Analysis Plan

6.1 Experiment Rationale

The BDM mechanism is one of the most widely used methods in experimental economics to elicit willingness to pay or willingness to accept for goods and services (Becker et al., 1964). Under the standard expected utility framework, changes in the support set should not impact a subject's willingness to accept (or willingness to pay) for a fixed induced value. In the BDM mechanism, submitting an offer equal to the induced value, i.e., providing a truthful response, is a weakly dominant strategy and always maximizes the expected payoff. In practice, however, systematic deviations from truth-telling are well documented (Bohm et al., 1997; Irwin et al., 1998; Noussair et al., 2004; Banerji and Gupta, 2014; Cason and Plott, 2014; Flynn et al., 2016; Mamadehussene and Sguera, 2023; Drichoutis et al., 2025). The theoretical incentive compatibility result holds independently of experimenter design decisions, namely the range used to draw the market offer and the item of interest's true value.

Our initial experiment (AEARCTR-0016635) tests the response-scale invariance property of the BDM mechanism by exogenously varying the upper bound of the market offer support set while holding the induced value fixed. Under incentive compatibility, these changes should not impact reported willingness to accept. However, we observe that the distribution of subject responses shifts systematically with the response scale: raising the upper bound increases average willingness-to-accept measures and the probability of overbidding relative to the induced value. These results indicate that elicited valuations depend on the mechanism's design elements beyond simply the incentive structure.

This experiment examines whether responses are sensitive to another design element, the item's value relative to the bounds of the support set. By varying the induced value while holding the support range fixed, we isolate shifts in the induced value's location within a static response scale, enabling us to test the value-location invariance prediction implied by the incentive compatibility of the BDM mechanism. The treatments mirror the relative induced value placements implemented in the first experiment without adjusting the support set. Finding similar patterns of deviation from the induced value would demonstrate the sensitivity of elicited valuations to experimenter design choices, highlighting the impact of context dependence in the mechanism.

6.2 Primary Analysis

- Mean Bid by Treatment:

  Let $WTA_{i,j}$, for $i \in \{3, 6, 9\}$ denote subject j's elicited willingness to accept for the treatment representing the induced value of \$3, \$6, and \$9, respectively. Define $d_{i,j} = WTA_{i,j} - i$ to represent the j-th subject's deviation from the induced value in the i-th treatment. Then, $\mathbb{E}[d_i]$ is the mean difference for each treatment.

  $H1$: $\mathbb{E}[d_3] > 0$

  $H2$: $\mathbb{E}[d_6] = 0$

  $H3$: $\mathbb{E}[d_9] < 0$

  $H1$ and $H3$ are evaluated using one-sided, single-sample Wilcoxon signed-rank tests (Wilcoxon, 1945) with $H2$ evaluated with a two-sided, single-sample Wilcoxon signed-rank test. We adopt the Holm-Bonferroni step-down procedure for multiple hypothesis testing (Holm, 1979).

  We first test whether exogenously changing the relative placement of the induced value within the support set is associated with deviations in the offer distribution away from the induced value. Our primary analysis examines shifts in the location of the willingness-to-accept distributions across treatments, while robustness checks evaluate corresponding hypotheses through regression analysis, difference in proportions, distributional skewness.

  The directional hypotheses reflect a predicted value-location effect: the position of the induced value within the support set may influence reported valuations. Behavioral evidence suggests that numeric judgments are sensitive to the framing of a stimulus' range and exhibit a propensity toward interior responses, often interpreted as anchoring or central tendency bias (Crosetto et al., 2020; Parducci, 1963; Poulton, 1979; Tversky and Kahneman, 1974). Interior attraction implies responses may drift away from salient boundaries. This predicts upward deviations when the induced value lies near the lower bound and downward deviations with an induced value close to the upper bound. Values near the midpoint would face weaker directional influence.

  The induced value is closer to the lower bound of the support range in the \$3 treatment, suggesting that average willingness-to-accept values will be above the induced value (overbidding). With the \$9 treatment, the induced value is nearer to the upper bound, implying the average offer level may fall below the induced value (underbidding). The \$6 treatment places the induced value at the midpoint of the support range, where the magnitude of the effect is expected to be minimal.

- Robustness Checks:

  1. Alternative Mean Bid Tests by Treatment:

     We also plan to test for statistically significant differences among the average treatment deviations. This enables us to make direct comparisons among the treatments in addition to the means relative to the IV:

$\mathbb{E}[d_3] > \mathbb{E}[d_6] > \mathbb{E}[d_9]$

These tests are evaluated jointly using a Friedman test with one-sided, paired Wilcoxon signed-rank tests used to make pairwise comparisons (Friedman, 1937; Wilcoxon, 1945).

2. Regression Analysis:

   As a robustness check of the primary analysis, we will use an OLS regression including additional demographic control variables. With $d_{i,j}$ representing the j-th subject's deviation from the induced value in treatment i, we will estimate the following specifications:

$$d_{i,j} = \alpha_0 + \alpha_1 \mathbf{1}[i = 3] + \alpha_2 \mathbf{1}[i = 9] + \varepsilon_{i,j} \tag{4}$$
$$d_{i,j} = \alpha_0 + \alpha_1 \mathbf{1}[i = 3] + \alpha_2 \mathbf{1}[i = 9] + \delta \mathbf{X}_j + \varepsilon_{i,j} \tag{5}$$
$$d_{i,j} = \alpha_1 \mathbf{1}[i = 3] + \alpha_2 \mathbf{1}[i = 9] + \eta_j + \varepsilon_{i,j} \tag{6}$$

   where $\mathbf{1}[i = k]$ is an indicator equal to 1 when observation (i, j) corresponds to treatment k. The \$6 treatment serves as the omitted baseline and $\mathbf{X}_j$ represents a vector of observable demographic controls for subject j including age, gender, ethnicity, risk preferences and income. The coefficients in Equation (4) then represent average deviations by treatment and, in Equation (5), the conditional mean deviation controlling for demographic factors. Finally, Equation (6) serves as a within-subject robustness specification, utilizing subject fixed effects, denoted by η_j, to capture time-invariant subject heterogeneity in lieu of the demographic control variables. We anticipate the predicted results of our primary analysis to hold:

   $\alpha_1 > 0$
   $\alpha_2 < 0$

   These hypotheses are evaluated t-tests with standard errors clustered at the subject level in all specifications.

3. Difference in Proportions:

   The use of only the willingness-to-accept distributions' means, either unconditional or conditional, in identifying treatment effects could conceal offsetting effects above and below the induced value. As an additional robustness check, we test the hypothesis that the probability of an offer above the induced value is decreasing across treatments. The approach is akin to the one adopted in Knetsch and Sinden (1984), which examines the change proportion of subjects willing to accept a given monetary offer in exchange for different endowed items. The difference in proportions test enables us to identify a shift in the number of subjects with offer levels above the induced value (overbidding) in the case where the overall willingness to accept distribution's mean remains unchanged if the portion of the sample bidding below the induced value moves closer to zero (underbidding).

   $P(d_{3,j} > 0) > P(d_{6,j} > 0)$
   $P(d_{6,j} > 0) > P(d_{9,j} > 0)$
   $P(d_{3,j} > 0) > P(d_{9,j} > 0)$

   These hypotheses are evaluated jointly with Cochran's Q test with pairwise comparisons using one-sided McNemar tests (Cochran, 1950; McNemar, 1947).

4. Skewness:

   Let $\hat{\gamma}_i, i \in \{3, 6, 9\}$ denote the sample skewness of the deviations from the induced value in each treatment.

   $\gamma_3 > 0$
   $\gamma_6 = 0$

$\gamma_9 < 0$

These hypotheses are evaluated using a large-sample normal approximation derived from the asymptotic variance of the sample skewness,

$$Z_i = \frac{\hat{\gamma}_i}{\sqrt{6/n_i}}, \tag{7}$$

where n_i denotes the number of observations per treatment. We conduct one-sided tests in the \$3 and \$9 treatments and utilize a two-sided test in the \$6 treatment to align with our predictions.

The skewness predictions follow from the same value-location hypothesis. In the \$3 treatment, the proximity of the induced value to the lower bound suggests a right skew with an upward response drift, generating more frequent positive deviations and a longer right tail. Conversely, we expect a negative skew in the \$9 treatment due to a response tail below the induced value. With the \$6 treatment at the midpoint of the support set, directional influence from the bounds offset, implying a symmetric distribution as positive and negative deviations are more balanced. Accordingly, we hypothesize $\gamma_3 > 0$, $\gamma_6 = 0$, and $\gamma_9 < 0$.

6.3 Heterogeneity Analyses

Our study includes several questions aimed at gaining insights into subjects' decision-making processes in the BDM task. We plan to investigate potential relationships between WTA and focuses on increasing the probability of "winning," maximizing payoff value, and desiring to control the task outcome. While not the primary focus of this paper, responses to these questions may inform future work. One of the questions is an open-ended response box requesting subjects to outline their strategy explicitly. Depending on the richness of the answers, we plan to use the techniques outlined in Hassan et al. (2025) to identify trends in text responses and establish behavioral patterns.

Our experimental survey requires subjects to correctly answer 11 comprehension questions covering both the structure of the BDM mechanism as well as payoff calculations in a variety of scenarios. We included the extensive suite of questions to improve subject understanding of the task; however, we cannot ensure with certainty that correct responses indicate understanding as opposed to trial and error. As a consequence, we track the number of incorrect attempts at answering the comprehension questions and plan to explore the robustness of the analyses to varying exclusion criteria based on the comprehension questions.

References for Analysis Plan

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

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

Bohm, P., Linden, J., & Sonnegard, J. (1997). Eliciting reservation prices: Becker-degroot-marschak mechanisms vs. markets. *The Economic Journal*, *107*(443).

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).

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*(3/4), 256–266.

Crosetto, P., Filippin, A., Katuščák, P., & Smith, J. (2020). Central tendency bias in belief elicitation. *Journal of Economic Psychology*, *78*, 102273.

Drichoutis, A. C., Palma, M. A., & Feldman, P. (2025). *Incentives and payment mechanisms in preference elicitation* [unpublished].

Flynn, N., Kah, C., & Kerschbamer, R. (2016). Vickery auction vs bdm: Difference in bidding behaviour and other-regarding motives. *Journal of the Economic Science Association*, *2*(2).

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, *32*(200), 675–701.

Hassan, T. A., Hollander, S., Kalyani, A., van Lent, L., Schwedeler, M., & Tahoun, A. (2025). Text as data in economic analysis. *Journal of Economic Perspectives*, *39*(3).

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Irwin, J. R., McClelland, G. H., McKee, M., Schulze, W. D., & Norden, N. E. (1998). Payoff dominance vs. cognitive transparency in decision making. *Economic Inquiry*, *36*(2).

Knetsch, J. L., & Sinden, J. A. (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *The Quarterly Journal of Economics*, *99*(3), 507–521.

Mamadehussene, S., & Sguera, F. (2023). On the reliability of the bdm mechanism. *Management Science*, *69*(2).

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157.

Noussair, C., Robin, S., & Ruffieux, B. (2004). Revealing consumers' willingness-to-pay: A comparison of the bdm mechanism and the vickery auction. *Journal of Economic Psychology*, *25*(6).

Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs: General and Applied*, *77*(2), 1.

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological bulletin*, *86*(4), 777.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124–1131.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, *1*(6), 80–83.