Stochastic choice and imperfect judgments of line lengths: What is hiding in the noise?

Sean, Duffy and John, Smith

17 February 2023

Stochastic choice and imperfect judgments of line lengths: What is hiding in the noise?*

Sean Duffy† John Smith‡

February 17, 2023

Abstract

Noise is a pervasive feature of economic choice. However, standard economics experiments are not well equipped to study the noise because experiments are constrained: preferences are either unknown or only imperfectly measured by experimenters. As a result of these designs–where the optimal choice is not observable to the analyst–many important questions about the noise in apparently random choice cannot be addressed. We design an experiment to better understand stochastic choice by directing subjects to make incentivized binary choices between lines. Subjects are paid a function of the length of the selected line, so subjects will attempt to select the longer of the lines. We find a gradual (not sudden) relationship between the difference in the lengths of the lines and the optimal choice. Our analysis suggests that the errors are better described as having a Gumbel distribution rather than a normal distribution, and our simulated data increase our confidence in this inference. We find evidence that suboptimal choices are associated with longer response times than optimal choices, which appears to be consistent with the predictions of Fudenberg, Strack, and Strzalecki (2018). Although we note that the relationship between response time and the optimality of choice becomes weaker across trials. In our experiment, 54 of 56 triples are consistent with Strong Stochastic Transitivity and this is the median outcome in our simulated data. Finally, we find a relationship between choice and attention, although we find strong evidence that the relationship is endogenous.

Keywords: Stochastic transitivity, choice theory, judgment, memory, search

JEL: C91, D91

***********PRELIMINARY AND INCOMPLETE************
***********DO NOT CIRCULATE************

*We thank Roberto Barbera, Cédric Bouquet, Utteeyo Dasgupta, Andreu Mas-Colell, J.J. Naddeo, and Michel Regenwetter for helpful comments. We thank Maureen Camper, Jackie Dunn, and Steven Gussman for help with the experiment. John Smith thanks Biblioteca de Catalunya. This research was supported by a Rutgers Library OAT award, a Rutgers University Research Council Grant #FPRC0000016, and a grant from the Rutgers University-Camden College of Arts and Sciences.

†Rutgers University-Camden, Department of Psychology

‡Corresponding Author; Rutgers University-Camden, Department of Economics, 311 North 5th Street, Camden, New Jersey, USA 08102; Email: smithj@camden.rutgers.edu; Phone: +1-856-225-6319; Fax: +1-856-225-6602

1 Introduction

Noise is a pervasive feature of economic choice.[1] However, standard economics experiments are not well equipped to study the noise because experiments are constrained: preferences are either unknown or only imperfectly measured by experimenters. As a result of these designs (where the optimal choice is not observable to the analyst), many important questions about the noise in apparently random choice cannot be addressed.

The standard designs lead to the following questions: Can we design a choice experiment where the objectively optimal choice is known to the analyst? Can we make small changes in the intensity of the preference for one choice over another? Can we gain insights on the stochastic distribution of the noise in apparently random choice? Can we supplement observations of the optimality of choice with non-choice data that provides clues on the deliberation? Can we use any insights on the noise to better understand stochastic transitivity?

We design an experiment that answers 'yes' to each of these questions. In other words, we design an experiment to help uncover what is hiding in the noise.

In our induced-values experiment, subjects make binary choices between lines that differ in length and they are paid a function of the length of their selected line.[2] Classic psychology research finds that objectively measurable properties are imperfectly perceived, which can generate stochastic choice (Weber, 1834; Fechner, 1860; Thurstone, 1927a,b).[3]

Length is an attractive attribute to incentivize because its measurement is easily understood by subjects, some research finds that it is nearly perceived linearly,[4] and there is a long history of research on the topic.[5] Further, length is an uncountable quantity, so time pressure is not required to prevent subjects from counting the stimuli so that they could select the optimal choice with certainty. Additionally, because length is a continuous measure, we can

[1] For example, see Butler and Loomes (2007), Hey (2005), Hey and Orme (1994), Loomes, Starmer, and Sugden (1989), Reutskaja, Nagel, Camerer, and Rangel (2011), and Sopher and Gigliotti (1993).

[2] Our experiment is an induced-values design, in that we impose the value on subjects. Induced-values experiments have been conducted since at least Smith (1976).

[3] These insights have lead to improvements in econometrics (McFadden, 1974,2001). See Brañas-Garza and Smith (2023) for more on imperfect perception and stochastic choice.

[4] For example, 10 cm is perceived to be roughly half of 20 cm (Stevens, 1957).

[5] See Münsterberg (1894) and Cattell (1902) for early efforts, and see Laming and Laming (1992) for a translation and a discussion of a German language paper from 1852.

make arbitrarily small changes in its magnitude.[6] Finally, our induced-values design permits presenting identical sets of lines, but offering different material incentives for the choices.

Subjects can only view one line at a time. This design yields clues about the deliberation process.[7] Specifically, we can observe both the response times and the number of instances within a trial that the subject clicked to reveal a line, which we refer to as *View clicks*.

Duffy, Gussman, and Smith (2021) and Duffy and Smith (2022) also employ incentivized line judgment tasks, but with choice sets that ranged between 2 and 6 items. However, our binary choice design here permits us to address a different set of questions.[8]

For example, here we are able to glean a better understanding of the relationship between the difference in the intensity of preferences (length of the lines) and the probability the optimal choice was made. Standard economic theory predicts that the more valuable option (longer line) will be selected with certainty, regardless of the difference between the values. Here we offer 12 distinct differences between the optimal and suboptimal line lengths. We find a gradual (not sudden) relationship between economic choice and the differences in the lengths of the lines. We also find a positive relationship between both response times and view clicks and the difference in length of the optimal and suboptimal lines.

Previous work also suggests that when controlling for the difference in the lengths of the lines in the choice set, the optimal choice is decreasing in the length of the longer line. This result could have been driven by the material incentives imbedded in the choice problem. Or it could have been that longer lines are more difficult to judge than shorter lines.[9] To help disentangle possible causes, we offer two payment treatments. There is a *high payments* treatment and a *low payments* treatment, where the former pays at a higher rate.

We find that response times are longer and there are more view clicks in the high payments treatment. We interpret this as suggesting that higher incentives prompt more effort. However, we do not find a higher likelihood of optimal choice in the high payments treatment. This

[6] We admit that we are constrained by the resolution of our computer monitors. Below we discuss this matter in more detail.

[7] This design is sometimes referred to as Mouselab. See Payne, Bettman, and Johnson (1993) for a classic reference and see Willemsen and Johnson (2019) for a recent overview.

[8] Below, we discuss the relationships between these previous efforts and this paper in more detail.

[9] This is an experimental regularity in the classic psychology literature (Weber, 1834). See Brañas-Garza and Smith (2023) for an overview of this literature, designed for experimental economists.

suggests that the additional effort is not improving choice in our setting.

Additionally, we find evidence that suboptimal choices are associated with longer response times than optimal choices. This result appears to be consistent with the predictions of Fudenberg, Strack, and Strzalecki (2018). There are specifications where this relationship appears to not be driven by endogeneity. However, we also note that the relationship that optimal choices are faster than suboptimal actions tends to diminish across trials. This suggests the need–even in simple, stable choice experiments–to examine data across trials rather than analyze data that have been averaged across trials.

McFadden (1974) and Yellot (1977) demonstrate that if choice errors have a Gumbel[10] (not normal) distribution then this implies the Luce (1959) logistic stochastic choice rule.[11] Despite the significance of the stochastic distribution of the noise, to our knowledge, the only papers to investigate this in a setting where incentivized choice is an increasing function of a single, objective measure are Duffy, Gussman, and Smith (2021) and Duffy and Smith (2022). Similar to these previous efforts, we find that the errors are better described as having a Gumbel distribution rather than a normal distribution. However, this paper is the first to scrutinize the performance of the technique of identifying the error distribution with simulated data from a known error distribution. Here we conduct simulations with normally distributed errors and errors drawn from a Gumbel distribution. Our methods correctly identify the error distribution in $92 - 97\%$ of our simulations.

In standard designs, violations of transitivity play a crucial role in identifying suboptimal choices. On the other hand, in our design, we can identify suboptimal choice, without having to find intransitive choices. Nonetheless, we offer choice pairs that could produce intransitive choices so that we can better understand intransitivity in standard designs.[12] In our design, we have 56 triples where subjects make choices on $\{a, b\}$, $\{b, c\}$, and $\{a.c\}$. Our results are broadly

[10]It is unfortunate that this distribution is also referred to as the Type I extreme-value distribution, the double exponential distribution, and the log-Weibull distribution. Below, we simply refer to the distribution as *Gumbel*.

[11]See Matějka and McKay (2015) for a rational inattention justification for this specification.

[12]See Tversky (1969) for an early transitivity reference. See Regenwetter, Dana, and Davis-Stober (2011) for a more recent reference and see Regenwetter et al. (2014) for a technique useful for studying intransitive behavior in standard designs.

consistent (54 out of 56 triples) with Strong Stochastic Transitivity (SST).[13] We simulate our experimental data to better understand the context of our observations. We conduct 100 simulations of our experiment. We find that SST consistency in 54 out of 56 triples is the median and the mode outcome of these simulations. However, when we interpret these 100 simulations as a single, large experiment, we do not find any SST violations. We interpret this as suggesting that stochastic choice in our experiment would converge to consistency with SST for an increasing number of observations per triple. Finally, we find a relationship between choice and attention, although we find strong evidence of an endogenous relationship.

Our results provide clues on what is hiding in the noise and we hope that these are helpful in informing models of stochastic choice.

2 Related literature

2.1 Random utility and random choice

Because of the accumulating evidence of the presence of stochastic choice in economic data, there has also been a growth in models that can help explain the apparent noise. For example, stochastic choice can be explained by consideration sets (Masatlioglu, Nakajima, and Ozbay, 2012; Manzini and Mariotti, 2014), the preference for randomization (Agranov and Ortoleva, 2017; Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella, 2019), the preference for flexibility (Ahn and Sarver, 2013; Saito, 2015), and still others. Undoubtedly, these help us understand stochastic choice. However, it would seem that the most general explanation of stochastic choice is that decision makers imperfectly perceive their own preferences. Since we can observe the objective measure of the subjectively perceived attribute, it is our view that our experimental design is well equipped to shed light on stochastic choice.

There is a large literature in psychology demonstrating a gradual (not sudden) relationship between behavior and objective features of a choice problem. Psychologists tend to characterize the gradual relationship with what is usually referred to as a *psychometric function*.[14] The

[13] See Luce and Suppes (1965) for a classic reference on stochastic transitivity.

[14] See Gescheider (1997, Ch. 4), Falmagne (2002, Ch. 6-7), and Kingdom and Prins (2016, Ch. 4) for more on psychometric functions.

economics literature contains research on the implications of the differences in the utility—of strength of preferences—on stochastic choice (Mosteller and Nogee, 1951; Debreu, 1958). Further, there appears to be renewed interest in the topic (Ballinger and Wilcox, 1997; Blavatskyy and Pogrebna, 2010; Gerasimou, 2021; Alós-Ferrer and Garagnani, 2021,2022a,2022b). The use of line lengths as a proxy for utility has the advantage that we can make small changes in the differences in the lines and observe the effect on optimal choice.

2.2 Incentivized choice involving imperfectly perceived objects

There is a growing list of references that employ incentivized judgments on imperfectly perceived stimuli in order to better understand choice. Caplin and Dean (2015), Dean and Neligh (2019), Dewan and Neligh (2020), and Caplin, Csaba, Leahy, and Nov (2020) incentivize judgments of countable stimuli to test models of rational inattention. Duffy, Gussman, and Smith (2021) and Duffy and Smith (2022) incentivize judgments of line length to better understand stochastic choice. Other authors incentivize judgments of static or dynamic dots to test models of decision making.[15]

Still others incentivize judgments of uncountable stimuli to test models of decision making: Tsetsos, Moran, Moreland, Chater, Usher, and Summerfield (2016) employ rectangles of dynamic size, Payzan-LeNestour and Woodford (2022) employ shades of grey, and Shevlin, Smith, Hausfeld, and Krajbich (2022) employ arrays of color shades. Finally, some incentivize judgments of imperfectly perceived stimuli to add realism to experimental settings: Corgnet, Hernán-González, and Kujal (2020) employ color shade in an experimental asset market and Goryunov and Rigos (2022) employ the spatial location of a dot to represent the state space in a game.[16]

Above we list papers that describe incentivized experiments with designs involving imperfectly perceived stimuli to study choice. We hope that authors continue to think about how such designs can inform our understanding of choice.

[15] See Bhui (2019a,b), Dutilh and Rieskamp (2016), Heng, Woodford, and Polania (2020), Oud, Krajbich, Miller, Cheong, Botvinick, and Fehr (2016), Pirrone, Wen, and Li (2018), Pleskac, Yu, Hopwood, and Liu (2019) and Zeigenfuse, Pleskac, and Liu (2014).

[16] See Brañas-Garza and Smith (2023) for a more exhaustive discussion of this literature.

As previously noted, Duffy, Gussman, and Smith (2021) and Duffy and Smith (2022) also describe incentivized judgments of length. These previous efforts had choice sets that varied in size (2-6 items), rather than the binary choice we study here. The binary choice design simplifies the specifications of subject-specific heterogeneity, simplifies the characterization of the choice sets (including the previous trials and the spatial locations of the lines), and renders simulated data more straightforward to produce. Additionally, the previous efforts only contained 3 differences between longest and next longest lines, whereas here we have 12 differences. We are therefore better equipped to study the sensitivity of the differences in intensity with choice. We also note that these previous efforts did not vary the material incentives and there were no simulated data. Hence, while some of the results overlap, the design that we employ here can address different matters than those in the previous designs.

2.3 Response times and choice

Research finds that response times tend to be increasing in decisions that are closer to indifference.[17] In our data, we find that larger absolute differences between the line lengths are associated with shorter response times.

Also in our data, we can observe whether the uniquely optimal choice was made. Therefore we can investigate the relationship between the optimality of choice and the response times. Research has found a negative relationship between the optimality of choice in perceptual judgments and response times.[18]

Fudenberg, Strack, and Strzalecki (2018) analyze a model of a decision maker with an unknown utility and a cost of acquiring information about the choice options. The authors derive a prediction in this setting: optimal choices will have shorter response times than suboptimal choices.

Consistent with these predictions, we find evidence that suboptimal decisions are associated

[17] For instance, see Henmon (1911), Volkmann (1934), Mosteller and Nogee (1951), Hey (1995), Moffatt (2005), Woodford (2014), Alós-Ferrer, Granić, Kern, and Wagner (2016), Echenique and Saito (2017), Konovalov and Krajbich (2019), Alós-Ferrer, Fehr, and Netzer (2021), Duffy, Gussman, and Smith (2021), Alós-Ferrer and Garagnani (2022a), and Duffy and Smith (2022).

[18] For instance, see Henmon (1911), Kellogg (1931), Bhui (2019b), Duffy, Gussman, and Smith (2021), and Duffy and Smith (2022).

with longer responses times than are optimal decisions. We note the possible presence of endogeneity in this econometric analysis. However, in some specifications we do not find evidence of endogeneity.

2.4 Payment treatments

There is a literature that examines the role of incentives by manipulating the material payoffs and observing the effects on behavior (Ariely, Gneezy, Loewenstein, and Mazar, 2009; Charness and Kuhn, 2011; Dickinson, 1999; Sillamaa, 1999a, 1999b). In fact, some research suggests a non-monotonic relationship between performance and payment (Gneezy and Rustichini, 2000; Pokorny, 2008). See Camerer and Hogarth (1999), Read (2005), and Kamenica (2012) for overviews on the effects of incentives.

There is also an existing literature where payments are manipulated in settings with imperfectly perceived but objectively measurable features.[19] Caplin, Csaba, Leahy, and Nov (2020) present subjects with a geometric shape judgment task and vary incentives. Dean and Neligh (2022) present subjects with a dot judgment task and vary incentives. Dewan and Neligh (2020) present subjects with a dot estimation task and vary incentives. Each set of authors finds a positive relationship between incentives and accuracy in the countable task.

On the other hand, Bhui (2019a) presents subjects with a dot motion task and finds that subjects are largely not sensitive to the material incentives treatments. While the dots are countable, they only appear for a short time, so that counting is not feasible.

Although we find evidence that subjects expend more effort in the high payment treatment, we do not find evidence of greater accuracy. It is not clear to us whether our payment treatments were not sufficient to prompt differences in accuracy or simply whether effort does not directly translate to accuracy in our line judgment task.

[19] Also see Civelli, Deck, and Tutino (2022) for evidence that payments affect attention.

3 Experimental design

3.1 Line selection task

Subjects made binary choices between lines.[20] There were 60 unique pairs of lines, which had lengths ranging from 120 pixels (5.59 cm) to 320 pixels (14.90 cm). The length of the lines in any pair was determined by subtracting various amounts from the *longer line*. The differences between the lines ranged from 1 pixel (0.0465 cm) to 40 pixels (1.86 cm).[21] For convenience, we hereafter refer to line lengths in pixels, rather than in cm.

There was a line situated on the left of the screen and a line situated on the right, although neither were visible at the beginning of each trial. The longer of the lines was programmed to be on the left side of the screen with an independent probability of 0.5. When subjects clicked on any point on the left (right) half of the screen, the left (right) line would appear and the right (left) line would disappear. There were no time or click restrictions: subjects could click back and forth to view the lines as many times and for as long as they preferred.

To indicate their choice, there was a box on the left half of the screen and a box on the right half of the screen. Both boxes were near the upper screen edge. At the beginning of each trial, both unselected boxes were empty. Subjects would indicate their choice by clicking on the box that corresponds to the line. A click on an empty box would fill that box with a smaller black box. Subjects could change their choice any number of times within the trial. When they were satisfied with their choice, they could end the trial by pressing Enter or space, and proceed to the next trial. Subjects were only able to end the trial if both lines were viewed at least once and a line was selected. Figures 1 and 2 offer screenshots from the line selection task.

[20] The lines each had a height of 0.36 cm and were the identical shade of grey.

[21] Our computer screens have a width of 47.664 cm and our program divided this width into 1024 pixels. We could have made our pixels smaller. However, the width of our screens have a maximum resolution of 1920 pixels. Therefore, the minimum possible pixel width would be 0.0244 cm.

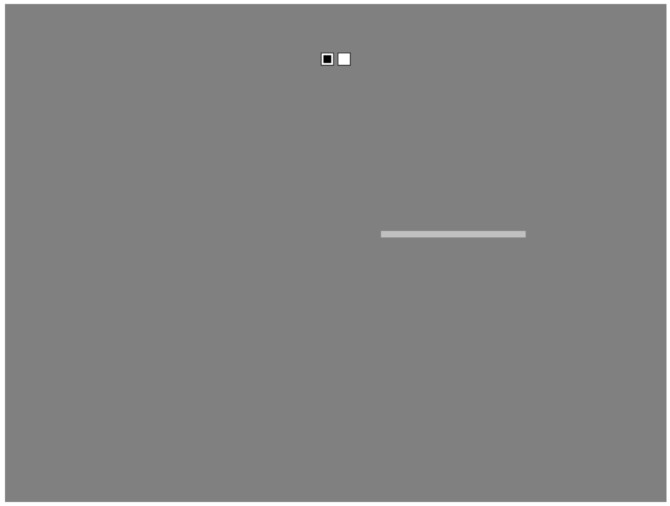

Figure 1: Shows that the right line is being viewed and the left line is selected.

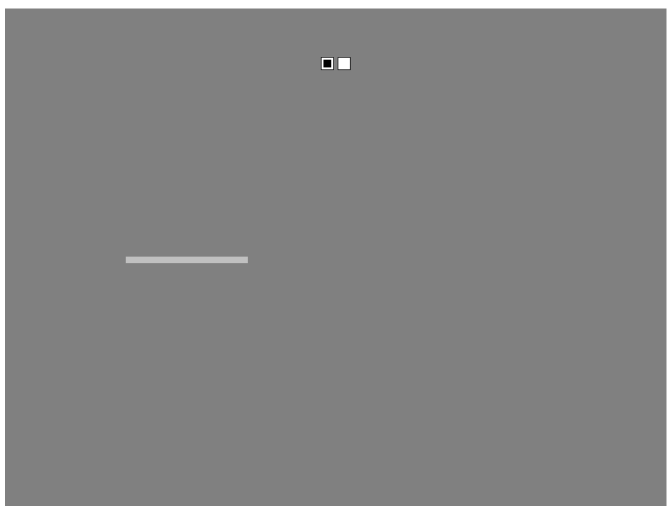

Figure 2: Shows that the left line is being viewed and also the left line is selected.

3.2 Payment treatments

Subjects made a choice from the 60 line pairs in the *high payment* treatment and a choice from the 60 line pairs in the *low payment* treatment, for a total of 120 choices.

Subjects were paid as a function of the length of the selected line. Payments ranged from \$0 to \$20 in the high payment treatment and from \$0 to \$10 in the low payment treatment. The terms "high" and "low" were not used in the experiment, but rather "\$0-\$20" or "\$0-\$10" respectively. Recall that the longest line was 320 pixels and the shortest line was 120 pixels. If a line of pixel length x is selected, the payment in that trial is:

$$y * \frac{(x - 120)}{(320 - 120)},$$

where $y = \$20$ in the high payment treatment and $y = \$10$ in the low payment treatment.

The payment treatments were arranged into alternating blocks of 20 trials. Subjects were given the high payment in the first block with an independent probability of 0.5. At the start of each block, subjects were presented with a screen for 20 seconds, informing them of the payment treatment for the next 20 trials. Further, before every line selection trial, subjects were presented with a screen for 2 seconds, reminding them of the payment treatment in the upcoming trial.

3.3 Survey questions

After every line trial was completed, but before the subjects were paid, subjects were given a set of survey questions, administered via paper.[22] We elicited the preferred pronoun of the subject, the handedness (right or left) of the subject, unincentivized responses to the standard version of the three Cognitive Reflection Test (CRT) questions (Frederick, 2005), and an optional estimate of their grade point average.[23]

3.4 Payment details

One of the 60 high payment treatment trials and one of the 60 low payment treatment trials were randomly selected for payment. Subjects were also paid a \$5 show-up fee. Immediately following the experiment, subjects were paid in cash, rounded up to the nearest \$0.25. The average payment was \$22.50.

[22] See https://osf.io/us9n6/ for the survey.
[23] There were 99 valid responses to this item.

3.5 Experimental Details

The experiment was programmed on E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA).[24] A total of 121 subjects participated in the experiment. However data from one subject did not properly record, and we were not able to recover any data on their line selections.[25] Further, we exclude from the analysis another subject who selected the longer line on less than half of the trials and did not complete the survey.[26] Therefore, our dataset has 119 subjects each making 120 pairwise line judgments, for a total of $14,280$ line judgments. The data and screenshots are available at https://osf.io/us9n6/.

3.6 Discussion of the design

Subjects must view both lines and must make a choice before the program will allow the subject to end the trial and proceed to the following trial. Beyond these restrictions, we do not place any restrictions on the number of line views, the response times, or the number of times the subject can change their preliminary selection before ending the trial. This design is intended to affect the response time as little as possible.

Our line pairs produce 56 triples with choice sets of $\{a,b\}$, $\{b,c\}$, and $\{a,c\}$, which allows us to study (stochastic) transitivity. See the appendix for a list of the line lengths and percent of optimal choices in the 60 line pairs (Table $A10$). Also see the appendix for a list of the line lengths and simulated SST violations in these 56 triples (Table $A11$).

4 Results

4.1 Gradual relationship between choice and differences in length

In every trial, subjects have an incentive to select the longer line. In order to study the effect of small changes in the differences in preference intensity, we present subjects with 12 absolute differences in the length of the longer and shorter lines, which range from 1 pixel to 40 pixels.

[24] See the appendix for details about the computer settings and the trial-specific random offsets of the line positions in Table $A8$.

[25] This subject (101) does not appear in the final dataset.

[26] This subject (126) does appear in the final dataset, but is excluded from the analysis.

Below we characterize the relationship between the difference in the lengths of the lines in the choice set and the probability of selecting a particular line. In Figure 3, on the horizontal axis, we plot the length of the line on the right minus the length of the line on the left. On the vertical axis we plot the fraction of trials where the right line was selected. We also include a best-fitting logistic *psychometric* function.

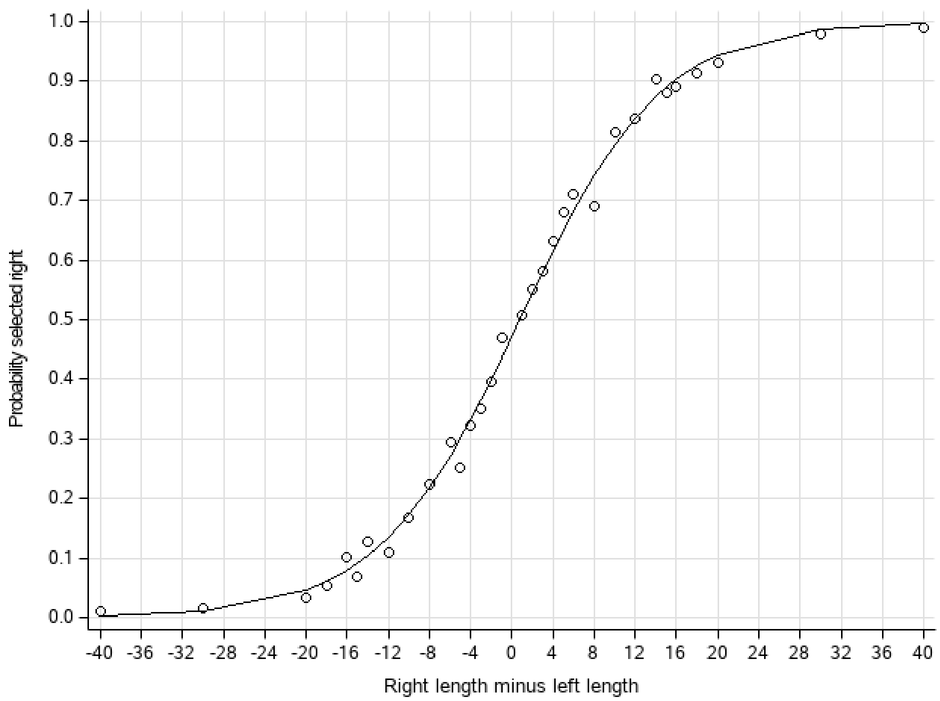

Figure 3: The relationship between the length in pixels of the right line minus the length in pixels of the left line and the probability of selecting the line on the right.

Standard economic theory predicts that subjects will select the more valuable option (longer line) with certainly. However, in contrast to these predictions, we observe a gradual relationship between choice and the differences in the lengths. Figure 3 also suggests that there is a slight overall bias toward selecting the left line over the right line.

4.2 Optimality of choice

Whereas the previous subsection characterized choice exclusively as a function of the differences in the lengths of the lines, here we conduct a more detailed analysis. We define the

Selected longer variable to be 1 if the choice was the longer line, and 0 otherwise. We conduct logistic regressions with the Selected longer variable as dependent variable. The *Absolute difference* variable is the absolute value of the difference between the line lengths in pixels. The *Longer length* variable is the length, in pixels, of the longer line. The *High payment* variable is a dummy variable that has a value of 1 in the high payment treatment, and 0 otherwise. We also include *Trial* and the Trial-High payment interaction as independent variables.

Due to the repeated nature of the observations, we offer fixed-effects specifications where we estimate dummy variables for each subject. These are designed to capture possible heterogeneity in subject-specific skill in performing the task. We also run fixed-effects specifications where we attempt to account for a possible subject-specific bias that favors the left or right lines. We summarize these regressions in Table 1.

Table 1: Logistic regressions of the Selected longer line variable

	(1)	(2)	(3)	(4)
Longer length	-0.00062^{***}	-0.00062^{***}	-0.00063^{***}	-0.00063^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Absolute difference	0.0234^{***}	0.0234^{***}	0.0229^{***}	0.0229^{***}
	(0.00063)	(0.00063)	(0.00058)	(0.00058)
High payment	-0.0033	-0.0192	-0.00134	-0.0088
	(0.0067)	(0.0142)	(0.0063)	(0.0133)
Trial	-0.00030^{**}	-0.00043^{**}	-0.00029^{**}	-0.00036^{**}
	(0.0001)	(0.00014)	(0.0001)	(0.00013)
High payment*Trial	-	0.00026	-	0.00012
		(0.00020)		(0.00019)
Subject fixed effects	Yes	Yes	Yes	Yes
Subject left dummies	No	No	Yes	Yes
Demographics	No	No	No	No
AIC	14038.37	14038.75	12670.91	12672.50

We provide the average marginal effects and the standard errors in parentheses. We do not provide the estimates of the intercepts or the subject-specific dummy estimates. AIC refers to the Akaike information criterion (Akaike, 1974). Each regression has $14,280$ observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Consistent with Figure 3, we observe that the differences in the lengths of the lines has a strong effect on selecting the longer line. We also note a negative relationship between the length of the longer line and optimal choice. We do not find a consistent statistically significant effect of the payment treatment. We also note that choice becomes less optimal across trials but we do not find evidence that this trajectory is affected by the payment treatment.[27]

In specifications (3) and (4) there appears to be heterogeneity in subjects having a bias toward the left or the right lines. However, in specifications where we do not control for subject-specific bias for favoring left or right, we can test whether there is an aggregate bias toward the left or right. We find that when the longer line is on the right, it is significantly less likely to be selected than when the longer line is on the left.[28] This provides evidence of the conjecture from Figure 3: there is an overall bias toward selecting the line on the left over the line on the right.[29] We also note that this left bias is not significantly different across trials.

We also test for other effects of the high payment treatment. We include specifications with the High payments-Absolute difference interaction and the High payments-Longer length interactions. These interactions are not significant.[30]

In every trial, the lines are randomly offset in the vertical and horizontal directions. See the appendix for more on these offsets. We conduct an analysis, similar to Table 1, but we include variables accounting for these offsets. We find a bias toward selecting the line closer to the nearest (left or right) screen edge and a bias toward selecting the line with the higher vertical presentation.[31]

Psychology research has also discovered sequential effects, whereby previous stimuli or previous responses might be related to subsequent responses (Verplanck, Collier, and Cotton,

[27] See the appendix for alternate specifications of Table 1. We run tobit regressions with a dependent variable of the longer line length minus the selected line length, which appears as Table $A1$. Our results are not changed. Also, see the appendix for the analysis involving the *demographics* of the subjects: whether the subjects reported being left handed, whether the subjects reported being preferring the 'her' pronoun, their CRT score, and self-reported GPA. This version of Table 1 appears as Table $A5$ and the results discussed here are not changed.

[28] These results are available from the corresponding author upon request.

[29] Research finds that subjects devote more attention to options on the left and they are more likely to select options on the left (Bowers and Heilman, 1980; Nicholls, Bradshaw, and Mattingley, 1999; Charles, Sahraie, and McGeorge, 2007). This bias toward the left is sometimes referred to as *pseudoneglect*.

[30] These results are available from the corresponding author upon request.

[31] This analysis is summarized in Table $A8$. See Olsson and Winman (1996) and Baranski and Petrusic (1999) for screen edge effects in spatial judgments. See Nicholls, Mattingley, Berberovic, Smith, and Bradshaw (2004) and Suavansri, Falchook, Williamson, and Heilman (2012) for the bias for items with a higher vertical presentation. This upward bias is sometimes referred to as *altitudinal pseudoneglect*.

1952; Jesteadt, Luce, and Green, 1977; Treisman and Williams, 1984; Petzold and Haubensak, 2004). We run an analysis similar to that summarized in Table 1, but we include variables accounting for details in the previous trial. We find some possibly interesting sequential effects.[32]

4.3 Non-choice data providing clues about deliberation

In an effort to interpret these optimal choice results, we perform similar analyses, but with two non-choice measures of effort. Our first measure is the number of times within a trial that the subject clicked to reveal one of the two lines. We refer to this variable as *View clicks*. Our second measure is the length of time from the start of the trial to the time which the subject ended the trial. Consistent with the literature, we perform our analyses by first taking the log of response times. Because these are continuous variables, we perform linear regressions. The regressions with View clicks as dependent variable are summarized in Table 2.

Table 2: Regressions of the View clicks variable

	(1)	(2)	(3)	(4)
Longer length	0.0026*	0.0026*	0.0024*	0.0024*
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Absolute difference	−0.111***	−0.111***	−0.111***	−0.111***
	(0.004)	(0.004)	(0.004)	(0.004)
High payment	0.173*	0.435*	0.155†	0.417*
	(0.087)	(0.180)	(0.087)	(0.180)
Trial	−0.0168***	−0.0146***	−0.0168***	−0.0147***
	(0.0013)	(0.0018)	(0.0013)	(0.0018)
High payment*Trial	-	−0.0043†	-	−0.0043†
		(0.0026)		(0.0026)
Subject fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
Subject left dummies	*No*	*No*	*Yes*	*Yes*
Demographics	*No*	*No*	*No*	*No*
AIC	87337.0	87344.3	86850.3	86857.6

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts or the subject-specific dummy estimates. AIC refers to the Akaike information criterion. Each regression has

[32] We report these results in Table *A*9. We note that some psychologists might interpret our results as consistent with a moving, temporary threshold between selecting the right or left line.

14,280 observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Consistent with the optimal choice results, we observe more View clicks when both the absolute difference between the lines is smaller and when the length of the longer line is longer. It seems as if more effort is expended when the decision is more difficult: smaller difference between the line lengths and when the lengths of the lines are longer. The coefficient estimate of High payments is positive and significant. With the measure of View clicks, it seems as if the high payment treatment prompts more effort. We also see that View clicks decrease across trials and that there is some evidence that this decrease is stronger in the high payment treatment.[33]

We next examine whether the analogous results hold with the log of Response time. The regressions are summarized in Table 3.

Table 3: Regressions of the log of Response times variable

	(1)	(2)	(3)	(4)
Longer length	0.00034***	0.00034***	0.00034***	0.00034***
	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Absolute difference	−0.00610***	−0.00611***	−0.0061***	−0.0061***
	(0.00018)	(0.00018)	(0.00017)	(0.00017)
High payment	0.00798*	0.0230**	0.00768*	0.0225**
	(0.00351)	(0.0073)	(0.00349)	(0.00726)
Trial	−0.0029***	−0.0027***	−0.0029***	−0.0027***
	(0.00005)	(0.00007)	(0.00005)	(0.00007)
High payment*Trial	-	−0.00025*	-	−0.00025*
		(0.00011)		(0.00011)
Subject fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
Subject left dummies	*No*	*No*	*Yes*	*Yes*
Demographics	*No*	*No*	*No*	*No*
AIC	−3486.8	−3475.9	−3298.3	−3287.2

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts or the subject-specific dummy estimates. AIC refers to the Akaike information criterion. Each regression has 14,280 observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

[33] See Table *A6* for the demographics version of Table 2. The results discussed here are not changed, with the exception that the High payment variable is not significant.

Similar to the results summarized in Table 2, we observe that subjects have longer response times for choice pairs with smaller differences between the lines and for lines with longer lengths. We also find that subjects expend a longer time deliberating about choices in the high payment treatment. Similarly, we find that response times decrease across trials and that this effect is stronger in the high payment treatment.[34]

These non-choice variables suggest that, while subjects do not successfully select the optimal choice in the high payment treatment more frequently than in the low payment treatment, they seem to expend more effort in these high payment trials.

4.4 The relationship between optimality and response time

Here we test the predicted negative relationship between response times and optimal choices (Fudenberg, Strack, and Strzalecki, 2018). We conduct an analysis, similar to that summarized in Table 3, but we include Selected longer as an independent variable. We also include the interaction of Selected longer and Trial. Table 4 summarizes this analysis.

Table 4: Regressions of the log of Response time variable

	(1)	(2)	(3)	(4)
Longer length	0.00032***	0.00032***	0.00033***	0.00033***
	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Absolute difference	−0.0056***	−0.0056***	−0.0057***	−0.0057***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
High payment	0.0078*	0.0077*	0.0077*	0.0075*
	(0.0035)	(0.0035)	(0.0035)	(0.0035)
Trial	−0.0029***	−0.0031***	−0.0029***	−0.0031***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Selected longer	−0.0418***	−0.0616***	−0.0267***	−0.0470***
	(0.0043)	(0.0084)	(0.0045)	(0.0085)
Trial*Selected longer	-	0.00032**	-	0.00033**
		(0.00012)		(0.00012)
Subject fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
Subject left dummies	*No*	*No*	*Yes*	*Yes*
Demographics	*No*	*No*	*No*	*No*
AIC	−3573.0	−3564.2	−3324.6	−3316.3

[34] See Table *A*2 for the version of Table 3 where response time, rather than the log of Response time, is the dependent variable. With the exception that the High payment variable is not significant, the results are not changed. See Table *A*7 for the demographics version of Table 3, and the results discussed here are not changed.

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts, the subject-specific dummy estimates, or the demographics estimates. AIC refers to the Akaike information criterion. Each regression has $14,280$ observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

In every specification, the Selected longer variable is negative and significant. This suggests that, when controlling for both the heterogeneity of the choice settings and the heterogeneity of the subjects, suboptimal choices tend to take longer than optimal choices. We also note that this effect diminishes across trials.[35]

Of course, we *should* be concerned about possible endogeneity. For example, it is possible that we have not sufficiently accounted for the subject heterogeneity. If not, then our results could be driven by more accurate subjects also tending to have shorter response times. It is also possible that we have not sufficiently accounted for the choice set heterogeneity. If not, then our results could be driven by more difficult choice problems are both less accurate and slower.

Due to clearly warranted concerns of endogeneity, we check whether the Selected longer variable is correlated with the errors in the regressions. When we conduct Spearman correlations between unstandardized residuals and the Selected longer variable in specifications $(1)-(4)$, the p-values (respectively) are 0.4213, 0.5334, 0.4450, and 0.5623. Our results are not changed if we use student residuals rather than the unstandardized residuals.[36] We interpret this as suggesting a lack of evidence of endogeneity in these regressions.

On the other hand, we note that the specifications offered in Table $A3$, with raw response times rather than log of response times, exhibit evidence of endogeneity: the 8 correlations are significant with p-values each less than 0.001. We suspect that the raw response time results might be differentially affected by long response times in a way that does not affect the specifications with log of response times. Regardless, we find some evidence that the relationship between response times and optimal choice is not driven by endogeneity.

[35] See Table $A3$ for the version of Table 4 where response time, rather than the log of response time, is the dependent variable. Again, the results are not changed, with exception that the high payment treatment is not significant.

[36] These Spearman correlations have p-values of 0.4211, 0.5331, 0.4550, and 0.5622.

4.5 Gumbel or normal errors better fit the data?

A common specification in a random utility models (RUM) is that there is a non-stochastic component of the utility function and an additive stochastic component.[37] For example, subject i on trial t, would have utility for option $j \in \{L, R\}$:

$$U_{i,t}^{j} = V_{i,t}^{j} + \epsilon_{i,t}^{j},$$

where $V_{i,t}$ is the non-stochastic component and $\epsilon_{i,t}$ is the random component. The subject i on trial t selects left when:

$$V_{i,t}^{L} + \epsilon_{i,t}^{L} \geq V_{i,t}^{R} + \epsilon_{i,t}^{R}.$$

We offer several specifications. For example, Specification A conducts a single estimate of β:

$$V_{i,t} = \beta * f(length_{i,t}).$$

Specification B, conducts a single estimate of β^L and β^R:

$$V_{i,t} = \beta^{\,j \in \{L,R\}} * f(length_{i,t}).$$

Specification C, conducts a single estimate of β and a single estimate of a Left Dummy:

$$V_{i,t} = \beta * f(length_{i,t}) + LeftDummy.$$

Specification D, conducts a single estimate of β and subject-specific Left Dummy estimates:

$$V_{i,t} = \beta * f(length_{i,t}) + LeftDummy_i.$$

[37] For instance, see McFadden (1974).

Specification E, conducts subject-specific estimates of β and a single Left Dummy:

$$V_{i,t} = \beta_i * f(length_{i,t}) + LeftDummy.$$

Finally, Specification F, conducts subject-specific estimates of β and subject-specific Left Dummy estimates:

$$V_{i,t} = \beta_i * f(length_{i,t}) + LeftDummy_i.$$

Within these general specifications, we also vary the specification of $f(.)$. We offer a linear specification:

$$f(x) = x.$$

However, it is not clear that the linear specification is the most appropriate. There are researchers who argue that there is not a linear relationship between stimuli and the perception of stimuli, but rather they are related by the log function (Fechner, 1860).[38] According to this classic view in psychology, the non-stochastic component of the utility function should have a specification of:

$$f(x) = \log(x).$$

There are other researchers who argue that the relationship between stimuli and the perception of the stimuli is neither linear nor logarithmic, but is rather described by the power function (Stevens, 1961).[39] Research suggests that the specification of this power function depends on the type of stimulus. Some research suggests that the exponent in the power function, when the stimulus is length, is 1.04 (Teghtsoonian, 1971). This research suggests a specification of:

[38] This is a version of what is sometimes referred to as Fechner's Law. See Falmagne (2002) for a sympathetic modern interpretation of Fechner's Law. See Brañas-Garza and Smith (2023) for more on Fechner's Law for the economist reader.

[39] See Brañas-Garza and Smith (2023) for more on Stevens' Power Law for the economist reader.

$$f(x) = (x)^{1.04}.$$

We conduct a probit (which assumes normally distributed errors) and a logit (which assumes Gumbel distributed errors) multinomial discrete choice (MDC) analysis on each of these 18 specifications. Table 5 reports the AICs for each specification.

Table 5: Comparisons of different multinomial discrete choice models

	Linear		Log		Power	
	Logit	Probit	Logit	Probit	Logit	Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Spec. A	14,169	14,221	14,091	14,133	14,176	14,229
Spec. B	14,130	14,181	14,063	14,106	14,137	14,189
Spec. C	14,145	14,196	14,067	14,111	14,152	14,205
Spec. D	13,051	13,119	12,959	13,019	13,060	13,129
Spec. E	13,927	13,931	13,846	13,904	13,934	13,939
Spec. F	12,841	12,847	12,749	12,748	12,849	12,855

We provide the Akaike Information Criterion (AIC) for the various models. Every specification has 14,280 pair-wise choices.

In all but the log version of Specification F, we find a smaller AIC for the logit than the probit. This analysis suggests to us that the errors are better described as having a Gumbel distribution, rather than a normal distribution.

One question is whether our MDC technique *could* successfully identify the distribution of the errors. To address this question, we simulate data with errors from a known distribution. We conduct 100 simulations of normally distributed errors with a zero mean and standard deviations of each of 7.5, 8.0, 8.5, and 9.0. We also conduct 100 simulations of errors with a Gumbel distribution with a zero mean and standard deviations of each of 7.5, 8.0, 8.5, and 9.0. These values were selected because distributions with a standard deviation of 7.5 tend to have AICs less than those in Specification A reported in Table 5 and distributions with standard deviation of 9.0 tend to have AICs larger than those in Specification A reported in Table 5. The subjects in our simulations are identical and do not have biases for the left or right lines. With this simulated data, we conduct a probit MDC, a logit MDC, and we compare the AICs.[40] We provide a summary of the analyses of these simulations in Table 6.

Table 6: Comparisons of different multinomial discrete choice models

	Probit AIC smaller	AICs equal	Logit AIC smaller
Normal errors ($SD = 7.5$)	94	1	5
Normal errors ($SD = 8.0$)	92	1	7
Normal errors ($SD = 8.5$)	96	1	3
Normal errors ($SD = 9.0$)	93	1	6
Gumbel errors ($SD = 7.5$)	2	1	97
Gumbel errors ($SD = 8.0$)	2	1	97
Gumbel errors ($SD = 8.5$)	3	3	94
Gumbel errors ($SD = 9.0$)	3	2	95

We provide the Akaike Information Criterion (AIC) for the various models with 100 simulations. Every simulation has $14,280$ pairwise choices.

When we generate errors with a normal distribution, the probit analyses have smaller AICs on $92\% - 96\%$ of the simulations. When we generate errors with a Gumbel distribution, the logit analyses have smaller AICs on $94\% - 97\%$ of the simulations. We interpret these simulations as suggesting that the analysis in Table 5 would likely correctly identify the underlying errors as either normally distributed or having a Gumbel distribution.

4.6 Stochastic transitivity

Recall that we have 56 triples where choices were made between $\{a, b\}$, $\{b, c\}$, and $\{a, c\}$, where $a > b > c$. This permits us to test various specifications of stochastic transitivity. Since we have 119 subjects responding twice to every pair, we have 238 observations per pair.

We find that 56 of these 56 triples satisfy moderate stochastic transitivity (MST):[41]

$$\begin{aligned} \text{If } \Pr(a &\succsim b) \geq 0.5 \text{ and } \Pr(b \succsim c) \geq 0.5 \text{ then} \\ \Pr(a &\succsim c) \geq \min\{\Pr(a \succsim b), \Pr(b \succsim c)\}. \end{aligned}$$

[40] These MDC simulations are available at https://osf.io/us9n6/.

[41] Hence each these 56 triples satisfy the *triangle inequality* ($\Pr(a \succsim b) + \Pr(b \succsim c) - \Pr(a \succsim c) \leq 1$) and *weak stochastic transitivity* (If $\Pr(a \succsim b) \geq 0.5$ and $\Pr(b \succsim c) \geq 0.5$ then $\Pr(a \succsim c) \geq 0.5$). See Luce and Suppes (1965) for a characterization of the relationships among the definitions of stochastic transitivity.

We also find that 54 of these 56 triples satisfy *strong stochastic transitivity (SST)*:[42]

$$\begin{aligned} \text{If } \Pr(a &\succsim b) \geq 0.5 \text{ and } \Pr(b \succsim c) \geq 0.5 \text{ then} \\ \Pr(a &\succsim c) \geq \max\{\Pr(a \succsim b), \Pr(b \succsim c)\}. \end{aligned}$$

In order to better understand the context of these stochastic transitivity results, we generate 100 simulations of the experiment. Specifically, we are interested to learn the likelihood of SST violations. We perform the simulations with errors drawn from a Gumbel distribution with zero mean and standard deviation of 8. We select this standard deviation because the AICs seem closest to those in the linear specifications in Table 5. As in our non-simulated data, each simulated pair choice has 238 observations in each simulated experiment. Note that subjects in our simulations are *a priori* identical and non-biased. Table 7 summarizes the distribution of SST violations in these 100 simulations.[43] For example, in 10 instances there are no SST violations, in 18 instances there is a single violation, etc.

Table 7: Distribution of 100 simulated experiments

56 out of 56 triples satisfy SST	10
55 out of 56 triples satisfy SST	18
54 out of 56 triples satisfy SST	31
53 out of 56 triples satisfy SST	19
52 out of 56 triples satisfy SST	14
51 out of 56 triples satisfy SST	6
50 out of 56 triples satisfy SST	0
49 out of 56 triples satisfy SST	2
$<$ 49 out of 56 triples satisfy SST	0

The results of the simulations suggest a stochastic distribution of the number of SST

[42] The Table $A10$ ID of these two triples are 30 $(238, 236)$, 29 $(236, 232)$, 32 $(238, 232)$ and 37 $(242, 236)$, 28 $(236, 234)$, 38 $(242, 234)$. The Table $A11$ ID of these two triples are 53 and 46.

[43] See Table $A11$ for the distribution of simulated SST violations for each triple. These SST simulations are available at https://osf.io/us9n6/.

violations. In the (non-simulated) experimental data, we found 54 out of 56 SST violations. Our simulations suggest that 54 out of 56 triples satisfying SST is both the median and the mode outcome.

On the other hand, rather than treat the simulations as 100 separate experiments, another interpretation is to pool the observations as regard them as a single, large experiment. In this large experiment, each choice pair is based on $23,800$ observations, not 238 observations. We therefore expect fewer idiosyncratic variations in the choice probabilities. In this large experiment interpretation, we do not find a single instance of an SST violation. This suggests to us that the processes in our experiment would converge to satisfying SST with increasing numbers of observations.

4.7 The relationship between attention and choice

Here we investigate the relationship between attention and choice. Some research suggests a positive relationship between the time that a subject spends being attentive to an object in a choice setting and the likelihood of selecting the object (Armel and Rangel, 2008; Krajbich, Armel, Rangel, 2010; Krajbich and Rangel, 2011). In fact, some authors assert a causal effect of attention on choice (Krajbich, 2019).

In our setting, one plausible measure of attention is the total time spent viewing a line. We perform an analysis, similar to Table 1, but with *Time viewing longer* as an independent variable. This variable is calculated by summing the time, possibly across multiple durations, which the subject viewed the longer line. We also include the Time viewing longer-Trial interaction. We summarize these regressions in Table 8:

Table 8: Logistic regressions of the Selected longer line variable

	(1)	(2)	(3)	(4)
Longer length	-0.00064^{***}	-0.00064^{***}	-0.00065^{***}	-0.00065^{***}
	(0.00009)	(0.00009)	(0.00009)	(0.00009)
Absolute difference	0.0236^{***}	0.0236^{***}	0.0230^{***}	0.0230^{***}
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
High payment	-0.0035	-0.0035	-0.0017	-0.0018
	(0.0067)	(0.0067)	(0.0063)	(0.0063)
Trial	-0.00005	-0.00074^{***}	-0.00002	-0.00068^{***}
	(0.00010)	(0.00014)	(0.00010)	(0.00013)
Time viewing longer	0.0081^{***}	0.0011	0.0083^{***}	0.0017
	(0.0012)	(0.0014)	(0.0011)	(0.0013)
Trial*Time viewing longer	-	0.00021^{***}	-	0.00020^{***}
		(0.00003)		(0.00003)
Subject fixed effects	Yes	Yes	Yes	Yes
Subject left dummies	No	No	Yes	Yes
Demographics	No	No	No	No
AIC	13987.02	13942.75	12605.08	12559.71

We provide the average marginal effects and the standard errors in parentheses. We do not provide the estimates of the intercepts, the subject-specific dummy estimates, or the demographics estimates. AIC refers to the Akaike information criterion (Akaike, 1974). Each regression has $14,280$ observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

The coefficient estimates for Time viewing longer are positive and significant. This suggests a relationship between the optimality of the choice and time viewing the longer line. A similar result is reported by Krajbich and Rangel (2011). We also see that the interaction estimates are positive and significant. This provides evidence that the relationship between attention and choice becomes stronger across trials.

Similar to the analysis summarized in Table 4, we need to exercise caution because of the possibility of endogeneity introduced by including the Time viewing longer variable. For example, it is possible that there is no causal effect of attention on choice. Rather it is possible that the Time viewing longer might simply be a reflection of the (possibly incorrectly) perceived value of the line. In this case, then there will be a correlation between the Time viewing longer variable and the errors.

When we conduct Spearman correlations between the unstandardized residuals and the Time viewing longer variable in specifications (1) – (4), p-values are each less than 0.001.

When we conduct Spearman correlations between Pearson standardized residuals and the Time Viewing Longer variable, we also find p-values each less than 0.001. This is strong evidence of endogeneity.

In our dataset, we find that a measure of attention is positively related to choice and this at this effect becomes stronger across trials. However, we find strong evidence of endogeneity in this relationship.

5 Conclusion

We present subjects with an incentivized, induced-values choice experiment where we can observe the optimal choice in addition to important non-choice data.

Unlike the prediction of classic economic theory, we find a gradual (not sudden) relationship between optimal choice and the differences intensity of the preference for one choice over the other. Additionally, we find that our payment treatments affect the non-choice behavior in a way that suggests that higher payment induces more effort. However, we do not find that higher payment significantly affects optimal choice.

Despite the important implications of the distribution of errors in models of stochastic choice (McFadden, 1974; Yellot, 1977) relatively little is known about the distribution of noise. Similar to Duffy, Gussman, and Smith (2021) and Duffy and Smith (2022), we find that errors in our data are better described as having a Gumbel distribution rather than a normal distribution. In order to learn whether our techniques would correctly identify the stochastic distribution, we simulate observations with both normally distributed errors and errors drawn from a Gumbel distribution. When we perform our analysis on this simulated data, we correctly identify the error distribution in over 92% of these simulations. Therefore, we have a greater confidence in our conclusions about the stochastic distribution of the noise.

Consistent with the predictions of Fudenberg, Strack, and Strzalecki (2018) we find evidence that, when controlling for subject heterogeneity and decision problem heterogeneity, suboptimal choices are associated with longer response times than optimal choices. We present specifications where this relationship does not demonstrate evidence of endogeneity. However, we also find that the relationship that optimal choices are faster than suboptimal actions becomes weaker across trials.

Our observations are largely consistent with definitions of stochastic transitivity. For example, we find that 54 of our 56 triples satisfy strong stochastic transitivity. In order to better understand these observations, we conduct simulations and we find that our observations are the median and mode of the simulated data. Finally, we find a relationship between choice and attention, although we find strong evidence of an endogenous relationship.

We find interesting dynamic effects that might have been missed if we did not consider the panel nature of our experimental data.[44] For example, choice becomes less optimal and response times become faster across trials. We also find that the high payments treatment affects response times but not the optimality of choice. The implications of these results in drift diffusion models are not clear to us.

We also find that the relationship between optimal choice and response times becomes weaker across trials. Again, the implications of this on drift diffusion models are not clear to us. Perhaps the decision boundaries become narrower across trials but they collapse at a slower rate within trials? Or perhaps the rate of drift varies across trials?

Our results provide clues–not available to standard experimental designs–on what is hiding in the noise. We hope that our results are helpful in informing models of choice. We also hope that authors continue to employ incentivized designs involving imperfectly perceived stimuli to help find what is hiding in the noise in other settings.

[44]See Regenwetter and Davis-Stober (2018) for more on the theoretical implications of non-stationarity in models of choice.

References

Agranov, Marina and Ortoleva, Pietro (2017): "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 125(1), 40–68.

Ahn, David S. and Sarver, Todd (2013): "Preference for flexibility and random choice," *Econometrica*, 81(1), 341–361.

Alós-Ferrer, Carlos, Fehr, Ernst, and Netzer, Nick (2021): "Time will tell: Recovering preferences when choices are noisy," *Journal of Political Economy*, 129(6), 1828–1877.

Alós-Ferrer, Carlos and Garagnani, Michele (2021): "Choice consistency and strength of preference," *Economics Letters*, 198, 109672.

Alós-Ferrer, Carlos and Garagnani, Michele (2022a): "Strength of preference and decisions under risk," *Journal of Risk and Uncertainty*, 64, 309–329.

Alós-Ferrer, Carlos and Garagnani, Michele (2022b): "The gradual nature of economic errors," *Journal of Economic Behavior and Organization*, 200, 55–66.

Alós-Ferrer, Carlos, Granić, Đura-Georg, Kern, Johannes, and Wagner, Alexander K. (2016): "Preference reversals: Time and again," *Journal of Risk and Uncertainty*, 52(1), 65–97.

Ariely, Dan, Gneezy, Uri, Loewenstein, George, and Mazar, Nina (2009): "Large Stakes and Big Mistakes," *Review of Economic Studies*, 76(2), 451–469.

Armel, K. Carrie and Rangel, Antonio (2008): "The Impact of Computation Time and Experience on Decision Values," *American Economic Review*, 98(2), 163–168.

Ballinger, T. Parker and Wilcox, Nathaniel T. (1997): "Decisions, error and heterogeneity," *Economic Journal*, 107(443), 1090–1105.

Baranski, Joseph V. and Petrusic, William M. (1999): "Realism of confidence in sensory discrimination," *Perception and Psychophysics*, 61(7), 1369–1383.

Bhui, Rahul (2019a): "A statistical test for the optimality of deliberative time allocation," *Psychonomic Bulletin and Review*, 26(3), 855–867.

Bhui, Rahul (2019b): "Testing optimal timing in value-linked decision making," *Computational Brain and Behavior*, 2(2), 85–94.

Blavatskyy, Pavlo R. and Pogrebna, Ganna (2010): "Models of stochastic choice and decision theories: Why both are important for analyzing decisions," *Journal of Applied Econometrics*, 25(6), 963–986.

Bowers, Dawn and Heilman, Kenneth M. (1980): "Pseudoneglect: Effects of hemispace on a tactile line bisection task," *Neuropsychologia*, 18(4-5), 491–498.

Brañas-Garza, Pablo and Smith, John (2023): "Imperfect perception and stochastic choice in experiments," working paper.

Butler, David J. and Loomes, Graham C. (2007): "Imprecision as an account of the preference reversal phenomenon," *American Economic Review*, 97(1), 277–297.

Camerer, Colin F. and Hogarth, Robin 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.

Caplin, Andrew, Csaba, Dániel, Leahy, John, and Nov, Oded (2020): "Rational inattention, competitive supply, and psychometrics," *Quarterly Journal of Economics*, 135(3), 1681–1724.

Caplin, Andrew and Dean, Mark (2015): "Revealed preference, rational inattention, and costly information acquisition," *American Economic Review*, 105(7), 2183–2203.

Cattell, J. McKeen (1902): "The time of perception as a measure of differences in intensity," *Philosophische Studien*, 19, 63–68.

Cerreia-Vioglio, Simone, Dillenberger, David, Ortoleva, Pietro, and Riella, Gil (2019): "Deliberately stochastic," *American Economic Review*, 109(7), 2425–2445.

Charles, Jennifer, Sahraie, Arash, and McGeorge, Peter (2007): "Hemispatial asymmetries in judgment of stimulus size," *Perception and Psychophysics*, 69, 687–698.

Charness, Gary, and Kuhn, Peter (2011): "Lab labor: What can labor economists learn from the lab?" *Handbook of Labor Economics*, 4, 229–330.

Civelli, Andrea, Deck, Cary and Tutino, Antonella (2022): "Attention and choices with multiple states and actions: A laboratory experiment," *Journal of Economic Behavior and Organization*, 199, 86–102.

Corgnet, Brice, Hernán-González, Roberto, and Kujal, Praveen (2020): "On booms that never bust: Ambiguity in experimental asset markets with bubbles," *Journal of Economic Dynamics and Control*, 110, 103754.

Dean, Mark, and Neligh, Nathaniel (2022): "Experimental tests of rational inattention," Working paper, Columbia University.

Debreu, Gerard (1958): "Stochastic choice and cardinal utility," *Econometrica*, 26(3), 440-444.

Dewan, Ambuj and Neligh, Nathaniel (2020): "Estimating information cost functions in models of rational inattention," *Journal of Economic Theory*, 187, 105011.

Dickinson, David L. (1999): "An experimental examination of labor supply and work intensities," *Journal of Labor Economics*, 17(4), 638–670.

Duffy, Sean, Gussman, Steven, and Smith, John (2021): "Visual judgments of length in the economics laboratory: Are there brains in stochastic choice?" *Journal of Behavioral and Experimental Economics*, 93, 101708.

Duffy, Sean and Smith, John (2022): "An economist and a psychologist form a line: What can imperfect perception of length tell us about stochastic choice?" Working paper, Rutgers University-Camden.

Dutilh, Gilles and Rieskamp, Jörg (2016): "Comparing perceptual and preferential decision making," *Psychonomic Bulletin and Review*, 23(3), 723–737.

Echenique, Federico and Saito, Kota (2017): "Response time and utility," *Journal of Economic Behavior and Organization*, 139, 49–59.

Falmagne, Jean-Claude (2002): *Elements of Psychophysical Theory.* Oxford University Press: New York.

Fechner, Gustav Theodor (1860): *Elemente der Psychophysik.* (*Elements of psychophysics*, translated 1966. Holt, Rinehart, and Winston, New York.)

Frederick, Shane (2005): "Cognitive reflection and decision making," *Journal of Economic Perspectives*, 19(4), 25–42.

Fudenberg, Drew, Strack, Philipp, and Strzalecki, Tomasz (2018): "Speed, accuracy, and the optimal timing of choices," *American Economic Review*, 108(12), 3651–3684.

Gerasimou, Georgios (2021): "Simple preference intensity comparisons," *Journal of Economic Theory*, 192, 105199.

Gescheider, George A. (1997): *Psychophysics: The Fundamentals.* Routledge Press, New York.

Gneezy, Uri and Rustichini, Aldo (2000): "Pay enough or don't pay at all," *Quarterly Journal of Economics*, 115(3), 791–810.

Goryunov, Maxim and Rigos, Alexandros (2022): "Discontinuous and continuous stochastic choice and coordination in the lab," *Journal of Economic Theory*, 206, 105557.

Heng, Joseph A., Woodford, Michael, and Polania, Rafael (2020): "Efficient sampling and noisy decisions," *eLife*, 9, e54962.

Henmon, V. A. C. (1911): "The relation of the time of a judgment to its accuracy," *Psychological Review*, 18(3), 186–201.

Hey, John D. (2005): "Why we should not be silent about noise," *Experimental Economics*, 8(4), 325–345.

Hey, John D. and Orme, Chris (1994): "Investigating generalizations of expected utility theory using experimental data," *Econometrica*, 62(6), 1291–1326.

Jesteadt, Walt, Luce, R. Duncan, and Green, David M. (1977): "Sequential effects in judgments of loudness," *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 92–104.

Kamenica, Emir (2012): "Behavioral economics and psychology of incentives," *Annual Review of Economics*, 4(1), 427–452.

Kellogg, W. N. (1931): "The time of judgment in psychometric measures," *American Journal of Psychology*, 43(1), 65–86.

Kingdom, Frederick A. and Prins, Nicolaas (2016): *Psychophysics: A Practical Introduction*. San Diego: Elsevier Science and Technology.

Konovalov, Arkady and Krajbich, Ian (2019): "Revealed strength of preference: Inference from response times," *Judgment and Decision Making*, 14(4), 381–394.

Krajbich, Ian (2019): "Accounting for attention in sequential sampling models of decision making," *Current Opinion in Psychology*, 29, 6–11.

Krajbich, Ian, Armel, Carrie, and Rangel, Antonio (2010): "Visual fixations and the computation and comparison of value in simple choice," *Nature Neuroscience*, 13(10), 1292–1298.

Krajbich, Ian and Rangel, Antonio (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.

Laming, Donald and Laming, Janet (1992): "F. Hegelmaier: On memory for the length of a line," *Psychological Research*, 54(4), 233–239.

Loomes, Graham, Starmer, Chris, and Sugden, Robert (1989): "Preference reversal: information-processing effect or rational non-transitive choice?" *Economic Journal*, 99(395), 140–151.

Luce, R. Duncan (1959): *Individual choice behavior: A theoretical analysis*. Wiley: New York.

Luce, R. Duncan and Suppes, Patrick (1965): "Preference, Utility and Subjective Probability," In *Handbook of Mathematical Psychology*, (Eds.) Luce, R. Duncan, Bush, Robert R., Galanter, Eugene, Wiley, New York, 3, 249–410.

Manzini, Paola, and Mariotti, Marco (2014): "Stochastic choice and consideration sets," *Econometrica*, 82(3), 1153–1176.

Masatlioglu, Yusufcan, Nakajima, Daisuke, and Ozbay, Erkut Y. (2012): "Revealed Attention," *American Economic Review*, 102(5), 2183–2205.

Matějka, Filip and McKay, Alisdair (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105(1), 272–98.

McFadden, Daniel (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, Zarembka, Paul (Ed.), New York, Academic Press, 105–142.

McFadden, Daniel (2001): "Economic choices," *American Economic Review*, 91(3), 351–378.

Moffatt, Peter G. (2005): "Stochastic choice and the allocation of cognitive effort," *Experimental Economics*, 8(4), 369–388.

Mosteller, Frederick and Nogee, Philip (1951): "An experimental measurement of utility," *Journal of Political Economy*, 59(5), 371–404.

Münsterberg, Hugo (1894): "Studies from the Harvard Psychological Laboratory: (I)," *Psychological Review*, 1(1), 34–60.

Nicholls, Michael ER, Bradshaw, John L., and Mattingley, Jason B. (1999): "Free-viewing perceptual asymmetries for the judgement of brightness, numerosity and size," *Neuropsychologia*, 37(3), 307–314.

Nicholls, Michael ER, Mattingley, Jason B., Berberovic, Nadja, Smith, Amanda, and Bradshaw, John L. (2004): "An investigation of the relationship between free-viewing perceptual asymmetries for vertical and horizontal stimuli," *Cognitive Brain Research*, 19(3), 289–301.

Olsson, Henrik and Winman, Anders (1996): "Underconfidence in sensory discrimination: The interaction between experimental setting and response strategies," *Perception and Psychophysics*, 58(3), 374–382.

Oud, Bastiaan, Krajbich, Ian, Miller, Kevin, Cheong, Jin Hyun, Botvinick, Matthew, and Fehr, Ernst (2016): "Irrational time allocation in decision-making," *Proceedings of the Royal Society B: Biological Sciences*, 283(1822), 20151439.

Payne, John W., Bettman, James R., and Johnson, Eric J. (1993): *The Adaptive Decision Maker*. Cambridge University Press, Cambridge UK.

Payzan-LeNestour, Elise and Woodford, Michael (2022): "Outlier blindness: A neurobiological foundation for neglect of financial risk," *Journal of Financial Economics*, 143(3), 1316–1343.

Petzold, Peter and Haubensak, Gert (2004): "The influence of category membership of stimuli on sequential effects in magnitude judgment," *Perception and Psychophysics*, 66(4), 665–678.

Pirrone, Angelo, Wen, Wen, and Li, Sheng (2018): "Single-trial dynamics explain magnitude sensitive decision making," *BMC Neuroscience*, 19, 54.

Pleskac, Timothy J., Yu, Shuli, Hopwood, Christopher, and Liu, Taosheng (2019): "Mechanisms of deliberation during preferential choice: Perspectives from computational modeling and individual differences," *Decision*, 6(1), 77–107.

Pokorny, Kathrin (2008): "Pay—but do not pay too much: An experimental study on the impact of incentives," *Journal of Economic Behavior and Organization*, 66(2), 251–264.

Read, Daniel (2005): "Monetary incentives, what are they good for?" *Journal of Economic Methodology*, 12(2), 265–276.

Regenwetter, Michel, Dana, Jason, and Davis-Stober, Clintin P. (2011): "Transitivity of preferences," *Psychological Review*, 118(1), 42–56.

Regenwetter, Michel and Davis-Stober, Clintin P. (2018): "The role of independence and stationarity in probabilistic models of binary choice," *Journal of Behavioral Decision Making*, 31(1), 100–114.

Regenwetter, Michel, Davis-Stober, Clintin. P., Lim, Shiau Hong, Guo, Ying, Popova, Anna, Zwilling, Chris, Cha, Yun-Shil, Messner, William (2014): "QTest: Quantitative testing of theories of binary choice," *Decision*, 1(1), 2–34.

Reutskaja, Elena, Nagel, Rosemarie, Camerer, Colin F., and Rangel, Antonio (2011): "Search dynamics in consumer choice under time pressure: An eye-tracking study." *American Economic Review*, 101(2), 900–926.

Saito, Kota (2015): "Preferences for flexibility and randomization under uncertainty," *American Economic Review*, 105(3), 1246–1271.

Shevlin, Blair R. K., Smith, Stephanie M., Hausfeld, Jan, and Krajbich, Ian (2022): "High-value decisions are fast and accurate, inconsistent with diminishing value sensitivity," *Proceedings of the National Academy of Sciences*, 119(6), e2101508119.

Sillamaa, Mary-Anne (1999a): "How work effort responds to wage taxation: An experimental test of a zero top marginal tax rate," *Journal of Public Economics*, 73(1), 125–134.

Sillamaa, Mary-Anne (1999b): "How work effort responds to wage taxation: A non-linear versus a linear tax experiment," *Journal of Economic Behavior and Organization*, 39(2), 219–233.

Smith, Vernon L. (1976): "Experimental economics: Induced value theory," *American Economic Review*, 66(2), 274–279.

Sopher, Barry and Gigliotti, Gary (1993): "Intransitive cycles: Rational Choice or random error? An answer based on estimation of error rates with experimental data," *Theory and Decision*, 35, 311–336.

Stevens, S. S. (1957): "On the Psychophysical Law," *Psychological Review*, 64(3), 153–181.

Stevens, S. S. (1961): "To honor Fechner and repeal his law," *Science*, 133(3446), 80–86.

Suavansri, Ketchai, Falchook, Adam D., Williamson, John B., and Heilman, Kenneth M. (2012): "Right up there: Hemispatial and hand asymmetries of altitudinal pseudoneglect," *Brain and Cognition*, 79(3), 216–220.

Teghtsoonian, Robert (1971): "On the exponents in Stevens' Law and the constant in Ekman's Law," *Psychological Review*, 78(1), 71–80.

Thurstone, L. L. (1927a): "A law of comparative judgment," *Psychological Review*, 34(4), 273–286.

Thurstone, L. L. (1927b): "Psychophysical Analysis," *American Journal of Psychology*, 38(3), 368–389.

Treisman, Michel and Williams, Thomas C. (1984): "A theory of criterion setting with an application to sequential dependencies," *Psychological Review*, 91(1), 68–111.

Tsetsos, Konstantinos, Moran, Rani, Moreland, James, Chater, Nick, Usher, Marius and Summerfield, Christopher (2016): "Economic irrationality is optimal during noisy decision making," *Proceedings of the National Academy of Sciences*, 113(11), 3102–3107.

Tversky, Amos (1969): "Intransitivity of preferences," *Psychological Review*, 76(1), 31–48.

Verplanck, William S., Collier, George H., and Cotton, John W. (1952): "Nonindependence of successive responses in measurements of the visual threshold," *Journal of Experimental Psychology*, 44(4), 273–283.

Volkmann, John (1934): "The relation of the time of judgment to the certainty of judgment," *Psychological Bulletin*, 31(9), 672–673.

Weber, Ernst (1834): *De Tactu.* (*The Sense of Touch*, translated 1978. Academic Press, New York.)

Willemsen, Martijn C. and Johnson Eric J. (2019): "(Re) visiting the decision factory: Observing cognition with MouselabWEB," in *A handbook of process tracing methods*, Schulte-Mecklenbeck, Michael, Kuehberger, Anton, and Johnson, Joseph G. (Eds.), Routledge, New York, 76–95.

Woodford, Michael (2014): "Stochastic choice: An optimizing neuroeconomic model," *American Economic Review*, 104(5), 495–500.

Yellott, John I. (1977): "The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution," *Journal of Mathematical Psychology*, 15(2), 109–144.

Zeigenfuse, Matthew D., Pleskac, Timothy J. and Liu, Taosheng (2014): "Rapid decisions from experience," *Cognition*, 131(2), 181–194.

Appendix-For Online Publication

Robustness of results: specifications

In this subsection, we examine the robustness of our results by offering different specifications than those in the main body. First, we study the robustness of the results in Table 1, by employing the *Longer minus selected* variable, which is the length of the longer line in that trial minus the length of the selected line. This variable is bounded below at 0, so we perform tobit regressions. The regressions, which are summarized in Table $A1$ are otherwise identical to those summarized in Table 1.

Table $A1$: Tobit regressions of the Longer minus selected variable

	(1)	(2)	(3)	(4)
Longer length	0.0233***	0.0233***	0.0226***	0.0227***
	(0.0026)	(0.0026)	(0.0025)	(0.0025)
Absolute difference	−0.362***	−0.362***	−0.349***	−0.349***
	(0.015)	(0.015)	(0.015)	(0.015)
High payment	0.012	0.490	−0.032	0.331
	(0.191)	(0.403)	(0.189)	(0.389)
Trial	0.0096***	0.0135***	0.0091***	0.0120**
	(0.0028)	(0.0040)	(0.0027)	(0.0093)
High payment*trial	-	−0.0078	-	−0.0059
		(0.0058)		(0.0056)
Subject fixed effects	Yes	Yes	Yes	Yes
Subject left dummies	No	No	Yes	Yes
Demographics	No	No	No	No
AIC	33643	33643	32497	32498

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts, the subject-specific dummy estimates, or demographics estimates. AIC refers to the Akaike information criterion (Akaike, 1974). Each regression has $14,280$ observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Keeping in mind the difference in the signs of the dependent variables, the results are qualitatively similar to those in Table 1: the optimality of choices are decreasing in the length of the longer line, increasing the absolute difference between the lines, and decreasing across trials.

Next, we study the robustness of the specification of response times. In Table 3, the dependent variable was the log of the response times. Below we examine the results of otherwise identical regressions, with the raw response times. These regressions are summarized in Table $A2$.

Table $A2$: Regressions of the Response times variable

	(1)	(2)	(3)	(4)
Longer length	0.0065***	0.0065***	0.0062***	0.0062***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Absolute difference	−0.129***	−0.129***	−0.129***	−0.129***
	(0.006)	(0.006)	(0.006)	(0.006)
High payment	0.0590	0.1723	0.0402	0.1608
	(0.1135)	(0.2357)	(0.1134)	(0.2355)
Trial	−0.0644***	−0.0634***	−0.0644***	−0.0634***
	(0.0016)	(0.0024)	(0.0016)	(0.0024)
High payment*Trial	-	−0.0019	-	−0.0020
		(0.0034)		(0.0034)
Subject fixed effects	Yes	Yes	Yes	Yes
Subject left dummies	No	No	Yes	Yes
Demographics	No	No	No	No
AIC	94946.4	94955.6	94397.5	94406.7

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts, subject-specific dummy estimates, or the demographics estimates. AIC refers to the Akaike information criterion. Each regression has $14,280$ observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

These results are similar to those in Table 3, with the exception of the effects of the high payments treatment. Here we do not find evidence that subjects in the high payments treatment had longer response times and we do not find a significant interaction between the High payments dummy and Trial variable.

Similarly, we examine the robustness of the results in Table 4, where the dependent variable was the log of the response times. In Table $A3$, we summarize the identical analysis where (raw) response time is the dependent variable.

Table $A3$: Regressions of the Response time variable

	(1)	(2)	(3)	(4)
Longer length	0.0060***	0.0060***	0.0058***	0.0058***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Absolute difference	−0.114***	−0.114***	−0.118***	−0.117***
	(0.006)	(0.006)	(0.006)	(0.006)
High payment	0.0555	0.0466	0.0397	0.0306
	(0.1132)	(0.1131)	(0.1133)	(0.1132)
Trial	−0.0647***	−0.0796***	−0.0646***	−0.0793***
	(0.0016)	(0.0033)	(0.0016)	(0.0033)
Selected longer	−1.092***	−2.309***	−0.844***	−2.037***
	(0.139)	(0.272)	(0.146)	(0.276)
Trial*Selected longer	-	0.0198***	-	0.0194***
		(0.0038)		(0.0038)
Subject fixed effects	Yes	Yes	Yes	Yes
Subject left dummies	No	No	Yes	Yes
Demographics	No	No	No	No
AIC	94886.5	94868.9	94365.9	94349.3

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts, the subject-specific dummy estimates, or the demographics estimates. AIC refers to the Akaike information criterion. Each regression has $14,280$ observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Again, these results are similar to those in Table 4, with the exception that the High payments coefficient is not significant. One advantage of the analysis summarized in Table $A3$ is that the Selected longer coefficients are easier to interpret because these are the estimated differences seconds between optimal and non-optimal trials. We also note that this difference appears to decrease across trials at a rate of just under 20 milliseconds.

Likewise for the analysis in Table 4, here we conduct a check on possible endogeneity on our analysis. We conduct Spearman correlations between the unstandardized residuals and the Selected longer variable, and we find p-values less than 0.001 for specifications $(1)-(4)$. Additionally, we obtain p-values less than 0.001 for specifications $(1)-(4)$ when we use the student residuals rather than the unstandardized residuals. This suggests to us that the lack of endogeneity claims following Table 4, are not robust to the specification of response times.

Robustness of results: demographics

Here we examine the robustness of our results by including demographic variables rather than estimating the fixed-effects.

The *Her pronoun* variable is 1 if the subject indicated she preferred the "her" pronoun, and a 0 otherwise. The CRT score variable is the sum of the correct responses on the Cognitive reflection test (Frederick, 2005). Every subject (119) offered a valid response to these items. The Left handed variable is 1 if the subject indicated they they are left handed, and a 0 otherwise. One subject did not offer a valid response, therefore we have observations for 118 subjects. Finally, we have 99 valid responses for an (optional) GPA estimate. Table $A4$ reports the Spearman correlation coefficients among the variables.

Table $A4$: Spearman non-parametric correlation coefficients

	1	2	3	4	5	6
1 Selected longer	1.00					
2 View clicks	-0.1021^{***}	1.00				
3 Response time	-0.0962^{***}	0.65450^{***}	1.00			
4 Her pronoun	-0.0014	-0.0911^{***}	0.0454^{***}	1.00		
5 CRT score	0.0004	0.1172^{***}	-0.0012	-0.0517	1.00	
6 Left handed	-0.0122	0.0491^{***}	0.0423^{***}	0.0418	0.0678	1.00
7 GPA (optional)	$0.0167^{\dagger}$	0.1307^{***}	0.0406^{***}	$0.1810^{\dagger}$	0.1463	0.0482

Each correlation between GPA (7) and (1) – (3) has 11,880 trial-level observations. Each correlation between GPA (7) and (4) – (6) has 99 observations. Each correlation between Left handed (6) and (1) – (3) has 14,160 trial-level observations. Each correlation between Left handed (6) and (4) – (5) has 118 observations. The correlation between Her pronoun (4) and CRT score (5) has 119 observations. The remaining correlations have 14,280 trial-level observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

We use these variables to construct analogous analyses as in the body of the paper. Below, we conduct an analysis similar to that in Table 1, but we employ the demographics variables rather than estimating fixed effects. This analysis is summarized in Table $A5$.

Table $A5$: Logistic regressions of the Selected longer line variable

	(1)	(2)	(3)	(4)
Longer length	-0.00063^{***}	-0.00063^{***}	-0.00057^{***}	-0.00057^{***}
	(0.0001)	(0.0001)	(0.00010)	(0.00010)
Absolute difference	0.0235^{***}	0.0235^{***}	0.0240^{***}	0.0240^{***}
	(0.0006)	(0.0006)	(0.0007)	(0.0007)
High payment	-0.0037	-0.0310^{*}	-0.0040	$-0.0261^{\dagger}$
	(0.0069)	(0.0140)	(0.0075)	(0.0152)
Trial	-0.00030^{**}	-0.00052^{***}	-0.00027^{*}	-0.00045^{**}
	(0.0001)	(0.00014)	(0.00011)	(0.00015)
High payment*Trial	-	0.00044^{*}	-	$0.00036^{\dagger}$
		(0.00020)		(0.00022)
Her pronoun	0.0009	0.0010	-0.0094	-0.0096
	(0.0076)	(0.0076)	(0.0086)	(0.0086)
CRT score	-0.0020	-0.0023	-0.0044	-0.0046
	(0.0037)	(0.0037)	(0.0039)	(0.0039)
Left handed	-0.0190	-0.0197	-0.0215	-0.0217
	(0.0127)	(0.0127)	(0.0135)	(0.0135)
GPA	-	-	0.0233^{*}	0.0234^{*}
			(0.0094)	(0.0094)
Observations	14,160	14,160	11,880	11,880
AIC	14030.04	14027.03	11690.00	11689.20

We provide the average marginal effects and the standard errors in parentheses. We do not provide the estimates of the intercepts. None of the specifications have subject-specific dummy estimates. AIC refers to the Akaike information criterion (Akaike, 1974). *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

These results are largely consistent with those in Table 1. However, we note a positive relationship between reported GPA and selecting the longer line.

We likewise construct the analogous version of Table 2, but with demographics variables and not fixed effects. This analysis is summarized in Table $A6$.

Table $A6$: Regressions of the View clicks variable

	(1)	(2)	(3)	(4)
Longer length	0.0026*	0.0026*	0.0030*	0.0030*
	(0.0012)	(0.0012)	(0.0014)	(0.0014)
Absolute difference	−0.1107***	−0.1107***	−0.1193***	−0.1193***
	(0.0051)	(0.0051)	(0.0058)	(0.0058)
High payment	0.1567	0.2711	0.1139	0.0756
	(0.1030)	(0.2074)	(0.1171)	(0.2358)
Trial	−0.0164***	−0.0154***	−0.0166***	−0.0170***
	(0.0015)	(0.0021)	(0.0017)	(0.0024)
High payment*Trial	-	−0.0019	-	0.0006
		(0.0030)		(0.0034)
Her pronoun	−1.805***	−1.806***	−2.4934***	−2.4937***
	(0.114)	(0.1139)	(0.1342)	(0.1342)
CRT score	0.7162***	0.7171***	0.4820***	0.4818***
	(0.0558)	(0.0558)	(0.0617)	(0.0617)
Left handed	0.5052**	0.5081**	0.7604***	0.7600***
	(0.1944)	(0.1944)	(0.2168)	(0.2168)
GPA	-	-	1.7462***	1.7462***
			(0.1486)	(0.1486)
Observations	14,160	14,160	11,880	11,880
AIC	91518.8	91528.2	77772.5	77782.0

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts. None of the specifications have subject-specific dummy estimates. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

These results are largely consistent with those in Table 2, however we note that High payment coefficient is not significant. We also note that subjects reporting being left handed select more view clicks than otherwise. Further, we note that subjects preferring the her pronoun tend to select fewer View clicks. Finally, we note that View clicks are increasing in both the CRT score and the reported GPA.

Likewise we offer a version of Table 3 but with the demographics variables rather than estimating fixed effects. We summarize this in Table $A7$.

Table $A7$: Regressions of the log of Response times variable

	(1)	(2)	(3)	(4)
Longer length	0.00035***	0.00035***	0.00037***	0.00037***
	(0.00005)	(0.00005)	(0.00006)	(0.00006)
Absolute difference	−0.0061***	−0.0061***	−0.0065***	−0.0065***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
High payment	0.0089*	0.0257**	0.0096*	0.0143
	(0.0043)	(0.0087)	(0.0047)	(0.0094)
Trial	−0.0029***	−0.0027***	−0.0030***	−0.0029***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
High payment*Trial	-	−0.00028*	-	−0.00008
		(0.00013)		(0.00014)
Her pronoun	0.0315***	0.0314***	$-0.0099^{\dagger}$	$-0.0098^{\dagger}$
	(0.0048)	(0.0048)	(0.0054)	(0.0054)
CRT score	−0.00098	−0.00085	−0.0070**	−0.0070**
	(0.00235)	(0.00235)	(0.0025)	(0.0025)
Left handed	0.0293***	0.0297***	0.0631***	0.0632***
	(0.0082)	(0.0082)	(0.0087)	(0.0087)
GPA	-	-	0.0268***	0.0268***
			(0.0059)	(0.0059)
Observations	14,160	14,160	11,880	11,880
AIC	1845.7	1856.9	1355.3	1370.9

We provide the coefficient estimates and the standard errors in parentheses. We do not provide the estimates of the intercepts. None of the specifications have subject-specific dummy estimates. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Again, our results are largely consistent with those summarized in Table 3. Some of the demographics results are analogous to those in Table $A6$. We find a positive relationship between reported GPA and response times. We also find that subjects who report being left handed have longer response times. However, we find some evidence of a negative relationship between CRT score and response time. We also find some evidence subjects who prefer the her pronoun have longer response times.

Computer details and analysis of offset

Computer details

The sessions were performed on standard 21.5 inch (54.6 cm) Dell EliteDisplay E221 monitors. E-Prime imposed a resolution of 1024 pixels by 768 pixels. Each computer had a double click speed setting of 6 out of 11. Each computer had a pointer speed setting of 6 out of 11.

Unincentivized practice details

The subjects had an unincentivized practice on the line selection task. If the subjects did not view both lines or did not select a line, the subjects were informed of this and were directed to repeat the practice line selection task. In this practice trial, the line on the left was 189 pixels and the line of the right was 224 pixels.

Offset details

Both lines appeared in rectangular regions on the screen. These rectangles were 400 pixels by 150 pixels. The boundaries of these rectangles were not visible to subject. There was a rectangular region on the left of the screen. The nearest edge of the screen was 56 pixels from both the left and right rectangles. There were 112 pixels separating the right and left boxes.

The lines were offset, both vertically and horizontally, within these rectangles. There was a minimum cushion of 10 pixels between the boundary of the rectangle and the line. Regardless of the number of line views, the offsetting was fixed for both lines throughout each trial.

Offset analysis: presentation effects

Although these offsets were randomized across trials, it is possible that they affect the optimal choice. One possibility is that lines that are closer to the nearest (left or right) screen edge are judged to be longer than lines that are farther from the nearest (left or right) screen edge. We define the *Edge distance shorter minus longer* variable to be the distance of the nearest edge of the shorter line to the screen edge minus the distance of the nearest edge of the longer line to the screen edge. This variable is increasing in the distance of the shorter line to the screen edge relative to the distance of the longer line to the screen edge, and could possibly provide a nudge to select the longer line.

Another possibility is that lines with a higher vertical position are judged to be longer than lines with a lower vertical position. We define the *Top distance shorter minus longer* variable to be the distance of the shorter line from the top of the screen minus the distance of the longer line from the top of the screen. This variable is increasing in the relative height of the longer line relative to the shorter line, and could possibly provide a nudge to select the longer line.

Below, we analyze the effects of these trial-specific offsets on choice. In order to account for as much subject-specific heterogeneity as possible, we run specification (3) of Table 1, with these offset variables. We summarize this analysis in Table *A*8.

Table *A*8: Logistic regressions of the Selected longer line variable

	(1)	(2)	(3)
Longer length	-0.00063^{***}	-0.00063^{***}	-0.00063^{***}
	(0.00009)	(0.00009)	(0.00009)
Absolute difference	0.0228^{***}	0.0229^{***}	0.0228^{***}
	(0.00058)	(0.00058)	(0.00058)
High payment	-0.0019	-0.0015	-0.0020
	(0.0063)	(0.0063)	(0.0063)
Trial	-0.00030^{**}	-0.00030^{**}	-0.00030^{**}
	(0.00009)	(0.00009)	(0.00009)
Edge distance shorter minus longer	0.00019^{***}	-	0.00018^{***}
	(0.00005)		(0.00005)
Top distance shorter minus longer	-	0.00022^{***}	0.00021^{***}
		(0.00006)	(0.00006)
Subject fixed effects	Yes	Yes	Yes
Subject left dummies	Yes	Yes	Yes
Demographics	No	No	No
AIC	12660.65	12661.09	12651.15

We provide the average marginal effects and the standard errors in parentheses. We do not provide the estimates of the intercepts or the subject-specific dummy estimates. AIC refers to the Akaike information criterion (Akaike, 1974). Each regression has $14,280$ observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

There appears to be a bias toward selecting lines that are closer to the edge of the screen.

There also appears to be a bias toward selecting lines with a higher vertical presentation. These results suggest that optimal choice is affected by its spatial presentation.[45]

Sequential effects

To examine possible sequential effects, we conduct an analysis, similar to that summarized in Table 1, however, we include two variables to characterize the previous trial.

One variable indicates the exogenous information that the longer line in the previous trial was on the same side of the screen as the longer line in the current trial. The *Previous longer and current longer same side* variable is 1 if the longer line in the previous trial and the longer line in the current trial are on the same side (left or right) of the screen, and 0 otherwise. The other variable indicates the endogenous information that the subject had a possibly temporary bias toward one side or the other. The *Previous choice and current longer same side* variable is 1 if the choice in the previous trial and the longer line in the current trial were on the same side of the screen, and 0 otherwise.

Each regression has $13,566$ because we exclude the first trial within each of the 6 blocks for each of the 119 subjects. We include the subject-specific estimates of a possible bias toward the left or right side of the screen. In this way, we are better able to estimate any local variations in the behavior of the subjects. These regressions are summarized in Table $A9$.

[45] We also note that it is possible that larger differences in the vertical or horizontal positions of the lines might make the lengths more difficult to compare. According to this view, the absolute values of the horizontal positions of the lines or the absolute values of the vertical positions of the lines might reduce the likelihood of optimal choice. However, we do not find evidence for either conjecture.

Table $A9$: Logistic regressions of the Selected longer line variable

	(1)	(2)	(3)
Longer length	-0.00066^{***}	-0.00066^{***}	-0.00066^{***}
	(0.00009)	(0.00009)	(0.00009)
Absolute difference	0.0228^{***}	0.0229^{***}	0.0228^{***}
	(0.0006)	(0.0006)	(0.0006)
High payment	-0.0009	-0.0009	-0.0009
	(0.0065)	(0.0065)	(0.0065)
Trial	-0.00026^{**}	-0.00027^{**}	-0.00026^{**}
	(0.00009)	(0.00009)	(0.00009)
Previous longer and current longer same side	-0.0067	-	-0.0238^{**}
	(0.0065)		(0.0075)
Previous choice and current longer same side	-	0.0226^{***}	0.0349^{***}
		(0.0067)	(0.0077)
Subject fixed effects	Yes	Yes	Yes
Subject left dummies	Yes	Yes	Yes
Demographics	No	No	No
AIC	12071.1	12060.7	12052.7

We provide the average marginal effects and the standard errors in parentheses. We do not provide the estimates of the intercepts or the subject-specific dummy estimates. AIC refers to the Akaike information criterion (Akaike, 1974). Each regression has 13,566 observations. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Even when controlling for an overall subject-specific bias for one side of the screen, we find evidence of temporary bias for one side of the screen. Specifically the positive and significant estimates of *Previous choice and current longer same side* suggest that subjects can have endogenous and temporary biases favoring one side over the other. Psychologists might interpret this as evidence of a moving, temporary threshold for the decision to select right or left. We also find some evidence that optimal choice is less likely when the longer line is on the same side as the previous trial than when they are on different sides of the screen.

Line pairs summary data

Subjects are presented with 60 unique line pairs. 24 pairs of lines had longest lines of 160, 240, and 320 pixels. The shorter lines were constructed by subtracting 1, 3, 5, 10, 15, 20, 30, 40 pixels from the longer line. 36 pairs of lines had a longest line ranging from 230 to 250 where the shorter lines were constructed by subtracting 2 through 20 pixels. The specific line pairs are listed below in pixels. The *Optimal* columns describe the percent of optimal choices and are based on 238 observations (119 subjects in both high and low payment treatments). This is summarized in Table $A10$.

Table $A10$

Longer	Shorter	ID	Optimal
160	159	01	55.46%
160	157	02	66.39%
160	155	03	76.05%
160	150	04	89.50%
160	145	05	95.80%
160	140	06	98.32%
160	130	07	99.58%
160	120	08	100%
240	239	09	50.84%
240	237	10	63.87%
240	235	11	69.33%
240	230	12	79.41%
240	225	13	91.60%
240	220	14	94.96%
240	210	15	99.16%

Longer	Shorter	ID	Optimal
240	200	16	99.58%
320	319	17	49.16%
320	317	18	55.04%
320	315	19	69.32%
320	310	20	78.15%
320	305	21	84.03%
320	300	22	90.76%
320	290	23	95.80%
320	280	24	97.06%
232	230	25	60.50%
234	232	26	52.52%
234	230	27	62.61%
236	234	28	60.08%
236	232	29	71.01%
238	236	30	55.46%

Table $A10$ continued

Longer	Shorter	ID	Optimal	Longer	Shorter	ID	Optimal
238	234	31	60.08%	246	236	46	81.09%
238	232	32	68.49%	246	234	47	86.13%
240	238	33	57.98%	246	232	48	89.08%
240	236	34	66.81%	248	246	49	58.82%
242	240	35	63.03%	248	244	50	63.03%
242	238	36	64.71%	250	248	51	52.94%
242	236	37	73.11%	250	246	52	70.59%
242	234	38	68.91%	250	244	53	70.59%
242	232	39	81.51%	250	242	54	76.47%
244	242	40	60.08%	250	240	55	84.45%
244	240	41	65.55%	250	238	56	86.55%
246	244	42	57.98%	250	236	57	88.66%
246	242	43	65.13%	250	234	58	89.50%
246	240	44	71.01%	250	232	59	92.86%
246	238	45	74.79%	250	230	60	95.80%

Stochastic transitivity simulations summary data

Table 7 summarized the distribution of the SST violations among our 100 simulations, where the noise is modeled as a Gumbel distribution with a 0 mean and standard deviation of 8. In Table $A11$, we present the number of SST violations within each of the 56 triples.

Table $A11$

ID	a	b	c	Sim. SST viol.	ID	a	b	c	Sim. SST viol.
1	250	232	230	19 out of 100	26	250	248	244	7 out of 100
2	250	234	230	3 out of 100	27	250	248	246	6 out of 100
3	250	240	230	0 out of 100	28	248	246	244	6 out of 100
4	250	234	232	13 out of 100	29	246	234	232	15 out of 100
5	250	236	232	3 out of 100	30	246	236	232	0 out of 100
6	250	238	232	0 out of 100	31	246	238	232	0 out of 100
7	250	242	232	0 out of 100	32	246	242	232	0 out of 100
8	250	246	232	1 out of 100	33	246	236	234	9 out of 100
9	250	236	234	16 out of 100	34	246	238	234	0 out of 100
10	250	238	234	1 out of 100	35	246	242	234	0 out of 100
11	250	242	234	0 out of 100	36	246	238	236	5 out of 100
12	250	246	234	1 out of 100	37	246	242	236	0 out of 100
13	250	238	236	10 out of 100	38	246	242	238	0 out of 100
14	250	240	236	2 out of 100	39	246	242	240	2 out of 100
15	250	242	236	0 out of 100	40	246	244	240	6 out of 100
16	250	246	236	1 out of 100	41	246	244	242	10 out of 100
17	250	240	238	9 out of 100	42	244	242	240	8 out of 100
18	250	242	238	0 out of 100	43	242	234	232	4 out of 100
19	250	246	238	1 out of 100	44	242	236	232	0 out of 100
20	250	242	240	8 out of 100	45	242	238	232	0 out of 100
21	250	244	240	0 out of 100	46	242	236	234	3 out of 100
22	250	246	240	0 out of 100	47	242	238	234	0 out of 100
23	250	244	242	2 out of 100	48	242	238	236	10 out of 100
24	250	246	242	0 out of 100	49	242	240	236	8 out of 100
25	250	246	244	6 out of 100	50	242	240	238	2 out of 100

Table $A11$ continued

ID	a	b	c	Sim. SST viol.
51	240	238	236	7 out of 100
52	238	234	232	3 out of 100
53	238	236	232	4 out of 100
54	238	236	234	11 out of 100
55	236	234	232	4 out of 100
56	234	232	230	11 out of 100