

A Double-Slit Experiment with Human Subjects¹

John Duffy*

Department of Economics, UC-Irvine

duffy@uci.edu

Ted Loch-Temzelides

Department of Economics, Rice University

tedt@rice.edu

Classification: Economic Sciences, Psychological and Cognitive Sciences

Keywords: Decision theory, Double-slit experiment, Repeated attribute measurements, Interference

Abstract

We study a sequence of “double-slit” experiments designed to perform repeated measurements of an attribute in a large pool of subjects using *Amazon Mechanical Turk*. Our findings contrast the prescriptions of decision theory in novel and interesting ways. The response to an identical sequel measurement of the same attribute can be at significant variance with the initial measurement. Furthermore, the response to the sequel measurement depends on whether the initial measurement has taken place. In the absence of the initial measurement, the sequel measurement reveals additional variability, leading to a multimodal frequency distribution reminiscent of an interference pattern, which is largely absent if the first measurement has taken place. This documents for the first time evidence of wave-like properties in human behavior. Such properties are not captured by existing theoretical models.

¹ The experiment reported on in this paper was funded by the UC Irvine School of Social Sciences. The experimental protocol was pre-approved by the UC Irvine Institutional Review Board, HS#: 2017-3650. The experimental design and hypotheses were pre-registered on the American Economic Association’s RCT registry, No. AEARCTR-0004122, pre-registration available at: <https://www.socialscienceregistry.org/trials/4122>. The data are available on openICPSR at <https://www.openicpsr.org/openicpsr/project/120284/version/V1/view/>. We thank Patrick Julius for excellent research assistance.

A cornerstone of axiomatic modeling of human decision-making holds that individuals have a well-defined ranking over possible deterministic or random alternatives.¹²³ As an example, this assumption implies that each individual has a full preference ranking among, say, the continuum of colors in the visible spectrum. Recent studies argue that it is plausible that the preference ranking (say, between two colors in the spectrum) emerges as part of the measurement, or elicitation process itself.⁴ Some experimental evidence provides support for this view.⁵ Related experiments typically assume a probabilistic framework where, for example, the subjects' Bayesian priors over gambles are refined through feedback information provided by the experimenter. Inferences about the subjects' posteriors are then contrasted to those implied by expected utility theory. A number of "paradoxes" have been identified through the resulting discrepancies. However, as probabilistic frameworks introduce an additional complexity layer, it is hard to distinguish whether the discrepancies from the prescriptions of decision theory are due to the absence of a preference ranking or due to the limited ability of subjects to form probabilistic assessments over an existing ranking. Even when experiments do not directly evoke lotteries over outcomes, they study whether a prior measurement of an attribute affects the sequel measurement of a different attribute (framing effects).⁶ Our investigation is motivated by the famous double-slit experiment, often used to demonstrate wave-like interference in physics. We perform repeated elicitations of the same attribute and study the final pattern resulting from subjects' responses conditional on whether previous measurements have occurred.

The double-slit experiment has a long tradition as both an idealized thought experiment and an expositional tool.⁷ While established in a physics context, the experiment can give rise to an elegant mathematical model via the introduction of non-commutative selective measurement operators.⁸ In other words, the mathematical foundation of quantum mechanics is a logical construction that is independent of its physical

interpretation. This modeling, in turn, can be interpreted and used in different contexts, including in the study of human subjects.⁹¹⁰ In the idealized physics framework, the double slit experiment involves a subatomic particle emitter firing particles in a general direction. The particles encounter a first barrier, which allows them to pass only through two slits. The final position and overall distribution of those particles which pass through the first barrier is recorded when they hit a second screen as illustrated in Figure 1. The most salient feature of the double-slit experiment in physics concerns the shape of the distribution describing the pattern created on the second screen. This depends critically on whether the experimenter elicits information about which slit a particle passes through the first barrier. For example, if detectors identify whether particles went through slit 1 or through slit 2 in Figure 1 then a bimodal distribution is observed on the second screen. On the other hand, if there is no detection of which slit the particles passed through, then a multimodal *interference pattern* emerges. As interference is a property of waves, the experiment reveals that subatomic particles demonstrate a form of “duality;” they may exhibit wave-like properties when not observed, and particle-like properties when they are observed.

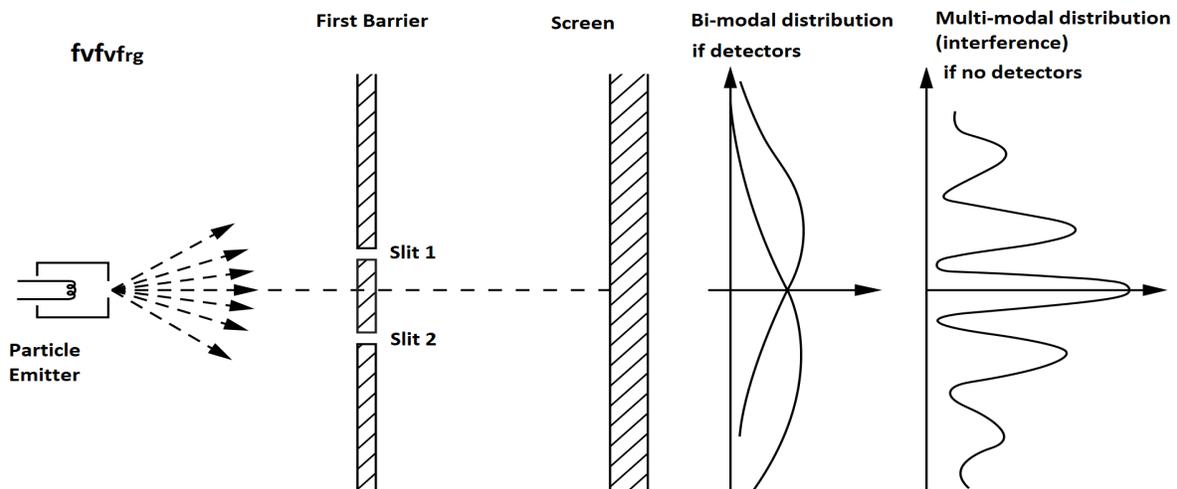


Figure 1: The Double Slit Experiment in Physics. Adapted from Feynman, Leighton, and Sands, 1965

Our ultimate goal is to determine if and under what conditions choices made by humans exhibit wave-like properties. Such a distinction would be surprising, but perhaps not more so than the claim, now universally accepted, that matter can exhibit a wave-particle duality. There are some interesting parallels already in place. First, as we mentioned already, it is plausible to think of the elicitation process as “creating” a preference ranking. Second, findings in psychology have highlighted that the order in which response alternatives are presented to respondents may lead to significantly different selections (non-commutativity).¹¹

We conduct a double-slit experiment with human subjects to investigate whether wave-particle duality is present in human choice data. While the double-slit experiment is often studied in the context of quantum mechanics, we do not take a stance on whether wave-like human behavior results from quantum mechanical processes in the brain. As interference may potentially play a role in a variety of cognitive processes, its origin could be entirely classical, for example, resulting from neural oscillators.¹²¹³

For our purposes, the double-slit structure provides a straightforward way of creating a bare-bones elicitation process that captures the interaction between subjects and the experimenter during repeated measurements of an attribute. A virtue of our approach is in that it does not rely on the interpretation of “ambiguous concepts,” or any probabilistic assessments from the subjects. We will report on successive measurements of the following type: from a scale 1-10 (in increments of 0.5) state your agreement with the statement “*I like the color green.*” The elicitation of preferences regarding a color seems neutral enough to protect our analysis from any significant “experimenter demand” or Hawthorne effects.¹⁴ Subjects were further instructed that 0.0 means “completely disagree,” 5.0 means “neither agree nor disagree,” and 10.0 means “completely agree.” Following the setup of the physics experiment, we distinguish between two main treatments, each containing two measurements. In both

treatments, we ask the above question twice of each subject. What distinguishes the two treatments is the *way* we elicit the attribute value initially. In the first stage of treatment I, the *monitored treatment*, each subject's individual choice was elicited and recorded by the computer program and reported back to the subject, so that it was very clear that their choice was being recorded. After seeing their first stage choice and clicking on a "next" button, subjects whose first stage response passed through either slit 1 (attribute value in $\{2, 2.5, 3\}$) or slit 2 (attribute value in $\{7, 7.5, 8\}$) proceeded to stage 2 where they were again asked their opinion of the same statement in stage 2 and where their choice was again recorded. By contrast, in treatment II, the *unmonitored treatment*, we do not elicit the value of subjects' opinion in the first stage. Instead, we ask subjects to think of their numerical choice and to "keep this number in your memory for a moment". Next we ask whether a subject's attribute choice belongs to the set $\{2, 2.5, 3, 7, 7.5, 8\}$ which comprises the union of the two slits: in $\{2, 2.5, 3\} \cup \{7, 7.5, 8\}$. Subjects simply answer Yes or No to this question. Subjects who answered Yes to this question immediately proceeded to stage 2 where they were again asked their opinion of the same statement in stage 2 but this time their choice was recorded just as in stage 2 of the monitored treatment. In both treatments, subjects who did not pass through the slit in stage 1 were immediately sent to complete a brief demographic survey and paid. Subjects who moved on to stage 2 completed that stage before moving on to the same demographic survey. Notice that there is no mention of slits in either treatment. Importantly, while we know something about subjects' first stage attribute choice in the unmonitored treatment, we do not know the precise value of their first stage choice or which slit it passed through. Using this design, subjects should understand that their first-stage choice was not known to the experimenters in the unmonitored treatment while it was known to the experimenters in the monitored treatment.

The details of the experiment are described in the *Methods* section of the article. We studied responses from a total of 1,620 subjects, recruited from

Amazon's Mechanical Turk. Following completion of the questionnaire, subjects were paid for their participation (USD\$0.50 per subject).

Subjects were randomly assigned to either the monitored (n=808) or to the unmonitored (n=812) treatments. A total of 702 subjects passed through the two slits to the second measurement stage, most of them from the {7,7.5,8} slit, indicating an overall favorable view of the color green. Subjects did not receive any additional information between the two stages. Decision theory puts no restrictions on the predicted agreement with the statement “*I like the color green,*” as some subjects might like green more than others. However, as no information or payoff-relevant characteristics are involved in this elicitation, standard theory predicts that the second stage answer should be identical to the first, regardless of whether the first stage answer was monitored (subjects’ first stage choice was recorded and reported back to them) or unmonitored (subjects’ first stage choice was not observed or recorded). Our results are at significant variance with this prediction.

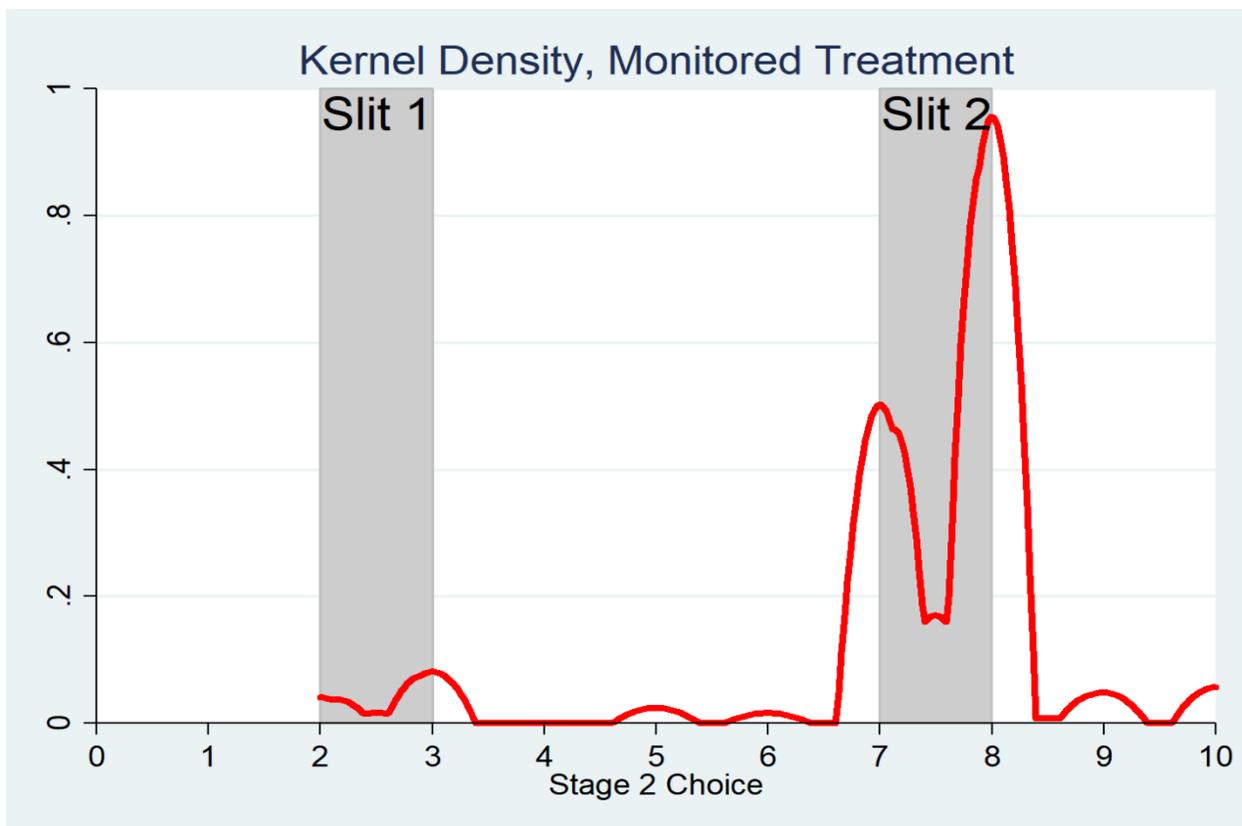


Figure 2: Estimated Kernel Density, Monitored Treatment

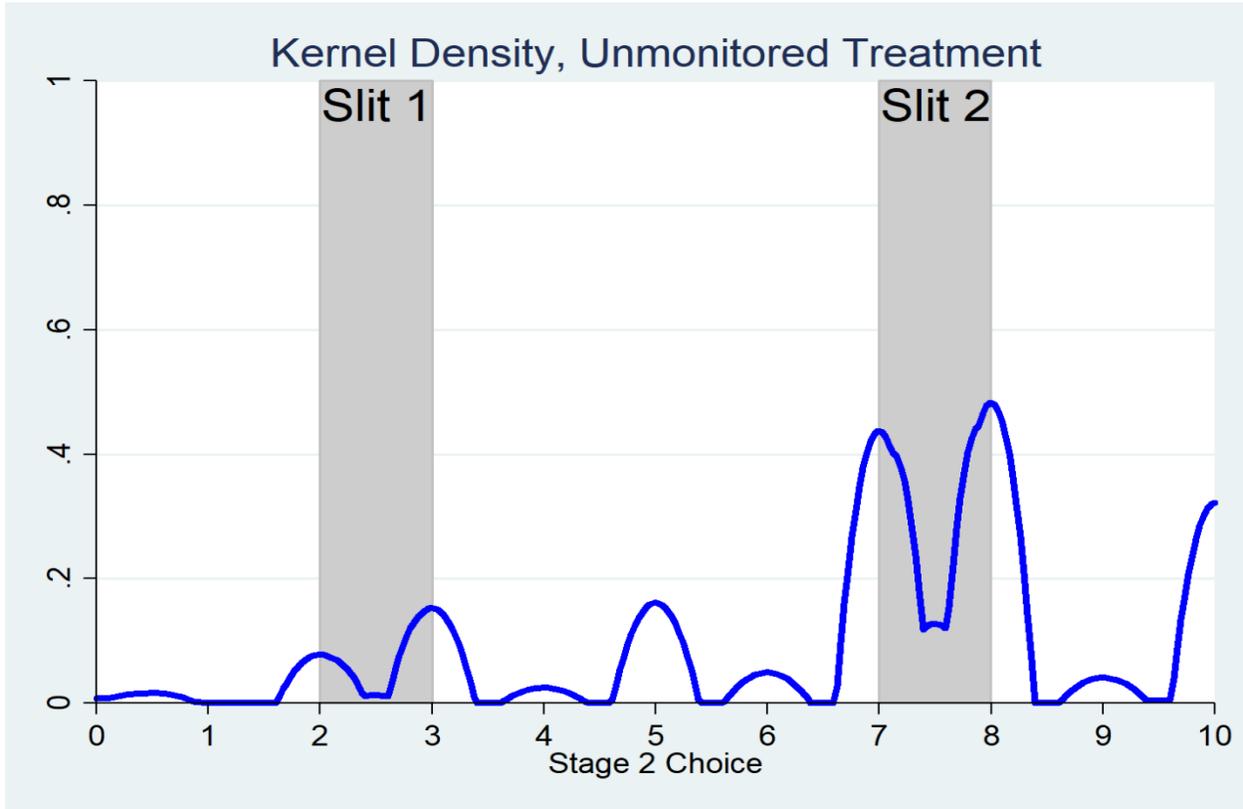


Figure 3: Estimated Kernel Density, Unmonitored Treatment

Figures 2 and 3 show kernel density plots of second stage choices (by those passing through the slits) in the Monitored and Unmonitored treatments respectively. These were estimated in Stata using the Epanechnikov kernel function, but other kernel functions give similar results

Notice that the estimated kernel density for the Unmonitored treatment is more diffused than for the Monitored treatment and the former covers the entire range of possible 2nd stage guesses, 0-10, whereas the estimated kernel density for the monitored treatment is more concentrated around the slit choices between 7-8 and does not cover the entire range of possible 2nd stage guesses, beginning only at around 1.5. The non-monotonicity of the density *within* each slit is present in both treatments and indicates a preference to report an integer value, as opposed to a decimal.

Confirming the impression given by Figures 2-3, we find that the distributions of second-stage responses made by those passing through the slits in the Monitored and Unmonitored treatments are significantly different from one-another according to a Kolmogorov-Smirnov test, $p < .0001$ (see also Figure A5 in the Methods section). Specifically, second-stage guesses are more dispersed over the admissible range in the Unmonitored treatment as compared with the Monitored treatment. Similarly, we also find a large difference in the dispersion of second stage responses between the two treatments, as measured by their standard deviations: the standard deviation in the Unmonitored treatment, 2.32, is more than 1.5 times greater than the standard deviation in the Monitored treatment, 1.49. This difference is significant according to an F-Test for equality of two standard deviations $p < .0001$. Finally, although an approximately equal number of subjects were randomly assigned to each treatment, 237 subjects passed to the second stage in the Monitored case, while 465 passed to the second stage in the Unmonitored one. This difference in proportions is statistically significant according to Pearson's Chi-Squared Test, $p < .0001$.

While the finding that differences in feedback from the experimenter may change subjects' responses is not new, the simplicity of the double-slit sequential measurement of a single attribute allows us to observe exactly *how* the pattern of responses changes from a first-stage measurement. It has not escaped our notice that the specific way in which the variance increased in the unmonitored treatment is the result of the emergence of a multi-modal frequency distribution that is reminiscent of an interference pattern. This is in contrast to the monitored case, which results in an essentially bi-modal distribution around the two slits. Documenting that human behavior can give rise to an interference pattern is a new and thought-provoking finding. To the extent that human behavior exhibits wave-like properties, these are not captured by existing models.¹⁵ Our results point to the need for additional research to further investigate this claim. While an affirmative conclusion would have profound implications

for modeling in decision theory, economics, and related fields, perhaps it would not be more surprising than the analogous conclusion we now know to be true in physics. Modeling the wave-particle duality of human behavior might account for some observed puzzles in human decision-making,¹⁶ but may also impose limits on our ability to measure economic attributes with an arbitrary degree of accuracy.

Methods

1. Pre-registration of study design and hypotheses and IRB approval

The design and hypotheses of this study were pre-registered on the American Economic Review's registry for randomized controlled trials (RCT) under the title "Measurement and Survey Response," *RCT ID AEARCTR-0004122*, on March 19, 2019, <https://www.socialscienceregistry.org/trials/4122>. The experimental protocol was pre-approved by the UC Irvine Institutional Review Board, HS#: 2017-3650.

2. Subjects, software, payments and data

Subjects were recruited from among *Amazon's Mechanical Turk* workforce residing in the U.S. In total, we recruited 1,846 subjects between April 3, 2019 and February 18, 2020. Subjects were only allowed to participate once. They first read an informed consent study page and if they agreed, proceeded on to the experiment. The experiment was computerized and programmed in *oTree*.¹⁷ Participation required only internet access and took no more than 10 minutes. Subjects who completed the task earned USD\$0.50. A total of 1,620 subjects completed the experiment, a completion rate of 87.8%. We report only on the 1,620 subjects who completed the study. The data collected for this study are available on *openICPSR*, at <https://www.openicpsr.org/openicpsr/project/120284/version/V1/view/>

3. Experimental Design

The experiment consists of two treatments: *Monitored* and *Unmonitored*. Subjects were randomly assigned to one (and only one) treatment. Each treatment consists of one or two stages followed by a questionnaire. In the first stage of both treatments, subjects were asked their opinion of the statement, “I like the color green” on a 0-10 point scale in increments of 0.5, with 0 (10) meaning complete disagreement (agreement) with the statement. Preliminary testing of 12 survey questions with undergraduate students at UC Irvine revealed that this particular question had a more uniform distribution of responses (on a 7-point *Likert scale*), as compared with other survey questions that we considered, though there was a skew in favor of the color green.

In our experiment, whether subjects passed from stage 1 to stage 2 or from stage 1 to the questionnaire depended on whether their first stage answer to the question lied within either of two slits.

Slit one was a numerical choice of either 2, 2.5, or 3. Slit 2 was a numerical choice of either 7, 7.5, or 8. If their stage 1 choice was contained within either slit, they passed on to stage 2.

Otherwise they transited from stage 1 directly to the questionnaire. Subjects were *not informed* of the procedure for determining whether they passed from stage 1 to stage 2 or directly to the questionnaire in either treatment.

A. Monitored Treatment

In the Monitored treatment subjects began stage 1 with the following question:

Question: Consider your opinion of this statement:

I like the color green

Think about a number that most closely matches your opinion of the statement above, on a scale of 0 to 10 inclusive in increments of 0.5 where 0.0 means “Completely Disagree” 5.0 means “Neither Disagree nor Agree and 10.0 means Completely Agree.

Your opinion

[Input box]

[Next button]

After entering their choice and clicking on a Next button, subjects saw a screen recording their choice:

You were asked to consider your opinion of this statement on a scale from 0 to 10 inclusive in increments of 0.5, where 0.0 means "Completely Disagree", 5.0 means "Neither Disagree nor Agree", and 10.0 means "Completely Agree".

I like the color green.

You responded: x.x

[Next button]

On this confirmation screen, x.x was the number between 0.0 and 10.0 inclusive corresponding to their first stage choice. This message made it clear to subjects that their first stage choice in the monitored treatment was being recorded by the computer program.

If their choice in the first stage ***did not*** lie in (pass through) one of the two slit intervals, {2, 2.5, 3} or {7, 7.5, 8}, then, after

clicking the next button they were immediately sent to complete the ex-post demographic survey.

If their choice in the first stage *did* pass through one of the two slits, {2, 2.5, 3} or {7, 7.5, 8} then, after clicking the next button they moved on to the stage 2 screen where they saw this question:

Question: You are again asked your opinion of the statement:

I like the color green.

Please enter the number that best corresponds to your opinion of this statement, on a scale from 0 to 10 inclusive in increments of 0.5, where 0.0 means "Completely Disagree", 5.0 means "Neither Disagree nor Agree", and 10.0 means "Completely Agree"

Your opinion.

[Input Box]

[Next button]

Once they entered a choice and clicked on the next button, their answer was recorded and they immediately moved on to the ex-post survey.

B. Unmonitored Treatment

In the Unmonitored Treatment, subjects begin stage 1 with the following question:

Question: Consider you opinion of this statement:

I like the color green

Think about a number that most closely matches your opinion of the statement above, on a scale of 0 to 10 inclusive in increments of 0.5 where 0.0 means “Completely Disagree” 5.0 means “Neither Disagree nor Agree and 10.0 means Completely Agree. Keep this number in your memory for a moment.

[Next button]

After clicking on the next button, they moved to a screen that posed this question:

Question: You were asked to consider your opinion of the statement

I like the color green

on a scale from 0 to 10 inclusive in increments of 0.5 where 0.0 means “Completely Disagree” 5.0 means “Neither Disagree nor Agree and 10.0 means “Completely Agree”

Is the number you chose in the set{2.0, 2.5, 3.0, 7.0, 7.5, 8.0}?

[Yes button] [No button]

If they clicked on No, they were immediately sent to complete the ex-post demographic survey. If they clicked on Yes, they moved on to the stage 2 screen where they saw this question:

Question: You are again asked your opinion of the statement:

I like the color green.

Please enter the number that best corresponds to your opinion of this statement, on a scale from 0 to 10 inclusive in increments of 0.5, where 0.0 means "Completely Disagree", 5.0 means

"Neither Disagree nor Agree", and 10.0 means "Completely Agree"

Your opinion.

[Input Box]

[Next button].

Once they entered a choice and clicked on the next button, their answer was recorded and they immediately moved on to the ex-post survey.

The ex-post survey asked subjects their country of citizenship, their postal code, their age, their gender and their race; the latter had a drop-down box.

Notice that the primary difference between the monitored and unmonitored treatment is that subjects' first stage choice in the monitored treatment *is elicited and recorded* while subjects' first stage choice in the unmonitored treatment is *not elicited*. While we did ask subjects in the unmonitored treatment whether their first stage choice was in the set {2.0, 2.5, 3.0, 7.0, 7.5, 8.0}, this was only to establish whether or not they would pass through the slits to stage 2. Their answer to this question *does not reveal* any additional information about their stage 1 choice. Subjects in the monitored treatment whose choice was in the same slit set {2.0, 2.5, 3.0, 7.0, 7.5, 8.0} automatically passed through to stage 2 and those whose choices were not in this set did not pass through, just as in the unmonitored treatment. Again, subjects were *not informed* of the procedure determining whether they passed from stage 1 to 2.

4. Data overview

We have data from 1,620 participants. The average age is 35.1 years. Males comprise 58.7% and females 41.3% of our subject sample. The breakdown by race is Asian, 8.99%; Black or African American, 9.77%; Hispanic or Latino, 6.18%; Mixed race, 3.21%; Pacific Islander, 0.25%; White, 72.59%. Figure A1 shows a heat map of the postal codes reported by participants in our study. It is clear from these data that our sample comprises a broad-based representation of people residing in the USA at the time of our study.

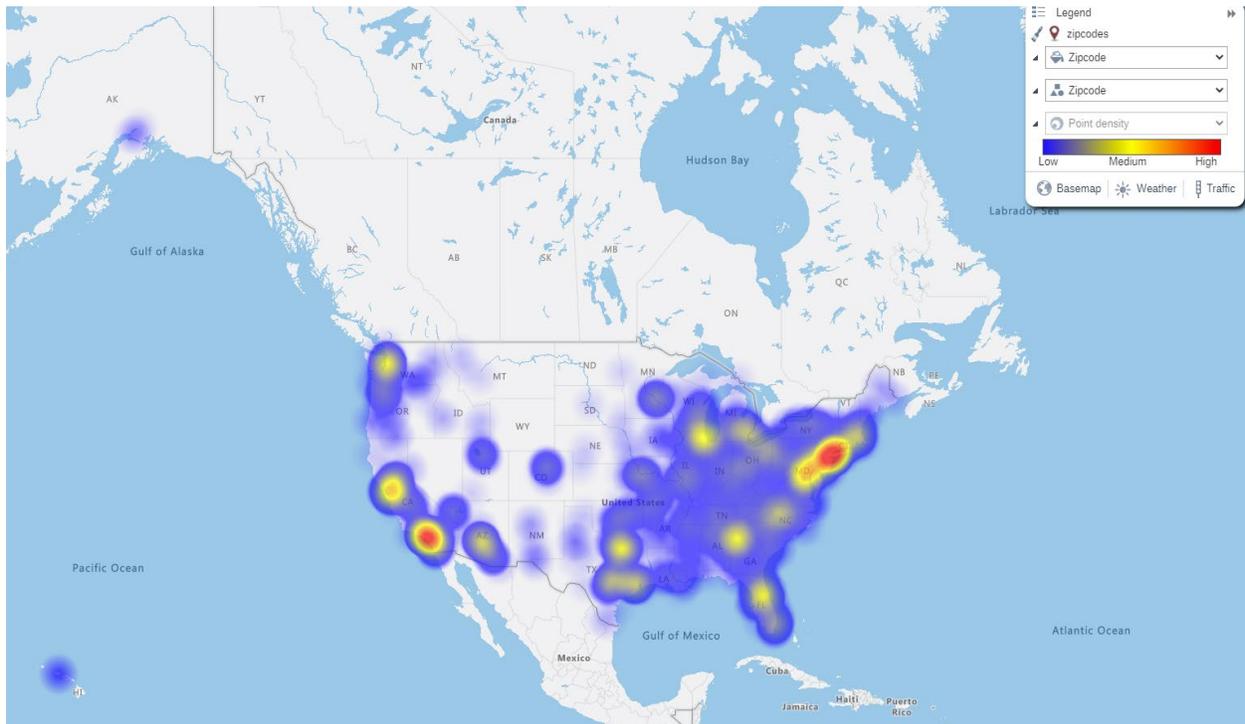


Figure A1: Heat map showing postal codes of study participants.

Of the 1,620 participants, 812 subjects were randomly assigned to the unmonitored treatment and 808 to the monitored treatment. Of the 812 subjects in the unmonitored treatment, 465 pass through one of the two slits, a pass-through rate of

57.3%. Of the 808 subjects in the monitored treatment, 237 pass through one of the two slits, a pass-through rate of 29.3%. Using Pearson's Chi-square test of proportions, the null hypothesis of no difference in pass through rates between the two treatments can be rejected $p < .0001$, in favor of the alternative that pass through rates were higher in the unmonitored treatment. We expect this finding is due to the fact that whether or not subjects pass through the two slits is self-reported in the unmonitored treatment but is explicitly verified in the monitored treatment. The kernel densities of second stage choices shown in Figures 2 and 3 of the text come from the 465 subjects of the unmonitored treatment and the 237 subjects of the monitored treatment, respectively, whose first stage choices were in one of the two slits.

5. Additional results and findings

In this section, we provide additional results and findings in support of what is reported in the main text of the paper.

A. Monitored treatment

For the monitored treatment we have all of the 808 subjects' first stage choices. Among the subjects in this treatment who passed through either slit, we can compare their first and second stage choices for consistency. Figure A2 shows a histogram of the first stage choices of the 808 subjects in the monitored treatment. Table A2 provides some characteristics of the first and second stage choices in the monitored treatment.

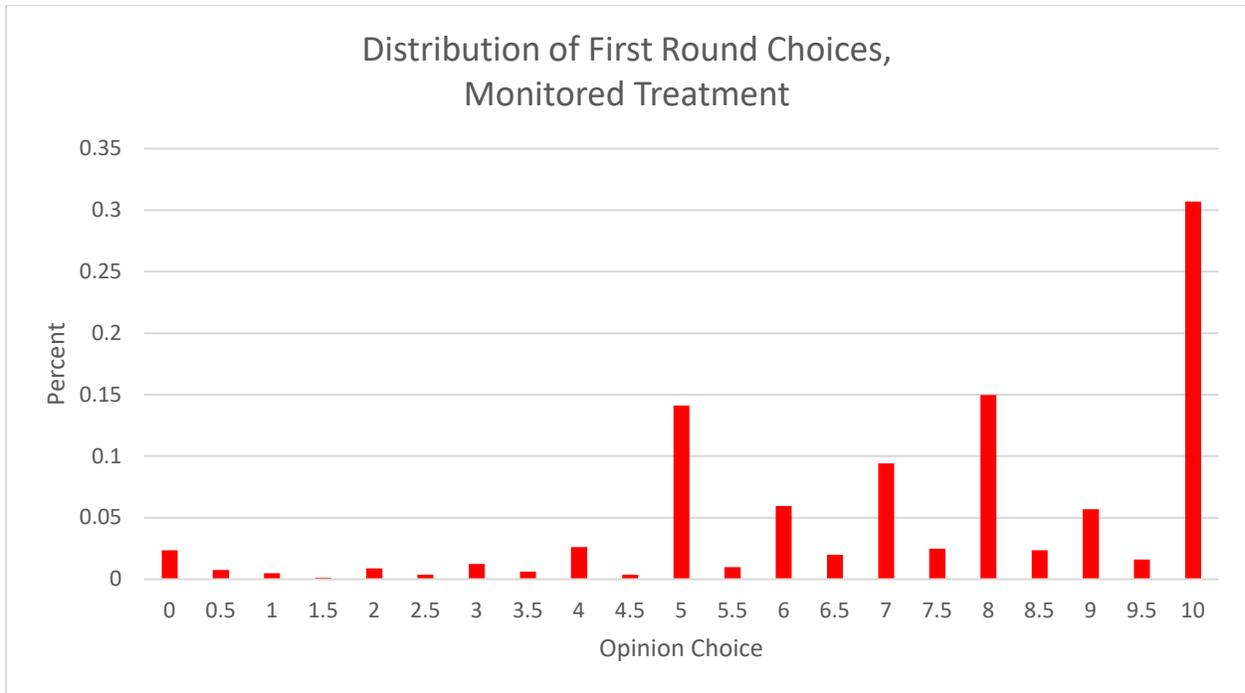


Figure A2: Histogram of first stage choices, monitored treatment

Table A1: Stage 1 and 2 Choices in the Monitored Treatment

Variable	Stage 1 Choice	Stage1 Choice of Those Passing Through a Slit	Stage 2 Choice
Mean	7.438	7.179	7.342
St. Dev.	2.504	1.475	1.491
Coeff. Var.	0.337	0.206	0.203
Median	8	8	8
Min	0	2	2
Max	10	8	10
Skewness	-0.953	-2.566	-2.128
Kurtosis	3.531	8.591	7.966
No. Obs.	808	237	237

The second column of Table A1 shows properties of all 808 first stage choices in the monitored treatment. Together, Figure A2

and Table A1 reveal that choices in the first stage of the monitored treatment are skewed to the right, with a mean of 7.4 and a median of 8, indicating a majority of participants agree with the statement “I like the color green”. While we do not observe such first stage choices for the unmonitored treatment, since subjects were randomly assigned to the two treatments, we conjecture that the distribution of first stage choices in the unmonitored treatment is similar to that of the monitored treatment.

We next compare the stage 1 choices of participants in the monitored treatment who passed through one of the two slits with their subsequent stage 2 choice. Properties of these 237 observations are shown in the last two columns of Table A2. Notice in particular that there is little difference in the mean, standard deviation and the coefficient of variation (the ratio of the standard deviation to the mean) as well as other measures between the stage 1 decisions of subjects who passed through the slit and their subsequent stage 2 choices. Indeed, the correlation coefficient between stage 1 and stage 2 choices of subjects passing through the two slits in stage 1 is 0.891, which is significantly different from 0, ($p < .0001$). Of the 237 subjects in the monitored treatment passing through one of the two slits, 202 of those subjects, or 85.2% chose the *exact same number* in stage 2 that they chose in stage 1. Of the remaining 35 subjects (14.8%) who passed through one of the two slits in stage 1 and chose different numbers in stage 2 than they chose in stage 1, 32 subjects adjusted their choice upwards and 3 adjusted their choice downwards in stage 2 relative to stage 1. Due to the

latter asymmetry, a Wilcoxon signed rank test for matched pairs indicates that we can reject the null hypothesis that the median deviation is 0 in favor of the alternative that stage 2 choices are higher than stage 1 choices ($p < .0001$), but this is because the test essentially excludes the 202 out of 237 or 85% of observations where the deviation was exactly 0. The overall mean deviation, i.e., stage 2 choice minus stage 1 choice (among those passing through a slit) is found to be just 0.16 with a standard error of .04. We note that of the 35 subjects who adjusted their stage 1 choice in stage 2, 16 (45.7%) chose numbers in stage 2 that were within the *same slit* that they chose in stage 1, e.g., adjusting from a stage 1 choice of 7 to a stage 2 choice of 8, so that only 19 or 54.3% of the 35 adjusting subjects chose stage 2 numbers that were outside of the slits they passed through in stage 1. Thus overall, 218 (202+16) subjects or 92.0% of all subjects in the monitored treatment who passed through one of the two slits in stage 1 chose numbers in stage 2 that were in the same slit as their stage 1 choice.

A further view of the latter finding is presented in Figure A3 which presents an XY bubble plot of stage 1 choices against stage 2 choices of all 237 subjects who passed through the 2 slits in the monitored treatment.

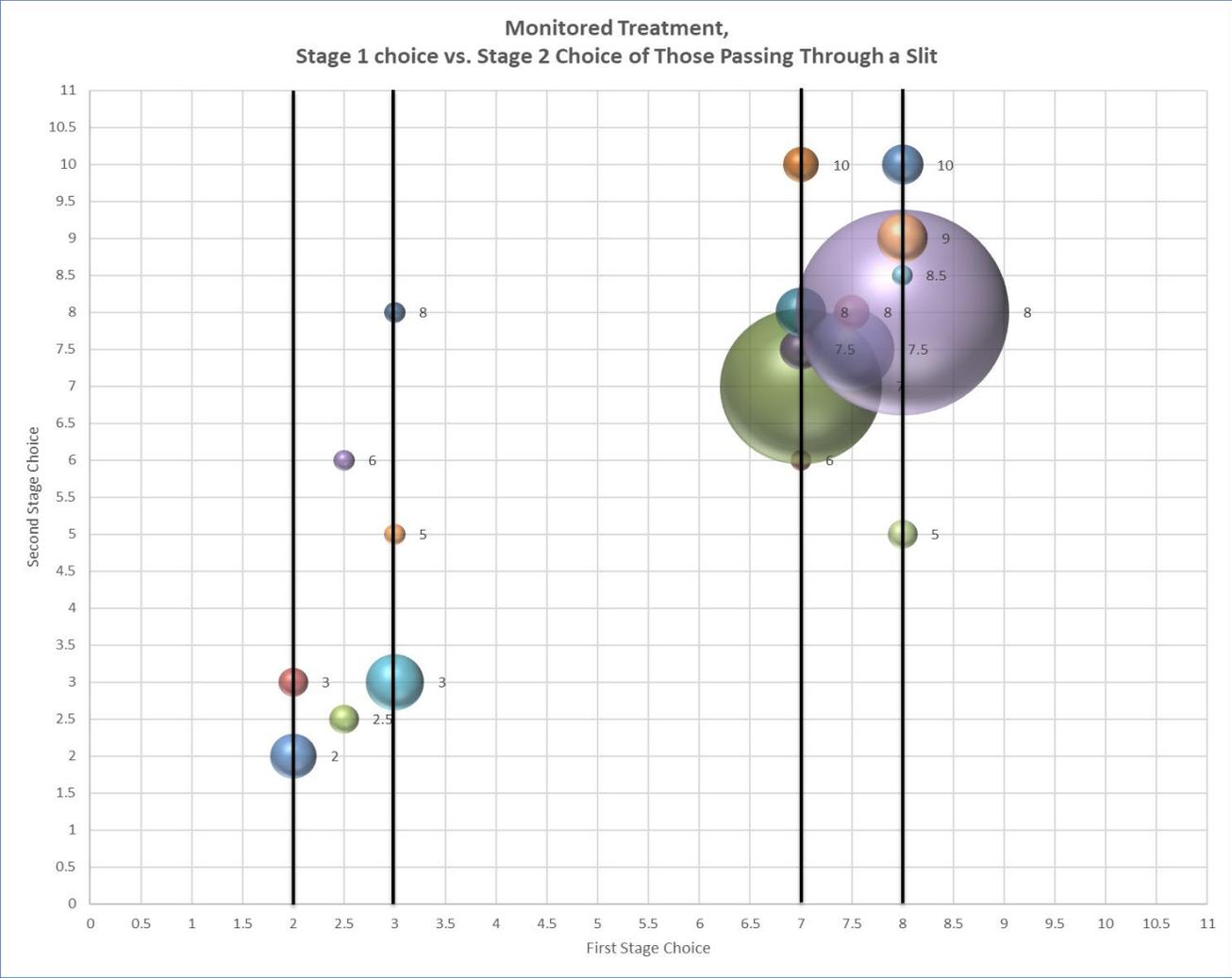


Figure A3: Stage 1 choices versus Stage 2 choices of those passing through the two slits, Monitored Treatment.

B. Unmonitored treatment

For the unmonitored treatment, we only have data on second stage choices, conditional on a self-report by the subjects that their first stage guess fell within one of the two slit intervals. As noted earlier, we have 465 such observations. Table A2 provides some characteristics of the stage 2 choices and for comparison purposes, repeats data from Table A2 on stage 2 choices in the monitored treatment.

Table A2: Stage 2 Choices in the Unmonitored Treatment

Variable	Unmonitored Treatment Stage 2 Choice	Monitored Treatment Stage 2 Choice (From Table A1)
Mean	6.967	7.342
St. Dev.	2.316	1.491
Coeff. Var.	0.332	0.203
Median	7.5	8
Min	0	2
Max	10	10
Skewness	-0.818	-2.128
Kurtosis	3.120	7.966
No. Obs.	465	237

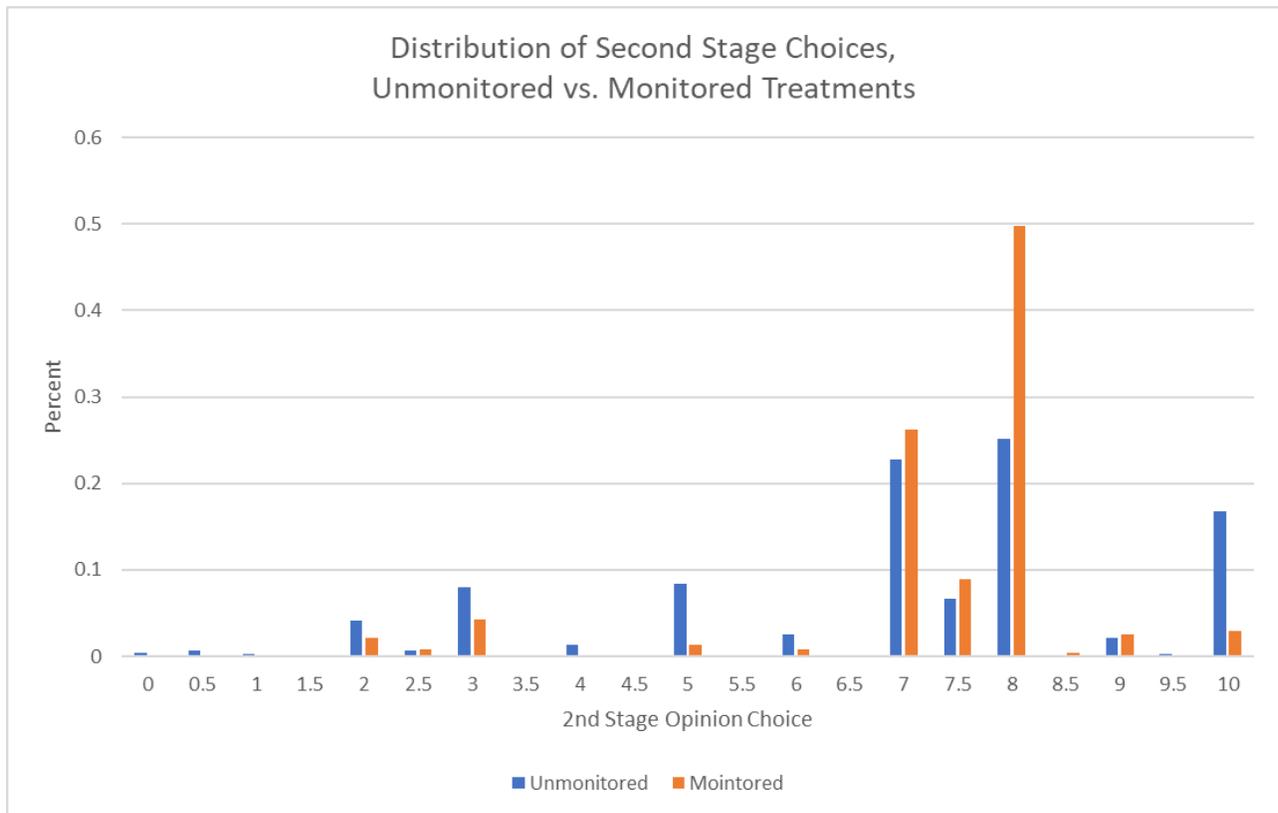


Figure A4: Histogram of second stage choices, unmonitored and monitored treatments

Figure A4 shows a histogram of stage 2 choices for both the unmonitored and monitored treatments. From Table A2 and Figure A4, we see clear evidence that the stage 2 choices in the unmonitored treatment are more dispersed over the admissible 0-10 choice range than are the stage 2 choices in the monitored treatment, which are more concentrated around the two slit intervals of 2-3 and 7-8.

Indeed, an F-test of the null hypothesis that the standard deviations of the second stage choices are equal between the two treatments is rejected ($f=0.4143$, $p<.0001$) in favor of the alternative that the standard deviation of stage 2 choices in the unmonitored treatment is greater than the standard deviation of stage 2 choices in the monitored treatment.

Finally, we consider whether the distribution of stage 2 choices in the unmonitored treatment differs from the distribution of stage 2 choices in the monitored treatment. The cumulative frequency distributions of stage 2 choices in both treatments are shown in Figure A5. As this figure reveals, choices in the unmonitored treatment are more widely dispersed than in the monitored treatment, where choices are concentrated around the two slits of 2-3 and 7-8. For this reason, we do not observe first order stochastic dominance of second stage choices in the monitored treatment, but instead, a single sharp crossing point.

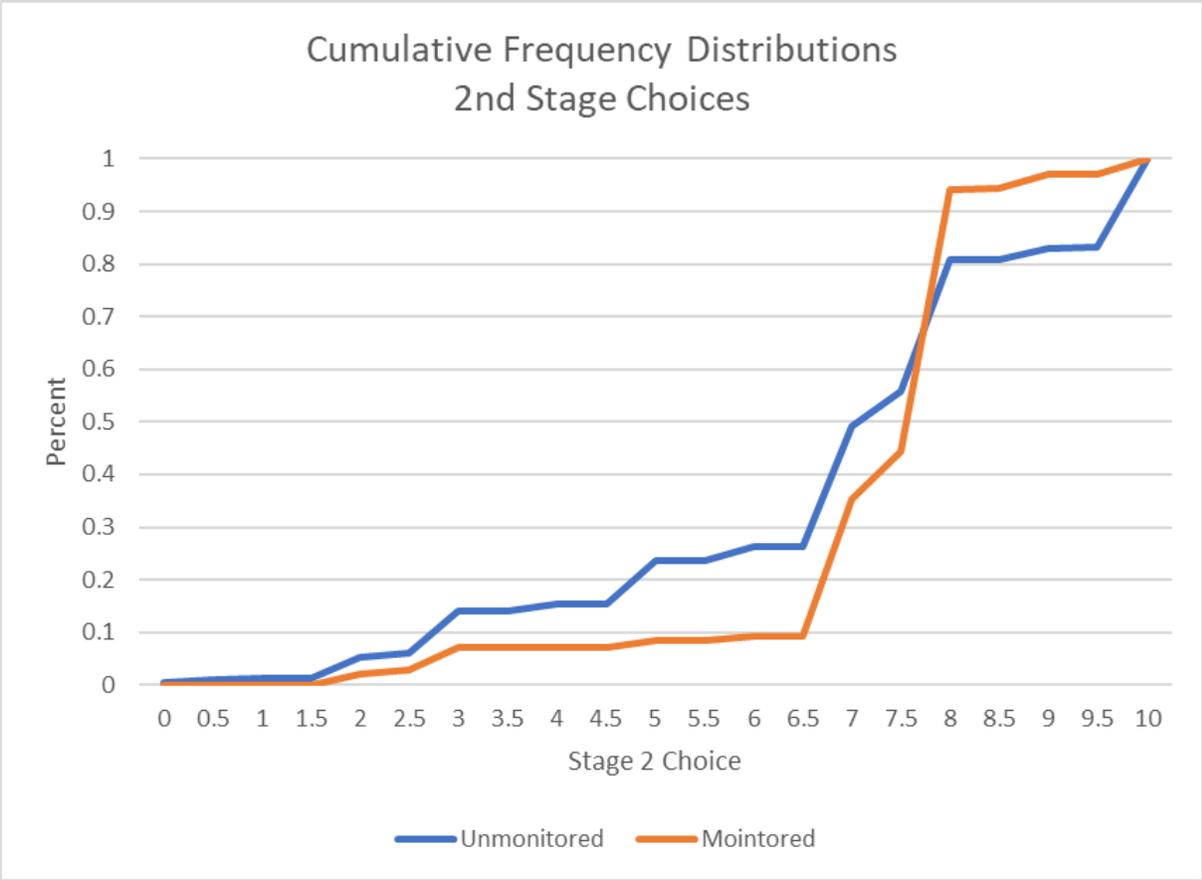


Figure A5: Cumulative frequency distributions of 2nd stage choices, unmonitored and monitored treatments

Indeed, a two-sample, two-sided, Kolmogorov-Smirnov test for the equality of distributions functions reveals that we can *strongly reject* the null hypothesis that the distributions are the same, ($D= 0.1695$, $p<0.001$). We view the latter test result as providing strong evidence that subjects’ stage 2 choices are greatly affected by whether they are monitored or not.

Further evidence that the distributions of stage 2 choices are different between the monitored and unmonitored treatments is presented in the kernel density estimates shown in Figures 2-3 of the text.

Author Contributions

Both authors contributed to the conception and design of the work and the data interpretation. In addition, Duffy contributed to the acquisition and analysis of the experimental data.

Materials and Correspondence

John Duffy

Competing Interests

None

Data Availability

The data we collected and analyzed for this paper are available for download at openIPCSR, an open access repository for social, behavioral, and health sciences research data.

A link is provided here.

<https://www.openicpsr.org/openicpsr/project/120284/version/V1/view/>

The data provided there were used to produce Figures 2 and 3 in the text and Figures A1, A2, A3, A4 and A5 and Tables A1 and A2 in the Methods Appendix.

We are happy to make available the programs we used to collect and analyze the data to any interested persons upon request.

References

- ¹ Debreu G. *Theory of Value: An Axiomatic Analysis of Economic equilibrium*. Yale University Press, 1959
- ² Neumann, John von and Morgenstern, Oskar, *Theory of Games and Economic Behavior*. Princeton, NJ. Princeton University Press, 1953.
- ³ Savage L.G. *The foundations of statistics*. John Wiley & Sons, Inc., 1954
- ⁴ Lambert Mogiliansky, A., Zamir, S., Zwirn, H., 2009. Type indeterminacy: a model of the KT (Kahneman–Tversky)-man. *Journal of Mathematical Psychology* **53(5)**, 349–361.
- ⁵ Tversky A. and E. Shafir. The Disjunction Effect in Choice under Uncertainty. *Psychological Science* **3**, 305-309
- ⁶ Wanga, Z., T. Sollowaya, R.M. Shiffrinb, and J.R. Busemeyer Context effects produced by question orders reveal quantum nature of human judgments *PNAS* **111(26)**, 2014
- ⁷ Feynman R.P., R.B. Leighton, and M. Sands *The Feynman Lectures on Physics*, Vol. III. Addison-Wesley, 1965
- ⁸ Schwinger, J., and B.G Englert. *Quantum Mechanics: Symbolism of Atomic Measurements*. Springer. 2001
- ⁹ Busemeyer J.R. and P.D. Bruza *Quantum Models of Cognition and Decision*. Cambridge University Press, 2012
- ¹⁰ Temzelides T. Modeling the Experimental Process in the Social Sciences. *Journal of Mathematical Economics* **46(4)**, 2010
- ¹¹ Krosnick, J.A., and D.F. Alwin. An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly*, **51(2)**, Pages 201–219, <https://doi.org/10.1086/269029>, 1987.
- ¹² Koch, C., and Hepp, K. Quantum mechanics in the brain. *Nature*, **440 (7084)**, 611-612, 2006.
- ¹³ De Barros J.A., and P. Suppes. Quantum mechanics, interference, and the brain, *Journal of Mathematical Psychology* **53**, 306-313, 2009.
- ¹⁴ Zizzo D. J. Experimenter Demand Effects in Economic Experiments, *Exp Econ* **13**, 75-98, 2010.
- ¹⁵ Guy C. Van Orden , Christopher T. Kello & John G. Holden. Situated Behavior and the Place of Measurement in Psychological Theory, *Ecological Psychology*, **22:1**, 24-43, DOI: 10.1080/10407410903493145, 2010.
- ¹⁶ Tversky A., and R.H. Thaler. Anomalies: Preference Reversals, *Journal of Economic Perspectives* **4(2)** 201–211, 1990
- ¹⁷ Chen, D.L., M. Schonger, C. Wickens. oTree—An open-source platform for laboratory, online, and field experiments. *J Behav. Exp Finance*, **9**,88-97, 2016.