Three Essays in Behavioural and Experimental Economics

by

Hanh Thi Tong

M.Sc., Aarhus University, 2013
B.Sc., Maastricht University, 2011

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Summer 2020

Approval

Name:	**Hanh Thi Tong**
Degree:	**Doctor of Philosophy (Economics)**
Title:	**Three Essays in Behavioural and Experimental Economics**
Examining Committee:	**Chair:** Alexander Karaivanov Professor
	David Freeman Senior Supervisor Associate Professor
	Luba Petersen Supervisor Associate Professor
	Shih En Lu Internal Examiner Associate Professor
	Sabine Kröger External Examiner Professor Department of Economics Laval University
Date Defended:	**July 30, 2020**

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

My thesis includes three chapters in Behavioural and Experimental Economics.

The first chapter – "Instructions", is co-authored with David J. Freeman, Erik O. Kimbrough and Garrett M. Petersen, and is published in the Journal of the Economic Science Association. In this paper, we experimentally compare how methods of delivering and reinforcing experiment instructions impact subjects' comprehension and retention of payoff-relevant information. We report a one-shot individual decision task in which non money-maximizing behavior can be unambiguously identified and find that such behavior is prevalent in our baseline treatment which uses plain, but relatively standard experimental instructions. We find combinations of reinforcement methods that can eliminate half of non money-maximizing behavior, and we find that we can induce a similar reduction via enhancements to the content of instructions. Residual non money-maximizing behavior suggests this may be an important source of noise in experimental studies.

The second chapter – "Anchors of Strategic Reasoning in the Traveler's Dilemma", is co-authored with David J. Freeman. In this paper, we experimentally study players' initial beliefs about non-strategic play that anchors their strategic reasoning in the traveler's dilemma, a game in which each player chooses a number and has the incentive to undercut their opponent by the minimal amount possible. In a within-subject design, each subject repeatedly plays variations of the traveler's dilemma game without feedback. To identify their strategic reasoning, we vary the upper and lower bounds of the strategy space in each round, and also vary the reward/penalty parameter for undercutting one's opponent. We find that players are both heterogeneous in the amount that they reason, and in their beliefs about non-strategic play. Notably, few players anchor their strategic reasoning on a non-strategic uniform random play. We also find ample evidence of non-strategic play being prevalent. Our results caution against the common practice of assuming the same anchor of initial reasoning for all players when estimating players' depths of strategic reasoning.

The third chapter – "Default-Setting and Default Bias: Does the Choice Architect Matter?", is co-authored with David J. Freeman and Lanny Zrill. The presence of pre-selected default options has been shown to influence individual decision making in various contexts including the choice of health insurance and retirement contributions. Even so, it is not well understood how individuals with heterogeneous preferences react to the procedure used to select the default options. We develop an econometric approach to test and compare default bias across default-setting rules that controls for heterogeneous individual preferences. In a within-subject experimental design studying lottery choices, we apply our approach to compare four different default-setting rules: Random defaults, Custom defaults selected based on an individual's own past choices, Social defaults selected based on others' choices, and Expert-set defaults. We find that the content of default-setting rules matters: we find default bias in all non-random default-setting regimes but not with Random defaults. Our subjects also tended to rank non-random default-setting regimes over choosing with No Default and Random defaults.

Keywords: Attention, Comprehension, Instructions, Non-strategic Play, Depth of Reasoning, Traveler's Dilemma, Default Bias, Choice Architecture

Dedication

To my loving family – Manh Tam Duong Diep Tom Kat
To my wonderful partner – Peter Stair

Acknowledgements

I am grateful for many people who have supported me through out the journey toward my PhD degree.

First and foremost, I would like to thank my senior supervisor, David Freeman, for his guidance and support, his generosity with his time, and his willingness to supervise me – his first PhD student – while working very hard under the tenure clock.

I am thankful to my supervisor Luba Petersen for her detailed feedback and her cheerful attitude, which I would say is a rare gem in academia. I'd also like to express my gratitude to faculty members - Erik Kimbrough, Arthur Robson, Shih En Lu, Simon Woodcock and Alex Karaivanov. I thank Gwen Wild, Kathleen Vieira-Ribeiro, Lisa Agosti, Christine Harper, Mike Perry, Tim Coram, Rebecca Ho and Azam Bhatti for their excellent works to keep things run smoothly.

I have been fortunate to have the company of excellent peers and friends who gave me a lot of advice and encouragement. Thank you – Veridiana de Andrade Nogueira, Wanzhu Zhang, Xiaowen Lei, Yaser Sattari, Meiyu Li, Eric Adebayo, Garrett Petersen, Ricardo Meilman Lomaz Cohn, Thomas Vigie, Kevin Laughren, Farouk Abdul-Salam, Rogayeh Tabrizi, Kevin Chen, Reza Sattari, Aidin Hajikhameneh, Edouard Djeutem, Alia M. and many more.

The poem "Try Try Again" captures much of the experience of this long intellectually challenging journey. Thank Lucy for sharing it.

Table of Contents

List of Tables

List of Figures

Chapter 1

Instructions

1.1 Introduction

Experiments start by providing instructions designed to ensure that subjects understand how their actions and others' actions determine payoffs. Such understanding is crucial to the economic interpretation of subjects' behavior – without it, the experimenter has lost control (Smith 1982). Almost from the field's inception, experimental economists have recognized that the effectiveness of instructions in establishing understanding may depend on how they are delivered and reinforced (Fouraker and Siegel 1963). Prominent textbooks give detailed guidelines on how to deliver instructions and suggest complementary methods to increase subjects' comprehension, including reading instructions aloud and using demonstrations, quizzes, and practice rounds (Friedman and Sunder 1994, Davis and Holt 1993, Cassar and Friedman 2004). Casual observation suggests wide variation in how practitioners deliver instructions and use reinforcement methods. We review the methods for delivering and reinforcing instructions as reported in experimental studies recently published in six leading journals and confirm this observation. We find that almost all experimenters complement their instructions with at least one reinforcement method, though the methods used vary substantially. This suggests that experimental economics lacks clear norms for how instructions ought to be delivered and reinforced. Troublingly, we were unable to classify roughly 22% of papers because they failed to provide sufficient details on their methods.

Despite observed variation in practices, there is scant evidence comparing their effectiveness. Thus we conduct an experiment to evaluate the impact of methods of delivering instructions and reinforcing their content on behavior. We study a one-shot timing decision in which each subject is performing a default Task 1 for money and must decide when (or whether) to switch over and complete Task 2. Task 2 can be performed at most once, and the subject is paid the most for doing it at the correct time and least for doing it earlier. Moreover, the subject is better off not doing Task 2 at all than doing it too early. This information is explicitly stated in the instructions. Doing the task too early – *non money-maximizing behavior* (NMB) – could reflect idiosyncratic preferences, or result from a failure to comprehend or retain information from the instructions. Variation in NMB across treatments, which hold the distribution of preferences constant in expectation, thus reflects variation in comprehension and retention. For most treatments, we hold constant the content of instructions and vary how instructions are delivered and reinforced. We include one additional treatment with enhanced instructions as a robustness check.

In our first treatment subjects complete self-paced computerized instructions including practice rounds and then take a comprehension quiz before beginning the study (providing us an alternative measure of their comprehension upon completion of the instructions). Nearly half of subjects in this treatment do the task too early, exhibiting NMB. A second treatment provides subjects with the quiz answers, and this generates a moderate, but statistically insignificant reduction in NMB. We thus study the additional impact of introducing monetary incentives for quiz performance, of going through the computerized instructions twice (both before and after the quiz), and of providing paper instructions alongside computerized instructions. We find that all three of these treatments lead to significant improvements relative to the baseline – but each only eliminates about half of the observed NMB, as does our treatment with enhanced instructions.

By studying an individual decision task, our experiment eliminates strategic and other-regarding motives that might confound the identification or interpretation of NMB. By studying a one-shot decision without feedback, we obtain a clean measure of understanding and retention of the instructions that is not confounded by learning. We are aware of two existing papers that have studied the impact of instruction delivery and reinforcement on play in repeated public goods games (Bigoni and Dragone 2012, Ramalingam et al. 2018).[1] The more relevant of these is Bigoni and Dragone (2012), who find that shortened on-screen instructions led to lower quiz scores and longer response times as compared to their baseline paper instructions, shortened paper instructions, and shortened on-screen instructions with active examples requiring subject input. However, they find no effect of instructions on observed behavior.

1.2 Literature Survey

We report how instructions are delivered and reinforced in 260 experimental studies published between January 2011 and December 2016 in Experimental Economics and five prominent general interest economics journals. We selected all papers in these journals that contained at least one lab experiment in which participants were given instructions on the

[1] Our discussion here is restricted to instruction delivery and reinforcement. We have little to say about how variation in the content of the instructions may affect behavior, by providing or failing to provide subjects with payoff-relevant information, or alternatively by influencing the framing of the experimental task. See Alekseev et al. (2017) for a discussion of the use of context in instructions. See also Converse and Presser (1986) for a discussion of effective survey design which offers potentially useful guidance for economists.

experimental procedure. For each paper, we checked whether instructions were delivered on paper, on screen, both, or neither. We also recorded the use of various practices intended to reinforce the content of the instructions, including reading the instructions aloud, demonstrations, practice rounds, and pre-experiment quizzes. Since ensuring subjects' initial comprehension may be particularly important when experiments are one-shot or provide limited feedback, we further classified the nature of each experiment based on whether or not a main task was one-shot, and whether or not subjects received feedback. This allows us to assess whether experimenters adapt their instruction protocols to the nature of the task being studied. Details of our classification procedure are given in Appendix A.1. The results of our survey are given in Table 1.1.

We were unable to determine how instructions were delivered in 22% of the studies we reviewed. If behavior is sensitive to how instructions are delivered, this oversight hampers replication. Of the remaining 204 studies, 61% deliver instructions exclusively on paper, 24% deliver instructions exclusively on screen, while another 5% use both. We find this noteworthy since the majority of these experiments are themselves computerized. The remaining 10% of these 204 studies use neither paper nor computer instructions. Most such studies are lab-in-the-field experiments studying non-student populations and deliver instructions orally along with some of the reinforcement methods discussed below. We suspect that experimental economists' revealed preference for paper instructions is driven by the fact that subjects can refer back to them throughout the experiment, which may not always be the case with computer instructions. This may mitigate subjects' tendency to forget important information.[2]

85% of all studies use at least one method of reinforcement which suggests that experimenters are almost universally concerned about subject comprehension and retention. Instructions are read aloud in 54% of studies.

We find that 57% of studies use demonstrations or practice rounds to reinforce subject understanding of the experiment. Examples of such practices include physical demonstrations of how risk will be resolved,[3] guided examples of possible actions and their consequent outcomes, and unpaid practice rounds. Of the studies that use at least one of these forms of reinforcement, 80% use guided demonstrations or guided practice rounds, and 42% use unguided practice rounds; some studies use both.

In addition to reinforcing the content of instructions, experiments can also test subjects' comprehension thereof with pre-experiment quizzes (39% of studies). At least 63% of these reinforced understandings and corrected misunderstandings by providing answers to the

[2]Reading instructions aloud and/or publicly distributing paper instructions may also help establish common information in strategic settings (Friedman and Sunder 1994, p. 77).

[3]Davis and Holt (1993, p. 23) and Friedman and Sunder (1994, p. 67) suggest that the use of physical randomization devices may enhance credibility.

Table 1.1: Instruction delivery and reinforcement in economics experiments

	Delivery method					
	Computer only	Paper only	Computer and Paper	Neither	Unclear	Total
Total	**48**	**124**	**11**	**21**	**56**	**260**
Reinforcement: Read aloud	19	79	4	21	17	140
Practice/Demonstration	30	63	10	15	29	147
Demo or guided practice	21	56	8	13	19	117
Unguided practice	16	22	4	4	16	62
Quiz	16	54	8	6	17	101
Feedback	10	35	5	4	10	64
Incentive	0	3	0	0	0	3
Require 100%	5	23	3	3	7	41
Feedback unclear	5	18	3	2	7	35
One-shot	15	43	4	12	10	84
Feedback between decisions	24	73	7	6	42	152

Each entry is the number of papers classified in that respective category. Indented categories are subsets of the preceding non-indented category.

quiz, and 41% required a perfect score to commence the experiment. Only three of the studies paid subjects for quiz performance. We note that 35% of studies that used a quiz did not clearly report whether or how subjects were given feedback on the quiz.

Given our prior that reinforcement may be especially important when feedback is limited, we find it surprising that one-shot experiments less frequently incorporate practice or demonstrations ($\rho = -.19$, $p < .01$, $n = 260$) and quizzes ($\rho = -.15$, $p = .02$, $n = 260$) in their instructions; see Appendix A.1 for more detail.

Our survey reveals wide variation in how experimenters deliver and reinforce instructions. Nevertheless, there are commonalities which seem to reflect some notion of 'best practices.' Few studies have tested whether current practices are effective – our experiment is designed to fill this gap.

1.3 Experimental Design

Overview of Experiment

We design a one-shot, individual choice experiment in which each subject performs two tasks, a base task which provides a low flow of payoffs throughout the experiment, and a second task which can only be completed once and results in a potentially large lump-sum payoff. The amount of the lump sum depends on the time at which they initiate the second task. Doing the second task too early results in a lower payoff than doing it at the right time (or not doing it at all).

Task 1 is the Poodle Jump game (based on a popular mobile game Doodle Jump), where players guide a bouncing poodle up a series of platforms by pressing two buttons. When a subject misses a platform, the poodle falls to the ground and the game restarts with no penalty. Each participant receives \$0.25 per period of Task 1, so long as they jump a minimum cumulative height. This height was chosen so that it would be trivially easy to complete but not automatic – effectively guaranteeing an attentive subject this payment each period.[4]

Task 2 is a simplified version of the slider task (Gill and Prowse 2012). Players can switch from Task 1 to Task 2 at any time by pressing the 'j' key, but they can only do this once. In the slider task, players are presented with four sliders which can be moved from zero to 100. The task is successfully completed when all four sliders are dragged to 50 and the player clicks "Continue."[5] Task 2's payoff depends on when the subject presses 'j'. For the first 21 periods, each period being one minute long, it pays \$0.20. However, in period 22 it jumps to \$7, falling to \$4 in period 23, then dropping by \$0.50 in every period thereafter

[4]Only 5 out of 308 subjects ever failed to attain the required height in a period; 4 did so once and one subject did so twice. These failures account for only 0.1% of all Poodle Jump periods.

[5]Only one subject started but failed to complete the slider task.

until period 30 when the experiment ends. These payoffs are demonstrated in Figure 1.1. Doing Task 2 in period 22 maximizes a subject's payoff; whereas, doing it before period 22 minimizes a subject's payoff. If a subject fails to do Task 2 in period 22, they would always earn higher payoffs by doing it as soon as possible thereafter.

Figure 1.1: Screenshot showing how payoffs were described to subjects

The challenge for subjects is to recognize and remember the correct time to press the 'j' key to complete Task 2, given the attention required to successfully complete Task 1 in each period. However, subjects have strong incentive to complete Task 2 at the right time: doing Task 2 at the right time raises payoffs by \$6.75 relative to not doing it at all, and by a minimum of \$3 compared to completing it at any other time. Moreover, doing it before period 22 leads the subject to forgo the opportunity to do it at the optimal period

Table 1.2: Summary of treatments

Treatment	Quiz	Answers	Additional Reinforcement	# of Subjects
NO QUIZ	No	No	No	43
QUIZ	Yes	No	No	76
ANSWERS	Yes	Yes	No	36
INCENTIVE	Yes	Yes	Pay 0.50 CAD per correct quiz answer	38
TWICE	Yes	Yes	Instructions restarted unexpectedly	38
PAPER	Yes	Yes	Instructions duplicated in paper printout	40
ENHANCED	Yes	No	Only through enhanced on-screen instructions	37

or thereafter, and also results in a lower payoff than never doing Task 2. Thus, doing Task 2 before period 22 precludes the subject from maximizing their monetary payoffs. We use the NMB acronym to refer to such behavior below.

NMB can thus reveal that a subject failed to comprehend or retain a particularly key piece of payoff-relevant information from the instructions.[6] As hinted at earlier, our design restricts the set of possible preference-based explanations for NMB. Moreover, since we sample subjects from the same distribution of preferences in each treatment, variation in NMB across treatments identifies changes in comprehension and retention.

Treatment Design

We employ a between-subjects design with seven treatments. We study the effectiveness of different ways of delivering and reinforcing the experiment's instructions on NMB using our aforementioned measure. Many experimenters implicitly assume that subjects fully understand their instructions. If this is true, we should not observe any difference between treatments. However, if subjects do not always comprehend or retain information from the instructions there is the potential for variation in delivery and additional reinforcement to reduce NMB. Our treatments test the impact of various more-or-less standard procedures employed by experimenters to improve comprehension and retention. All treatments are summarized in Table 1.2. All treatments started with a common set of self-paced on-screen instructions, which included a graphical explanation of payoffs as well as reinforcement from practice rounds for both tasks and practice switching between tasks.

The NO QUIZ treatment presents the instructions on screen with no additional reinforcement. The NO QUIZ treatment gives us information on NMB when subjects read instructions on their own.

[6] We note that neither a subject who understood and retained this information but simply forgot to switch nor a subject who (for whatever reason) did not understand this information but only switched at or after period 22 would be coded as exhibiting NMB by this measure.

The QUIZ treatment was identical to the NO QUIZ except that each subject completed a six question comprehension quiz on paper at the end of the on-screen instructions; subjects were informed that there would be a quiz prior to beginning the instructions, but no feedback was given on the quiz. The QUIZ treatment allows us to assess whether the presence of the quiz affects NMB, and the quiz itself gives a secondary measure of comprehension. When we analyze our data, we use this as our baseline treatment for comparison to the other treatments below.

The ANSWERS treatment was identical to the QUIZ treatment, except that subjects were presented the answers to the quiz orally after all had completed it. This corrected possible misunderstandings revealed in quiz answers and reinforced key pieces of information from the instructions. As noted by Cassar and Friedman (2004), a quiz is a good way to "make sure that the subjects understand the rules" (p. 71); thus we expect providing the answers to the quiz will correct failures of comprehension or retention and reduce NMB.

The TWICE treatment was identical to the ANSWERS treatment except that after completing the quiz and answers, the experimenter unexpectedly restarted the instructions for the participants to work through a second time. This allowed subjects to further review any content they missed on the first go and provided additional reinforcement. As noted by Friedman and Sunder (1994), "[when] a subject does not seem to understand the instructions [...] the experimenter may reread the relevant part of the instructions or go through an example" (p. 77). Repeating the instructions TWICE achieves both of these objectives and thus should reduce NMB.

The INCENTIVE treatment was identical to the ANSWERS treatment except that subjects were paid $0.50 for each correct quiz answer, and were informed of this before starting the instructions. We hypothesized that this would lead subjects to pay more attention to the material in the instructions, and make any mistakes from the quiz more salient, thereby improving understanding. Pay for performance is standard in experimental economics because economists believe it motivates subjects to think carefully and participate actively in experiments (Hertwig and Ortmann, 2001). By paying for performance on the quiz, we anticipate that subjects will exert more effort in carefully reading the instructions, thereby reducing NMB.

The PAPER treatment was identical to the ANSWERS treatment except that the experimenter also distributed paper printouts of the instructions (in addition to the on-screen instructions), which participants could keep and reference at any time, even while completing the quiz.[7] We thus expect PAPER to improve comprehension as measured by quiz scores and reduce NMB both for this reason, and through improving retention given the quiz score since written instructions are available throughout the session.

[7]The PAPER treatment potentially reduces forgetfulness since all relevant information is accessible throughout the experiment.

The ENHANCED treatment was identical to the QUIZ treatment but with enhanced on-screen instructions.[8] Compared to the other treatments, the on-screen instructions were lengthened from five to seven screens in length. In these enhanced instructions, Figure 1.1 appeared four times (instead of only once), and subjects were presented with four worked-out examples that explained the payoff that would result from different possible switching times. Unlike in our other treatments, the last page of the enhanced instructions included Figure 1.1, and each subject waited on that page while other subjects completed the instructions and while they completed the quiz. With the benefit of hindsight, we emphasized the details we knew past subjects had failed to grasp. This treatment is also consistent with the advice of Friedman and Sunder (1994), applied between-subjects, and we expect the ENHANCED instructions to similarly reduce NMB.

For reasons explained above, we hypothesize that each additional form of reinforcement reduces NMB. Specifically, we conjectured that having a QUIZ would have a similar level of NMB as NO QUIZ, but relative to these treatments, ANSWERS would reduce NMB, each of our remaining interventions on top of that (INCENTIVE, TWICE, and PAPER) would further reduce NMB, and ENHANCED would also reduce NMB relative to QUIZ. We hypothesized that higher quiz scores will be associated with lower rates of NMB, and that in the INCENTIVE, PAPER, and ENHANCED treatments most or all reductions in NMB are reflected in higher quiz scores, while the ANSWERS and TWICE treatments reduce NMB given quiz scores.

Our experiment differs from existing studies on instructions in two regards. First, this is an individual decision task, so there is neither complexity from strategic behavior nor other-regarding concerns. Second, it is a one-shot task – each subject can only press ‘j’ once – so participants who fail to understand the instructions cannot learn through trial and error. These features allow us to cleanly identify NMB and attribute variation in NMB to variation in the delivery and reinforcement of instructions. Nonetheless, we believe that our experiment provides a good analogy to other experiments, particularly those where a decision of interest is only one of multiple decisions the subject makes. We also conjecture that more complicated experiments face at least as much risk of misunderstanding as exists in our simple experiment (even if most existing experiments are unable to diagnose it).

Procedures

Upon entering the lab, the experimenter assigned participants to visually isolated computer terminals. Participants were told not to interact with one another for the duration of the experiment. In all treatments, participants were informed that they would be given a set of instructions followed by an experiment in which they could potentially earn a significant amount of money; in the treatments with a quiz, they were also informed that there would be

[8]The ENHANCED treatment was added later on a suggestion from the editor.

a quiz at the end of the instructions; subjects in the INCENTIVE treatment were informed that they would be paid for their quiz performance above and beyond their earnings from the experiment. The experimenter then started the self-paced on-screen instructions which included a written description of the tasks and the payoff structure, practice rounds of both tasks, practice switching between tasks, and a graphical illustration of the payoffs to both tasks in each period (a full copy of the instructions are presented in Appendix A.2). Once all participants completed the instructions, the experimenter distributed the quiz in the QUIZ, ANSWERS, INCENTIVE, TWICE, PAPER, and ENHANCED treatments; the correct answers were revealed after all participants had completed the quiz except in the QUIZ and ENHANCED treatments. In the TWICE treatment, subjects completed the on-screen instructions a second time, including practice rounds. Then the experiment started. At the end of some sessions, we conducted a post-experiment questionnaire (Appendix A.4).[9]

We recruited 308 participants to 45 sessions through Simon Fraser University's CRABE recruiting system, with no subject participating in more than one session. Each session lasted under an hour. Average earnings were 18.37 CAD including a 7 CAD show-up payment. We collected no other demographic data nor other behavioral measures.

1.4 Results

We use a subject's decision to do Task 2 at any time before period 22 as NMB, which is our behavioral measure of their failure to pay attention to, comprehend, absorb, or retain information from the instructions. Table 1.3 shows the share of NMB by treatment. All p-values reported below are two-sided.

Finding 1: *NMB* is prevalent.

In our NO QUIZ and QUIZ treatments, 44% and 47% of subjects exhibited NMB by doing Task 2 before period 22. This is despite the fact that these treatments include both demonstrations and practice periods. Even in our most effective treatment, the corresponding share is 18%. These findings suggest that failures to comprehend or retain information from instructions may be an important source of noise.[10] This justifies concern about the effectiveness of instruction delivery and reinforcement methods.

Finding 2: Combining reinforcement methods reduces *NMB*.

We find that additional reinforcement reduces NMB: we reject the joint hypothesis that NMB occurs at the same rate across all treatments (Fisher's exact test, $p < .01$, $n = 308$).

[9] We have responses from 72 subjects because this was added at the suggestion of a referee.

[10] In Appendix A.3, we show that we find similar results if we account for trembles by defining NMB based on doing Task 2 before period 21.

Table 1.3: *Non Money-maximizing Behavior* across treatments

	NO QUIZ	QUIZ	ANSWERS	INCENTIVE	TWICE	PAPER	ENHANCED
NMB	.442	.474	.333	.237	.184	.225	.216
Quiz Score (avg.)	n/a	4.10	4.06	4.32	4.53	5.43	4.59
QUIZ	.849						
ANSWERS	.362	.220					
INCENTIVE	.064	.016	.442				
TWICE	.017	.004	.186	.779			
PAPER	.062	.010	.316	1.00	.781		
ENHANCED	.056	.013	.302	1.00	.779	1.00	

First row reports the fraction of NMB by treatment.
Second row reports the average quiz score by treatment. Remaining entries report a p-value from a Fisher's exact test of differences in NMB between treatments.

Compared to NO QUIZ and QUIZ, we observe somewhat less NMB in the ANSWERS treatment (33%), but we do not detect any statistically significant differences between these treatments (Fisher's exact test of equal NMB rates across these treatments, $p = .35$, $n = 155$). In each of the INCENTIVE (24%), TWICE (18%), and PAPER (23%) treatments that provide additional reinforcement, subjects exhibited significantly less NMB than in the QUIZ treatment (Fisher's exact tests, $p < .02, .01, .01$, $n = 114, 114, 116$ respectively). While the ENHANCED treatment (22%) reduces NMB (Fisher's exact test, $p = .01$, $n = 113$), it does not eliminate it.[11] Our findings suggest that more detailed instructions and extensive reinforcement each improve comprehension and retention of the instructions.

Finding 3: Lower quiz scores are associated with NMB. Providing quiz answers while also making incorrect answers salient can reduce NMB among lower performers.

Quiz scores provide an alternative measure of subject comprehension immediately after the instructions. In the QUIZ treatment which provides neither feedback nor additional reinforcement, quiz score and NMB are negatively related (Goodman-Kruskal γ, $p < 0.01$, $n = 76$); indeed 13 of 76 subjects had a perfect score on the quiz, and none of them subsequently exhibited NMB in the experiment. In fact, across all of our treatments we find it striking that only one of the 73 people with a perfect quiz score exhibited NMB.[12] This indicates that full *comprehension* at the completion of the instructions appears to be a sufficient condition for avoiding NMB in our experiment and that *retention* is a second-order issue.

Our quiz score data enable us to test whether the INCENTIVE, PAPER, and ENHANCED treatments improved subjects' comprehension as demonstrated on the quiz, compared to the pooled distribution of quiz scores from the QUIZ, ANSWERS, and TWICE treatments, which followed identical procedures up to the collection of the quiz.[13] Average quiz scores by treatment are reported in Table 1.3. To our surprise, neither the INCENTIVE nor the ENHANCED treatment significantly improved quiz scores (rank-sum tests, $p = .59, .14$, $n = 188, 187$, respectively). The PAPER treatment, which made the answers accessible to subjects during the quiz, improved scores significantly (rank-sum test, $p < 0.01$,

[11] We cannot reject the hypothesis INCENTIVE, TWICE, PAPER, and ENHANCED lead to similar improvements (Fisher's exact test of no association, $p = .96$, $n = 153$).

[12] One person with a perfect quiz score in the TWICE treatment switched 28 seconds too early.

[13] We find no significant differences in the distribution of quiz scores in the QUIZ, ANSWER, and TWICE treatments (Kruskal-Wallis test, $p = .15$, $n = 150$).

Table 1.4: Treatment effects on Non Money-maximizing Behavior and Quiz Scores

	Dependent variable			Mediation analysis	
	NMB		Quiz Score	NMB	
	(1)	(2)	(3)	(4)	
NO QUIZ	-0.128				n
	(-0.889, 0.632)				
ANSWERS	-0.588	0.051	-0.050	-0.138	112
	(-1.424, 0.248)	(-2.884, 2.987)	(-0.586, 0.487)	(-0.308, 0.048)	
ANSWERS × Quiz Score		-0.206		0.00715	
		(-0.941, 0.528)		(-0.069, 0.085)	
INCENTIVE	-1.065**	-2.531*	0.211	-0.219	114
	(-1.948, -0.182)	(-5.332, 0.271)	(-0.354, 0.775)	(-0.392, -0.030)	
INCENTIVE × Quiz Score		0.361		-0.028	
		(-0.257, 0.978)		(-0.111, 0.049)	
TWICE	-1.383***	-2.181	0.421	-0.255***	114
	(-2.329, -0.436)	(-5.235, 0.873)	(-0.156, 0.998)	(-0.425, -0.065)	
TWICE × Quiz Score		0.207		-0.057	
		(-0.467, 0.880)		(-0.145, 0.021)	
PAPER	-1.131**	7.334*	1.320***	0.134	116
	(-2.009, -0.253)	(-0.810, 15.478)	(0.922, 1.718)	(-0.189, 0.343)	
PAPER × Quiz Score		-1.485*		-0.177***	
		(-2.970, 0.001)		(-0.273, -0.085)	
ENHANCED	-1.182**	-0.557	0.489*	-0.188*	113
	(-2.096, -0.269)	(-4.596, 3.482)	(-0.030, 1.008)	(-0.382, 0.013)	
ENHANCED × Quiz Score		-0.116		-0.067*	
		(-0.993, 0.760)		(-0.151, 0.004)	
Quiz Score		-0.679***			
		(-1.053, -0.306)			
Intercept	-0.105	2.683***	4.105***		
	(-0.561, 0.350)	(0.993, 4.374)	(3.789, 4.422)		
Observations	308	265	265		

QUIZ is the omitted category. *, **, and *** respectively denote $p < .1$, $p < .05$, $p < .01$. Robust (HC1) 95% confidence intervals are in parentheses in Columns (1)-(4). Mediation column reports estimated "direct effects" in the row of a treatment dummy, and mediated effects in the row of the interaction term between Quiz Score and that treatment dummy, both evaluated relative to the QUIZ baseline. That is, the direct effect of a treatment corresponds to $\mathbb{E}[\text{NMB}|\text{Treatment}, \text{Quiz Score} = 4.1] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$, while the mediated effect corresponds to $\mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = \mathbb{E}[\text{Quiz Score}|\text{Treatment}]] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$.

$n = 190$), and the linear regression in Table 1.4, column 3 shows that PAPER had the largest effect on quiz score of all of our treatments.[14]

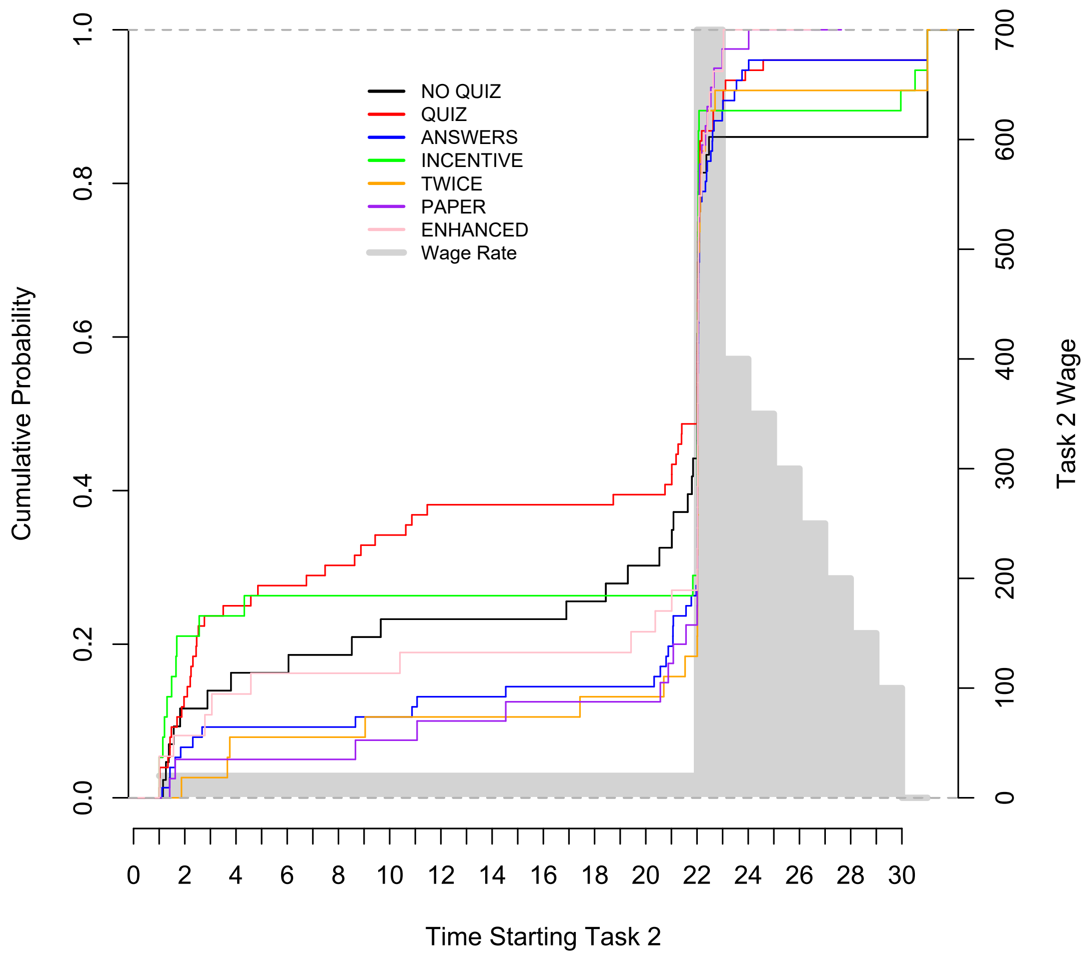
Quiz score data also allow us to further assess *how* our treatments reduce NMB. Goodman-Kruskal γ tests revealed that quiz score had a significant ($p < .05$ in each test, $n = 76, 36, 38, 40, 37$ respectively for each of QUIZ, ANSWERS, TWICE, PAPER, and ENHANCED) negative relationship with NMB in each treatment except INCENTIVE (where $p = .054$, $n = 38$) and NO QUIZ (where scores were not available). To decompose the extent to which treatment effects operate via (i.e. are *mediated* through) improved comprehension demonstrated on the quiz, we perform mediation analysis (applying the approach of Imai, Keele, and Yamamoto 2010) in column 4 of Table 1.4, based on a model of NMB as a logistic-linear function of quiz score, treatment, and their interactions (column 2), and a linear regression to model treatment effects on quiz scores (column 3). The INCENTIVE and TWICE treatments have sizable and significant direct effects but insignificant and small mediated effects.[15] This indicates that these treatments primarily reduce NMB by clearing up (TWICE) and making salient (INCENTIVE) failures of comprehension demonstrated on the quiz. In contrast, the PAPER treatment has the largest mediated effect of all treatments, which is statistically significant, but only a small and insignificant direct effect beyond that. Mediated and direct effects of the ENHANCED treatment are each borderline insignificant, indicating a mix of both types of effects, but point estimates indicate a larger direct effect.

Robustness Checks Figure 1.2 shows empirical CDFs of completion times for Task 2, by treatment. For robustness, we show in Appendix A.3 that we would arrive at similar qualitative conclusions to those reported in Table 1.4 using any of three alternative measures of NMB which vary the strictness of the criteria by which we classify behavior as NMB.

[14]The positive effect of paper instructions on quiz performance is consistent with the evidence reported in Bigoni and Dragone (2012).

[15]In the case of TWICE, this is reassuring since any mediated effect can only arise due to sampling variation.

Figure 1.2: Empirical CDFs of Task 2 completion times, by treatment



Our post-experiment questionnaire was only partially able to diagnose causes of NMB in our experiment (see Appendix A.4 for a full analysis). While subjects' responses are correlated with behavior and quiz scores, they fail to provide any indication of the differences between the QUIZ and ENHANCED treatments in NMB revealed in the experiment.

1.5 Discussion

Our experiments indicate that even when using combinations of reinforcement methods including demonstrations, practice periods, and a quiz, many subjects' behavior reveals that they fail to pay attention to, understand, or retain information from the instructions. Combining these with further reinforcement methods reduced NMB, as did increasing the level of detail in the instructions. Each of these methods leads to a similar improvement but does not eliminate NMB.

In our setting, we feel confident attributing variation in the anomalous behavior that we observe to a variation in the failure to understand or absorb the instructions. In other experiments designed to test for anomalous behavior, the distinction between truly anomalous behavior of interest and a failure to understand the instructions may not be so clearcut. This justifies a concern with how instructions are given and the use of behavioral checks of understanding. Our findings broadly suggest that experimenters' attempts to reinforce the instructions or make them more salient can be effective at reducing NMB. Note that though we are able to reduce NMB in our design, some residual NMB persists even in the best case. While the extent of such NMB is likely to vary with experimental context (e.g. subject pool, design, feedback), its presence is noteworthy and has implications for the power and interpretation of experimental tests.

Finally, our findings motivate advice on how to report and deliver instructions. First, experimenters should be aware that the way instructions are delivered and reinforced has consequences for behavior. Second, we suggest providing paper instructions when possible, since this requires no extra lab time, is almost free, and is about as effective in reducing NMB in our experiment as other reinforcement methods. Third, we suggest that all experimental papers should clearly report how they deliver and reinforce instructions, as this can be crucial for close replication and interpretation.[16] Journals' efforts to require experimenters to share copies of their instructions are laudable, and these could be complemented by standardized reporting of how instructions are delivered and reinforced.

[16]For example, recent work by Chen et al. (2018) demonstrates, via new experiments following different instructions protocols, that a recent failed replication attempt arose because of differences in how instructions were delivered.

Chapter 2

Anchors of Strategic Reasoning in the Traveler's Dilemma

2.1 Introduction

Decades of lab experiments have documented that play by human subjects often violates the predictions of Nash equilibrium in games without feedback. The most prominent class of models in behavioral game theory explains these deviations as a result of limited strategic reasoning (Nagel 1995, Stahl and Wilson 1995). In these models, a player anchors her strategic reasoning with initial beliefs about the play of a non-strategic player. To select an action, she iteratively calculates best-replies a finite number of times starting with non-strategic play. Much of the behavioral game theory literature has focused on estimating players' strategic sophistication while controlling for initial beliefs, and does so by studying games where different plausible initial models of non-strategic play lead to identical best replies (e.g. Nagel 1995, Arad and Rubinstein 2012) or using an identification strategy that is insensitive to the details of how a player anchors her strategic reasoning (Kneeland 2015). In general, when a player only reasons a finite number of steps, the anchor of a player's strategic reasoning may matter. However, the extant literature in behavioral game theory has barely attempted to study how players form the initial beliefs about non-strategic play that anchor their strategic reasoning.

We thus conduct an experiment to test the separate predictions of three plausible models of non-strategic play in the traveler's dilemma game (Basu 1994). In the traveler's dilemma game, two players each make a monetary "claim" that must lie between an upper and a lower bound. Each player receives the lower of the two claims, but if the claims are unequal, then the player who made the lower claim receives a transfer from their opponent. In the traveler's dilemma, the upper bound, lower bound, and middle of the strategy space are salient and plausible specifications of non-strategic play, as is uniform randomization across the strategy space. Each player has the incentive to undercut the claim they expect the other player to make by as little as possible. Given a lack of knowledge of the opponent's claim, a player thus faces a trade-off – undercutting more reduces one's payoff, but risks being undercut and not obtaining the transfer. Given a fixed number of steps of strategic reasoning, each of these models of initial beliefs about non-strategic play makes distinct predictions about play given fixed game parameters. By observing how an individual's play varies as we vary the upper and lower bounds of the strategy space and the transfer, we can test the competing predictions of these four models of how they anchor their strategic reasoning while simultaneously measuring their strategic sophistication.

We thus classify each subject based on their play in 30 rounds of the traveler's dilemma with different game parameters. We find that 27% of players tend to play Nash equilibrium (or observationally equivalently, anchor their reasoning at the lower bound), 33% are boundedly strategic, while 40% are non-strategic. A majority of boundedly strategic players (55%) anchor their strategic reasoning at the upper bound, and a similar fraction of non-strategic players (58%) tend to non-strategically play the upper bound. Some strategic players (35%) anchor their reasoning on and some non-strategic players (33%) tend to play the middle of the strategy space. Notably, we classify only one strategic subject as anchoring on uniformly random initial beliefs. We thus conclude that players are both heterogeneous in the amount that they reason, and how they anchor their strategic reasoning, with anchoring on uniform randomization being rare.

Related Literature. Our paper contributes to the existing literature in behavioral game theory by providing individual-level estimates of how players initiate their strategic reasoning. Our work takes leading models of limited strategic reasoning as a starting point, namely, the level k model (Nagel 1995, Stahl and Wilson 1994, 1995) and the noisy introspection model (Goeree and Holt 2004).[1] These models assume that each player anchors their strategic reasoning on beliefs about how a non-strategic player, denoted $L0$, would play. In this literature "L0 is usually assumed to be uniform random over others' possible decisions" (Crawford et al. 2013); examples include Stahl and Wilson (1994, 1995), or estimates players' strategic sophistication in games where most plausible models of $L0$ generate to the same best responses (Nagel 1995, Costa-Gomes and Crawford 2006 , Arad and Rubinstein 2012, Kneeland 2015). This literature finds that the majority of players behave as if they compute 1-4 levels of reasoning, while almost no one is classified as $L0$. We instead study the traveler's dilemma (Basu 1994) because it has multiple plausible $L0$ specifications that each generate different best replies, allowing us to test between $L0$ specifications. Past work showed that subjects frequently play non-rationalizable strategies in this game, even after playing multiple rounds with feedback (Capra et al. 1999). Our study of how players anchor their strategic reasoning complements three existing strands of experimental work that estimate how people reason in strategic settings.

[1] Crawford et al. (2013) survey this literature.

A first strand of work estimates how players reason in two player normal form games using econometric specifications of level k models. This literature typically assumes that $L0$ uniformly randomizes. Early work by Stahl and Wilson (1994) showed that the vast majority of players are best classified as $L1$ or $L2$. Costa-Gomes et al. (2001) use players' information look-ups to disentangle a broader class of models, but nevertheless find similarly. Wright and Leyton-Brown (2019) estimate alternative specifications for $L0$ using data from experimental studies of normal form games. They find that the most predictive $L0$ specification deviates from uniform randomization by putting additional weight on actions based on their minmax, maxmax, and maxmax fairness evaluations, and by their distance from the symmetric maxmax action when available. Their best performing model puts weight 45% on uniform randomization but results on an approximately 80% weight on uniform randomization when applied to the traveler's dilemma.[2] However, unlike our study, none of the aforementioned papers study heterogeneity in $L0$ across individuals. In addition, our focus on a particular class of games – the traveler's dilemma – allows us to specify a class of $L0$ specifications a priori.

A second strand of work that uses a non-neutral framing to induce a particular pattern to $L0$ (following Rubinstein et al. (1997)) and estimates that $L0$ is sensitive to salience Crawford and Iriberri (2007). Bardsley et al. (2010) measure non-strategic play and beliefs about non-strategic play in coordination games by studying play in two ancillary games, one where players "pick" an action from the games strategy space but without any incentives, another where players are paid to "guess" the actions of the pickers. They compare the actions chosen by pickers, guessers, and subjects who play the underlying coordination game to test for alignment between non-strategic play, guessers' beliefs thereof, and coordinators' beliefs. Hargreaves-Heap et al. (2014) study experimental hide-and-seek games, where, following Rubinstein et al. (1997) and Crawford and Iriberri (2007), non-neutral labels are designed to induce a non-uniform $L0$. They reject the assumption that players' implied anchors of their strategic reasoning depend only on non-strategic features of the game. Penczynski (2016) infers $L0$ and levels of reasoning from text communication between partners in hide-and-seek games and finds evidence of role-asymmetric $L0$s that respond to non-neutral frames. In contrast, the neutrally-framed traveler's dilemma motivates a different class of $L0$ specifications.

Our paper also complements a third strand of work on belief formation in experimental guessing games and related games with incentives to undercut, following Nagel (1995)'s pioneering study of the 2/3 beauty contest. Costa-Gomes and Crawford (2006) use guesses and information look-ups in a variety of two player guessing games to estimate levels of

[2]Approximate percentages are from Figure 6 of Wright and Leyton-Brown (2019). Since any individual action can achieve a symmetric payoff when the other player takes the same action, their notion of maxmax fairness puts each weight on all actions in the traveler's dilemma.

sophisticaton while assuming a uniform random $L0$, and classify most subjects as $L1$ or $L2$, but with up to 38% of subjects not well classified by any strategic type. In a similar class of games, Fragiadakis et al. (2016) fail to classify 70% of subjects and find that most of these subjects cannot mimic or best-reply to their own past play in identical game, and suggest that they may be poorly described by any model based around deterministic strategic reasoning. Burchardi and Penczynski (2014) apply the method used in Penczynski (2016) to Nagel (1995)'s 2/3 beauty contest and find that the modal communicated $L0$ belief is in the middle of strategy space (50), but the average is slightly higher (55); 20% of their subjects are classified as $L0$. Agranov et al. (2015) use incentivized choice process data to study the process of reasoning in the 2/3 beauty contest, and classify 45% of subjects as $L0$ – in contrast to earlier findings (reviewed in Crawford et al. (2013)) that classifies almost no one to $L0$. Fragiadakis et al. (2019) elicit subjects' beliefs about others' in an undercutting game. They find that most subjects put positive weight on multiple other behavioral types, consistent with the cognitive hierarchy model of Camerer et al. (2004) but inconsistent with models like level k in which each player makes a point prediction about their opponent's play. Our paper contributes by providing individual-level tests of concrete alternative specifications of how players anchor their strategic reasoning in an undercutting game where a uniform random $L0$ generates different predictions for behavior than a $L0$ that plays in the middle of the action space.

The rest of the paper is organized as follows. Section 2.2 presents the theoretical predictions. Section 2.3 presents the experimental design. Section 2.4 presents the results. Section 2.5 concludes.

2.2 Theoretical Predictions

In the traveler's dilemma game, each of two subjects simultaneously and independently makes a "claim", which is an integer $x_i \in \{\underline{x},\ldots,\bar{x}\}$ from a range specified by a lower bound $\underline{x}$ and an upper bound $\bar{x}$. Subject i's payoff, π_i, is given by

$$\pi_i = \begin{cases} x_i + R & \text{if } x_i < x_{-i} \\ x_i & \text{if } x_i = x_{-i} \\ x_{-i} - R & \text{if } x_i > x_{-i} \end{cases}$$

where R is a reward/penalty parameter.If the claims are different, both players receive the lower of the two claims, and the player who made the higher claim transfers R to the player who made the lower claim. In case of equal claims, both players receive what they claimed. Thus, if $R > 0$, each player has the incentive to undercut the other player's claim by 1. The triple $\langle \underline{x}, \bar{x}, R \rangle$ fully describes a parameterization of the traveler's dilemma game.

We consider the following four leading models from play in games: Nash Equilibrium, Quantal Response Equilibrium (QRE; Mckelvy et al. 1995, Goeree et al. 2016), Level k (Nagel 1995; Stahl and Wilson 1995), and Noisy Introspection (NI; Goeree and Holt 2004). Nash Equilibrium and Quantal Response Equilibrium are both equilibrium models in which play is respectively a deterministic and stochastic best-reply to the behavior of other players. In these equilibrium models, the anchor of strategic reasoning does not matter. In contrast, Level k and NI are models in which a player respectively best replies to a deterministic or stochastic less sophisticated player(s). Our interest is in the distinction between different models of anchors of strategic reasoning, and not on distinguishing between Level k and NI given any fixed specification of how players anchor their strategic reasoning, and thus we consider both models.

Nash Equilibrium. Since $R > 0$, each player has the incentive to undercut the other's claim. That is, given the conjecture that the other player never plays actions above $\hat{x}$, then $\hat{x}$ is dominated by $\hat{x} - 1$. By induction, the game is dominance solvable and claiming $\underline{x}$ is the unique rationalizable strategy (and unique Nash Equilibrium strategy) for each player.

Level k. The level k model is a non-equilibrium model of limited strategic reasoning wherein a player iteratively best-replies to a model of non-strategic play, referred to as $L0$. A $L1$ player best-replies to a $L0$ play, and more generally, a Lk player best-replies to $L(k-1)$ play. The parameter k captures the number of steps, or level, that a player reasons. Unlike in most games studied in lab experiments, the predictions of the level k model across different parameterizations of the traveler's dilemma depend on the specification of $L0$. This gives us the ideal setting to test the contrasting predictions of different models of $L0$.

We consider three models of $L0$:

1. "Top", where $L0$ plays $\bar{x}$;
2. "Middle", where $L0$ plays $\frac{\bar{x}+\underline{x}}{2}$; and
3. "Uniform", where $L0$ uniformly randomizes.

Each model of $L0$ makes a separate set of predictions for strategic players:[3] under $L0$-top, Lk claims $\bar{x} - k$. Under $L0$-middle, Lk claims $\frac{\bar{x}+\underline{x}}{2} - k$. Under $L0$-uniform, Lk is indifferent between claiming $\overline{x} + 1 - 2R - k$ and $\overline{x} + 2 - 2R - k$. We note that a fourth model of $L0$ where $L0$ plays $\underline{x}$ makes identical predictions to Nash Equilibrium.

Quantal Response Equilibrium. QRE assumes that each player noisily best responds to the equilibrium distribution of play. Specifically, let p denote a probability distribution

[3] We provide full derivations in Appendix B.1.

over actions, and let $U(x_i = c|p)$ denote the expected payoff to playing $x_i = c$ if the other player's distribution over actions is given by p. Then, p is a QRE if

$$p_c \equiv \text{Prob}(x_i = c|\gamma) = \frac{\exp\left(\gamma U(x_i = c|p)\right)}{\sum_{j=\underline{x}}^{\bar{x}} \exp\left(\gamma U(x_i = j|p)\right)}$$

The parameter γ captures the precision of play – with higher γ indicating less noisy play.[4] In the QRE model, a player's claims will have a smooth distribution and centered on their modal claim, which must be strictly below $\bar{x}$.

Noisy Introspection. NI is a non-equilibrium model and can be viewed as a noisy version of level k where a player iteratively and noisily best-replies to a model of non-strategic play, referred to as $NI0$ (which plays an identical role to $L0$ in the level k model). An $NI1$ player noisily best-replies to $NI0$ play, and more generally, a NIk player noisily best-replies to $NI(k-1)$ play. As in QRE, noisy behavior is modeled using a logit formula. Let p denote a probability distribution over actions. Then, for a NIk player,

$$p_c^k \equiv \text{Prob}(x_i = c|\gamma) = \frac{\exp\left(\gamma U(x_i = c|p^{k-1})\right)}{\sum_{j=\underline{x}}^{\bar{x}} \exp\left(\gamma U(x_i = j|p^{k-1})\right)}$$

As in level k model, we consider three models of $NI0$: "top", where $NI0$ plays $\bar{x}$; "middle", where $NI0$ plays $\frac{\bar{x}+\underline{x}}{2}$; and "uniform", where $NI0$ uniformly randomizes. Each model of $NI0$ also makes a separate set of predictions for higher levels, for reasons analogous to the level k model.

Error Structure. QRE and NI models predict a distribution over claims. In contrast, Nash and level k models (with a fixed k and $L0$) make point predictions. Since each subject's behavior may be noisy, we estimate two noisy versions of these two models. In the action tremble specifications, each player plays according to the model with probability $1 - \epsilon$ and uniformly randomizes with probability ϵ. In the payoff tremble specification, subjects are assumed to compute the expected distribution of other's behavior, p, according to the model, but stochastically best replies to this according to the logit model, that is, $\text{Prob}(x_i = c|p) = \frac{\exp(\gamma U(x_i=c|p))}{\sum_{j=\underline{x}}^{\bar{x}} \exp(\gamma U(x_i=j|p))}$.[5]

[4]As $\gamma \to \infty$, the model converges to Nash Equilibrium play, whereas lower levels of γ induce noisier behaviors.

[5]Unlike in the QRE, p is not required to be in equilibrium in our level k and NI specifications.

2.3 Experimental Design

Each subject was randomly and anonymously matched with another subject to play 30 rounds of the traveler's dilemma game with different parameters, divided into three blocks. Each block has the same ten pairs of lower bound-upper bound parameters but a different reward/penalty parameter R (Table 2.1); the first and third blocks have $R = 5$, while the second block has $R = 20$. Within each block, we vary the lower and upper bounds across rounds in the following four ways: varying the upper bound only (e.g. round 1 versus 6), varying the lower bound only (e.g. round 5 versus 8), varying both the lower and upper bounds while keeping the middle of the range comparable to another pair of lower-upper bound (e.g round 4 versus 9), varying both the lower and upper bounds while keeping the gap between the bounds comparable to another pair of lower-upper bound (e.g. round 2 versus 3). Subjects received no feedback between decisions. One round was randomly chosen for payment at the end of the experiment.

Table 2.1: Round parameters

Round	Lower bound	Upper bound
1, 11, 21	20	120
2, 12, 22	80	200
3, 13, 23	40	160
4, 14, 24	20	180
5, 15, 25	50	200
6, 16, 26	20	160
7, 17, 27	60	180
8, 18, 28	100	200
9, 19, 29	50	150
10, 20, 30	40	200

$R = 5$ for rounds 1 to 10 and rounds 21 to 30

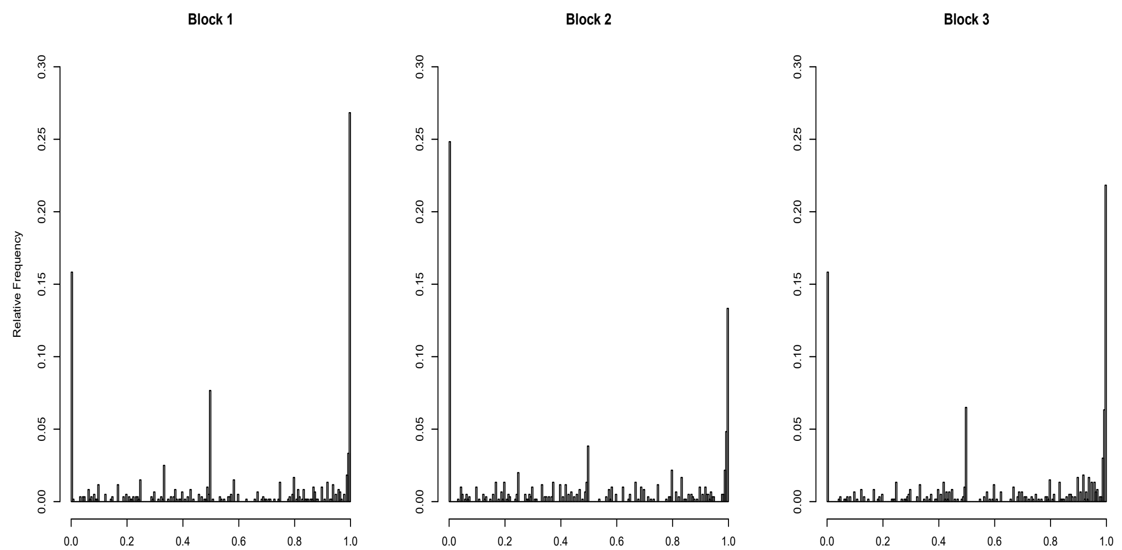
$R = 20$ for rounds 11 to 20

We recruited 60 subjects from the experimental economics recruitment pool at Simon Fraser University to participate in experimental sessions between July 2018 and February 2019. Each session lasted approximately 45 minutes. Each subject received a minimum \$7 (CAD) show-up fee in addition to their experiments earnings, which were converted from experimental currency units to dollars at a rate of 1 ECU = \$0.10; the average payment was \$15.50.

2.4 Results

To compare claims across rounds with different lower and upper bounds, we compute a normalized claim $x^n_{ig} = \frac{x_{ig}}{\bar{x} - \underline{x}} \in [0, 1]$ in each game g for each subject i. Figure 2.1 shows that normalized claims are spread out over the feasible ranges with three spikes at exactly the lower bound (19% of the data), the middle (6% of the data) and the upper bound (21% of the data). These three spikes account for 46% of the data.

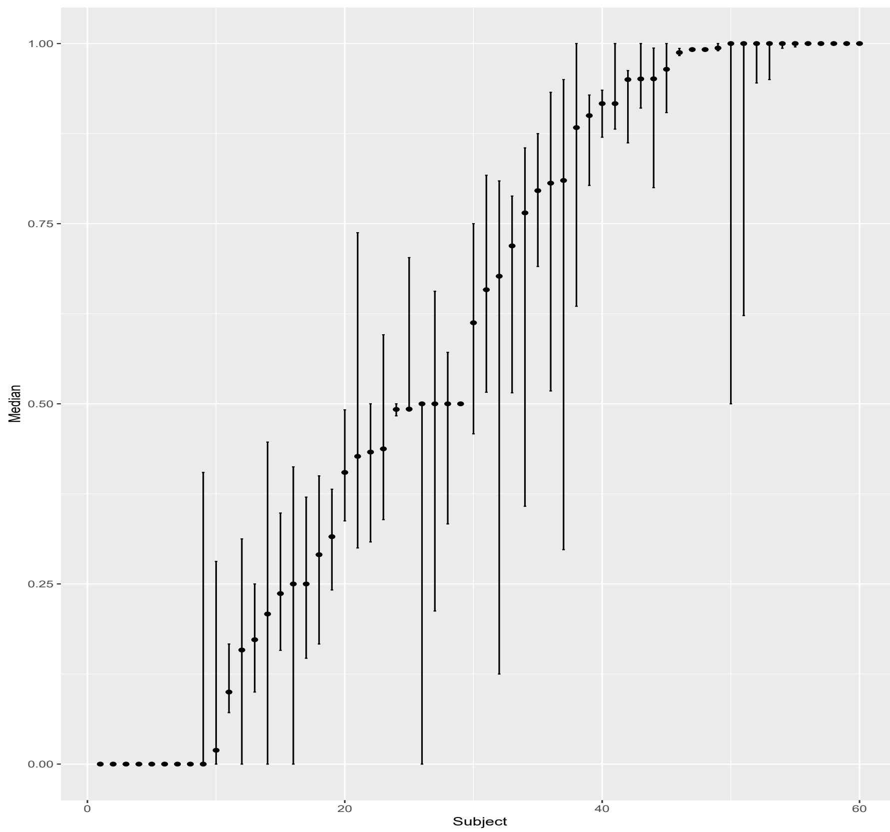
Figure 2.1: Frequency of normalized claims



Claims of the lower bound $\underline{x}$ are exactly consistent with Nash equilibrium play – though we cannot distinguish whether a subject is a $L0$ who follows a heuristic of choosing the lower bound or a subject plays the Nash equilibrium strategy. The remaining data are difficult to reconcile with leading behavioral models of limited strategic reasoning. This is because level k and NI k models predict that when $k \geq 1$, subjects will undercut below a deterministic model of non-strategic (i.e. $L0/NI0$) play – which would lead to spikes in the claim distribution *below*, rather than on focal claims like the top and the middle. Similarly, neither model predicts claims exactly at focal points under a model of uniformly random $L0/NI0$. By the same logic, QRE would always predict that a person's modal claim will be strictly below $\bar{x}$, and also predicts a smooth distribution of responses without isolated spikes at focal points. These models are thus inconsistent with our observation that the top and middle claims are frequently made. This observation suggests that some of our subjects may follow a heuristic of choosing a focal point instead of being strategic (in the sense of undercutting). We will further analyze this behavior at the subject level.

Individual behavior. Figure 2.2 plots median normalized claims across all rounds by subject, in increasing order of the median, and their corresponding 25^{th} and 75^{th} percentiles. Of the 60 subjects, there are 11, 4 and 9 subjects whose medians are exactly at the three spikes - upper bound, middle, and lower bound, respectively. We also notice the variation of choices within subject, a significant portion of subjects seem to have claims spread out across the normalized range.

Figure 2.2: Median normalized claims in increasing order, by subject



To understand individual behavior, we classify each subject to a best-fitting model as follows. For each subject, first we estimate each of fifteen theoretical models considered separately by maximum likelihood. We estimate twelve strategic models based on Nash equilibrium (2 models), QRE (1 model), level k (6 models), and NI (3 models). For the level k and NI models, we separately estimate versions with middle, top, and uniform $L0/NI0$ specifications, and we model a player's individual level of reasoning k as being drawn from a truncated Poisson distribution with parameter τ. We estimate both an action tremble (ϵ) version that allows random and payoff-independent mistakes and a payoff tremble (γ) version that assumes that a person's best replies to their model-derived beliefs for both Nash equilibrium and level k models. This structural modeling approach is similar to that of Goeree, Louis, and Zhang (2018) except in our specification of $L0$. This modeling approach for level k and NI models allows for within-subject heterogeneity in level of reasoning across rounds.

Let θ be the set of parameters of interest. For example, in a Nash model with action trembles, $\theta = \{\epsilon\}$, in a NI model, $\theta = \{\gamma, \tau\}$. An individual's likelihood function (we drop subscript i for individuals), given the observed choices x, in the set of game G is:

- For Nash and QRE: $L(\theta \mid x, G) = \prod_{g \in G} p(x \mid \theta, g)$
- For level k and NI: $L(\theta \mid x, G) = \sum_k f(k;\tau) \prod_{g \in G} p(x \mid \theta, g)$. The probability of a player being type k is $f(k;\tau) = \frac{(\frac{e^{-\tau} * \tau^k}{k!})}{\sum_l (\frac{e^{-\tau} * \tau^l}{l!})}$. $f(k;\tau)$ is a truncated Poisson distribution.

Here, $p(x \mid \theta, g)$ denotes the probability of claiming x in game g when the model under consideration has parameters θ. In addition, we consider three models of non-strategic play. First, uniform random play (which assigns the same likelihood to any choices), and two models of non-strategic play with action trembles: one where a subject plays the middle of the claim range and one where they play the top. We explain estimation details in Appendix B.3.

Among the fifteen models (including uniform randomization), for each subject, we pick the model with the highest log-likelihood and classify that subject to the best fitting of the above specifications whenever we can reject the null hypothesis of random or non-strategic behavior, according to the procedure we describe below. For each subject, we use a likelihood ratio test to compare the best-fitting strategic model to uniform randomization, and we use Vuong tests (Vuong 1989) to compare it to the other two models of non-strategic play. When the best-fitting strategic model is a significant improvement (at the 5% level) over the models of non-strategic play, we classify a subject to that strategic model. Otherwise, if either of the two non-strategic models is a significant improvement over uniform randomization, we classify them to the best fitting non-strategic model. In the event of a tie between strategic models, we classify the subject to the model with the fewest parameters. We detail this procedure in Appendix B.3.

Table 2.2 summarizes our subject classification.

Table 2.2: Subject classification

Nash	QRE	Strategic (Lk and NIk for $k \geq 1$)			Non-Strategic		
		Top $L0$	Middle $L0$	Uniform $L0$	Top	Middle	Uniform
16	1	11	8	1	14	8	1

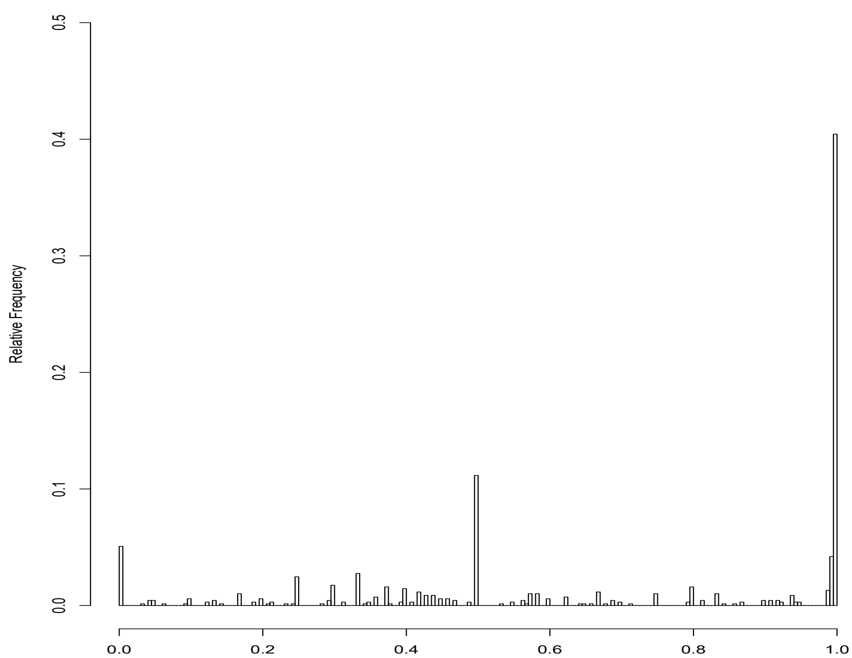
Finding 1: Non-strategic play is prevalent.

We classify a large proportion – 40% – of subjects as non-strategic (Table 2.2). Specifically, 23% tend to play the top and 13% tend to make the claim in the middle of the range. We classify 27% as playing the lower bound (i.e. Nash equilibrium) and 33% as (boundedly) strategic. This finding suggests that non-strategic play is prevalent and not just a fictitious anchor for strategic reasoning.

Finding 2: There is heterogeneity in non-strategic behavior, but almost no uniform randomization.

We observe heterogeneity in non-strategic play. The most common (58% of non-strategic subjects) is to claim at the upper bound of the range. Furthermore, we do not find evidence for the common assumption of non-strategic play as uniform randomization. Only one of our subjects is categorized into non-strategic heuristic of uniformly randomizing. At the aggregate level, the choice distribution of non-strategic subjects is different from a uniform distribution, as depicted in Figure 2.3 ($p < 0.01$, Uniform $(0, 1)$ Kolmogorov-Smirnov test).

Figure 2.3: Choice distribution of subjects classified as non-strategic



Finding 3: There is heterogeneity in how boundedly strategic players anchor their strategic reasoning.

Among subjects classified as strategic, 55% anchor their reasoning at the top of the strategy space, 40% anchor at the middle of the strategy space, and only 5% (1 subject) anchors their reasoning with uniform random play. The heterogeneity we observe suggests that in similar structural estimation exercises, imposing the same $L0$ assumption on all people may lead to a misidentification of strategic reasoning.

Effect of the reward/penalty parameter. In line with previous studies, we find a negative correlation between the reward/penalty parameter and average choices. Average normalized choices in Block 1, Block 2 and Block 3 are 0.595, 0.482 and 0.624, respectively; subject average normalized claims are similar in Block 1 and Block 3 (paired Wilcoxon test, $p = 0.15$). That is, we find a moderate effect of varying the reward/penalty parameter on claims (paired Wilcoxon test for Block 1 and Block 2, $p < 0.01$ in favor of the alternative "greater" for Block 1, also $p < 0.01$ for Block 3 and Block 2.)[6], but we do not find evidence of learning between Block 1 and Block 3.

2.5 Discussion

By carefully varying the game parameters in the traveler's dilemma, we uncover how strategic players anchor their strategic reasoning and how non-strategic players play. We classify 27% of subjects as tending to play Nash equilibrium (or tending to choose the lower bound), 33% as strategic and 40% as non-strategic. Contrary to our initial expectations based on assumptions and results in the prior literature, we do not find much evidence for the common assumption of non-strategic play as uniform randomization and we do find substantial evidence for non-strategic play. In addition, we find substantial heterogeneity in how subjects anchor their strategic reasoning.

For experimental and empirical applications of level k or noisy introspection models, our results suggest that common practice of identifying levels of strategic reasoning conditional on a prespecified assumption about $L0$ that is common for all subjects may lead to misleading estimates whenever the $L0$ specification matters. Since the way that people anchor their strategic reasoning matters in many games, we thus believe that future work in experimental game theory on this topic will be important for understanding how real people play games and how to model it.

[6]When the reward/penalty parameter increases from 5 to 20, the changes in median average normalized claims are 0.02, 0.09, and 0.11 for subjects classified as Nash, strategic players, and non-strategic players, respectively. Although the changes are small, they suggest that the models may not be completely specified.

Chapter 3

Default-Setting and Default Bias: Does the Choice Architect Matter?

3.1 Introduction

In a variety of decisions ranging from choice of health plan or retirement contributions to shipping methods for online shopping, one option is pre-selected as the "default alternative" and thus will be selected unless the decision-maker actively changes to another option. While pre-selecting a default alternative does not in itself affect the set of alternatives available to the decision-maker, behavioural economists have documented that it can have a substantial impact on actual choices – biasing decision-makers towards the default (Samuelson and Zeckhauser 1988; Madrian and Shea 2001; Johnson and Goldstein 2003; Handel 2013; Ericson 2014).

This observation motivates firms and governments to intentionally set defaults – often with the aim of improving decisions. In principle and in practice, there are many different ways to select a default option from a choice set. However, in a setting where people's preferences are heterogeneous, any selected default option will be undesirable for some people. In such a setting, it is unclear whether and, if so, how people's response to defaults depends on the rule used to select the default option. It is also not obvious how people would want defaults to be set if they were given the choice.

This paper provides the first systematic study of how people both respond to and rank different rules for selecting a default option from a choice set in a setting where the desirability of choice alternatives is not objectively ranked. We consider four default-setting rules: (i) defaults set randomly ("Random"), (ii) defaults based on the decisions of others ("Social"), (iii) defaults selected by an expert ("Expert"), and (iv) defaults custom-selected for each person based on their past choices ("Custom"). These four default-setting rules are motivated from discussions of choice architecture in the behavioural economics literature (Thaler and Sunstein 2003, Johnson et al. 2012) and real-world examples of default-setting in practice.

To compare these default-setting rules in a setting where preferences are heterogeneous, we conduct an experiment in which each subject makes choices among risky lotteries both without a default option and with a default option selected according to each of the four different default-setting rules. Comparing choices under a default-setting rule versus choices without a default allows us to measure and test for the presence of default bias separately for each rule. We additionally use our experiment to compare the strength of default bias across different default-setting rules. We then ask each subject to rank these four default-setting rules and No Default, thereby observing their preferences over default-setting rules.

An important contribution of our paper is to provide generally applicable approaches to measure and test for absolute and comparative default bias based on "apples-to-apples" comparisons that control for subjective default quality at the individual level. Under standard economic preferences, people tend to choose options they prefer. Thus a decision-maker will tend to choose the default option more often under a the default-setting rule that tends to pick options they prefer. This confounds the measurement and comparison of default bias across default-setting rules that sometimes select different defaults. Our key conceptual contribution is to design econometric methods to control for this confound that require minimal assumptions. Our experiment is designed to allow us to apply these methods to compare default bias under the default-setting rules we study while controlling for differences in the quality of defaults they assign, which is subjectively determined by each person's preferences.

We find significant evidence of default bias under the Social, Expert, and Custom default-setting rules with slightly more such bias under the latter two rules. However, we find no noticeable default bias with Random defaults. Our data additionally confirms the importance of controlling for default quality, as subjects are substantially more likely to choose a default option that they had also chosen when there was no default. Subjects' rankings of default-setting rules reveals that they tend to prefer rules that set subjectively higher quality defaults, with a noticeable preference for Expert default-setting rule. However, 21% of subjects most prefer to choose with No Default, indicating a noticeable aversion to the presence of defaults for a notable minority of subjects.

Following our discussion of related literature, we present our approach for measuring and comparing default bias across default-setting rules while controlling for default quality at the individual level (Section 3.2), our experimental design and procedure (Section 3.3), our results (Section 3.4), and discussion (Section 3.5).

Related Literature

Our paper builds on the broad literature on choice architecture (Thaler and Sunstein 2003) and default bias (Samuelson and Zeckhauser 1988; Madrian and Shea 2001; Johnson and Goldstein 2003; Handel 2013; Ericson 2014).

A related line of experiments study the effects of default quality on subjects' propensity to choose the default option in settings where the quality of choice options can be objectively ranked and ought to be the same across all subjects. This literature consistently finds that when default quality is manipulated, subjects are more prone to choose the default option when the default-setting rule has a greater tendency to select higher quality defaults (Caplin and Martin 2017; de Haan and de Linde 2018; Altmann et al. 2019; Altmann et al. 2019).[1] In each of these papers, each choice alternative is a monetary payment, but is presented so that computing each monetary payment is costly (de Haan and de Linde 2018; Altmann et al. 2019) or requires a probabilistic inference about the default setter's information (Altmann et al. 2019). Unlike in these papers, in most settings of interest (as in our experiment), preferences are subjective, heterogeneous, and not directly observed, thus the decision to follow or abandon a default takes on another dimension. Our study measures and compares default bias and decision quality in and across default-setting rules while controlling for subjective preferences.

In a related line of work, Arad and Rubinstein (2018) study people's attitudes to soft interventions that have been proposed in the behavioural economics literature, and find that a substantial fraction of respondents are averse to government-mandated default savings rates and other interventions. However, they only study attitudes and do not elicit actual choice behaviour.

A large literature in economics studies choice with a reference point (e.g. Kahneman 1979, 1991) or a status quo (e.g. Masatlioglu and Ok 2005, 2014). However, this literature generally assumes that the impact of the reference point or status quo on choice does not depend on how it was set.

3.2 Defining and Comparing Default Bias across Default-setting Rules

We study decision-makers who make choices with and without default options under different default-setting regimes. Our conceptual framework extends the existing work on default bias in choice (e.g. Masatlioglu and Ok 2005) to allow choice to depend on the rule used to select the default from the choice set, which we assume is known by the decision-maker.

Formally, let X be the set of all possible options, let $\mathcal{A}$ denote the set of all choice sets, which are non-empty subsets of X. Let $\mathcal{T}$ denote the set of functions, called default-setting

[1] Altmann et al. (2019) study how the presence of a background tasks affects choice of the default option, and find that subjects tend to choose the default option more often when faced a more difficult background task. Altmann et al. (2019) study a game where the default is set by a partially informed player whose incentives are either aligned, misaligned, or partially aligned with the partially informed decision-maker, and find that the default was chosen more often when incentives were more aligned. de Haan and de Linde (2018) study how the quality of defaults in earlier decisions affects subsequent default bias and find that it does.

rules, that select a probability distribution over defaults (or no default) for each set in $\mathcal{A}$; $\emptyset$ denotes "no default". A triple $(A, d, T) \in \mathcal{A} \times X \cup \{\emptyset\} \times \mathcal{T}$ defines a choice problem whenever d is in the support of $T(A)$.

In principle, observable behaviour can be described by a random choice function $p : X \times \mathcal{A} \times X \cup \{\emptyset\} \times \mathcal{T} \to [0, 1]$, where $p(x|A, d, T)$ denotes the probability that the decision-maker chooses x from choice set A when the default is d and was selected according to rule T.[2] Let $T = \text{ND}$ denote the regime that always assigns no default, $\emptyset$.

Next, we can define default bias as a higher probability of choosing x from A when x is the default than when A is faced with no default. Note that in our definition, default bias is evaluated under specific default-setting rules – a decision-maker may exhibit default bias under one rule but not another.

Definition. p exhibits default bias under rule T if $p(x|A, x, T) \geq p(x|A, \emptyset, ND)$ for every $A \in \mathcal{A}$ and x in the support of $T(A)$, with strict inequality for at least one such choice problem.

The definition of default bias compares the choice of defaults under rule T to the choice of the same options when there is no default. We may also wish to compare the strength of default bias in different default-setting regimes to each other. We formally define how to properly make such a comparison below, adapting the main idea from our definition of default bias to control for default quality by only comparing rules when they prescribe the same defaults.

Definition. p exhibits a stronger default bias under default-setting rule T than under T' if $p(x|A, x, T) \geq p(x|A, x, T')$, with strict inequality for at least one such choice problem, for all choice problems (A, x) such that x is the default for A under both T and T'.

Statistical tests of absolute and comparative default bias

In most settings, including our experiment, we only observe a finite number of choices per person and not the entire random choice function for each person. Thus we seek methods of testing for absolute and comparative default bias that allow us to aggregate data from all subjects while still carefully controlling for heterogeneity.

Our definitions of absolute and comparative default bias are each based on the comparisons of pairs of choice problems where each paired comparison controls for heterogeneity across individual and choice sets. Testing the null hypothesis of "no default bias under T" is equivalent to testing that $p(x|A, x, T) = p(x|A, \emptyset, ND)$ for every $A \in \mathcal{A}$ and x in the support of $T(A)$. Similarly, testing the null hypothesis of "equal default bias under T and

[2] There exists substantial evidence that behaviour has a substantial random element, even when studied at the individual level (e.g. Hey 1995).

under T'" is equivalent to testing that $p(x|A, x, T) = p(x|A, x, T')$ for every $A \in \mathcal{A}$ and every x that is in the support of both $T(A)$ and $T'(A)$. Each test is thus based on paired observations that each consist of a pair of two choices by the same person from the same choice set. For testing absolute default bias, each pair consists of an observed choice under a default-setting rule as compared to a choice made with no default. For comparative default bias tests, each pair consists of two observed choices, each with the same default, but where the default was selected under different default-setting rules in each case.

With data from many different choice sets from a single individual, or from one choice set each for many individuals, we could apply a McNemar's test to non-parametrically test the null hypothesis without having to estimate an entire random choice function. When we observe many different choice sets for each individual under both T and ND, we use an Obuchowski test (Obuchowski 1998) to aggregate across individuals and choice sets. Like a McNemar's test, the Obuchowski non-parametric test uses paired data to compare the estimated proportions with which the default is chosen. However, the Obuchowski test accounts for both intra- and inter- subject correlations to account for the fact that any pairs of observations from the same subject cannot be viewed as independent. This adjustment is analogous to the use of clustered standard errors in a panel regression.

3.3 Experimental Design

Our individual choice experiment consists of 84 rounds of lottery choice tasks with monetary outcomes and with no feedback between decisions (Table 3.1). We use lotteries as simple-to-implement choice objects where values are subjective and based on the experimental literature on decision-making under risk we expect a substantial variation in tastes between individuals (e.g. Hey and Orme 1994).

The experiment involves choices from 24 unique choice sets comprised of five lotteries in each. In the first 24 rounds – the No Default treatment – each subject makes a choice from each choice set without any option being designated as the default option (Figure 3.1A). In the following rounds, subjects proceed through the four default-setting rules (Random, Social, Expert, and Custom), completing 12 rounds for each.[3]

Default-setting rules

The description of each of the four default-setting rules to subjects is given in Table 3.1. We set these defaults according to the following procedures.

Random. A lottery was randomly selected from each choice set.

[3]To address potential concerns about treatment order effects, we varied the order of three treatments with defaults - Social, Expert, and Custom, at the session level. Subjects went through No Default, Random, one of the six treatment orders, then finally Choice of Default. We also varied starting choice sets and the order of choice sets in each treatment to control for order effects.

Table 3.1: Summary of treatments

Treatment	Description	Timeline
No Default	“In each of the next 24 decisions, no option will be selected as the default option.”	Round 1 - 24
Random	“In each of the next 12 decisions, one option will be selected as the default option. The default was selected **randomly** from the available lotteries.”	Round 25 - 36
Expert	“In each of the next 12 decisions, one option will be selected as the default option. The default was selected **by an expert** from the available lotteries.”	Round 37 - 48
Social	“In each of the next 12 decisions, one option will be selected as the default option. The default is the option that was most often selected by **a group of previous participants**.”	or Round 49 - 60,
Custom	“In each of the next 12 decisions, one option will be selected as the default option. The default was **custom-selected** for you based on your past choices.”	or Round 61 - 72 depending on session
Choice of Default	Default regime ranking	Round 73 - 84

“Description” describes how defaults are set in default-setting “Treatment”, exactly as worded to subjects at the beginning of each treatment.

Figure 3.1: Choice screens for No Default and Custom

(A) Choice screen for No Default

(B) Choice screen for Custom

Social. We ran a pilot experiment with nine participants in October 2019 with the same 24 choice sets. We set their modal choice in No Default as the default lottery for the Social default-setting rule.

Expert. We use an expected utility model with constant relative risk averse utility-for-income function $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$ for $\gamma = \frac{3}{4}$ to select the lottery with the highest expected utility as the default for each choice set.[4]

Custom. We coarsely scored each subject's risk aversion based on their No Default choices in the three Eckel and Grossman (2002) style choice problems with equal likelihood of both outcomes for each lottery. In each of these choice sets, lotteries were scored from 1 (safest) to 5 (riskiest) and we added these scores to obtain a final score S between 5 (i.e. always choosing the safest option) and 15 (i.e. always choosing the riskiest option). Based on the score S, we assigned each subject to one of three groups, and assigned a different constant relative risk aversion parameter γ to each group that would generate a score in that range. Specifically, we assigned $\gamma = 2, 1.25$, and 0.5 respectively for the cases $S \leq 6$, $7 \leq S \leq 9$, and $S \geq 10$. Then, for each choice set, the expected utility maximizing lottery was selected as the default option.

Prior to making choices under a default-setting rule, the rule was described to subjects on a screen (as in Table 3.1); it was also described on each waiting screen between successive choices and on the top of each decision screen (e.g. see Figure 3.1B). In each of these rounds, one available lottery is set by the default-setting rule as the default lottery, and is prominently displayed at the top of the screen and appears pre-selected (Figure 3.1B).

We chose our default-setting rules to mimic real-world default-setting rules, albeit subjects had exactly the information about each described in Table 3.1. Random defaults is a useful theoretical benchmark since defaults will be transparently uncorrelated with individual preferences. Random defaults are used in past experiments like Samuelson and Zeckhauser (1988) and in the United States of America for assigning a default health insurance plan under the Medicare Part D program (Ericson 2014). More broadly, we view it as a benchmark for real-world cases where a historically-set default is unlikely to be correlated with a person's preferences. Social defaults are one possibility suggested by Thaler and Sunstein (2003). The discussion of choice architecture tends to views defaults as something that experts can intentionally set. Our design intentionally provides no information on how the expert chooses defaults – which we view as consistent with real-world examples. For example, the default allocation for pension plan contributions at Simon Fraser University is to a "balanced" fund designed by the pension fund trustees and pension administrator to be a good and balanced option for a large number of plan members. Customized defaults

[4]Expected utility is a normatively appealing model, and expected utility with constant-relative risk aversion is widely used by experimental economists to describe choices involving risks. This is the particular parameter value found by Zrill (2020) as best describing the median subject in an portfolio choice experiment and we thus felt that this was a good default for most subjects.

have been suggested as one way to improve choices (Smith et al. 2013), including health plan choices (Zhang et al. 2015); "sensory defaults" on a website based on cookies or other information are an example of where custom defaults are used in practice (Johnson et al. 2012).

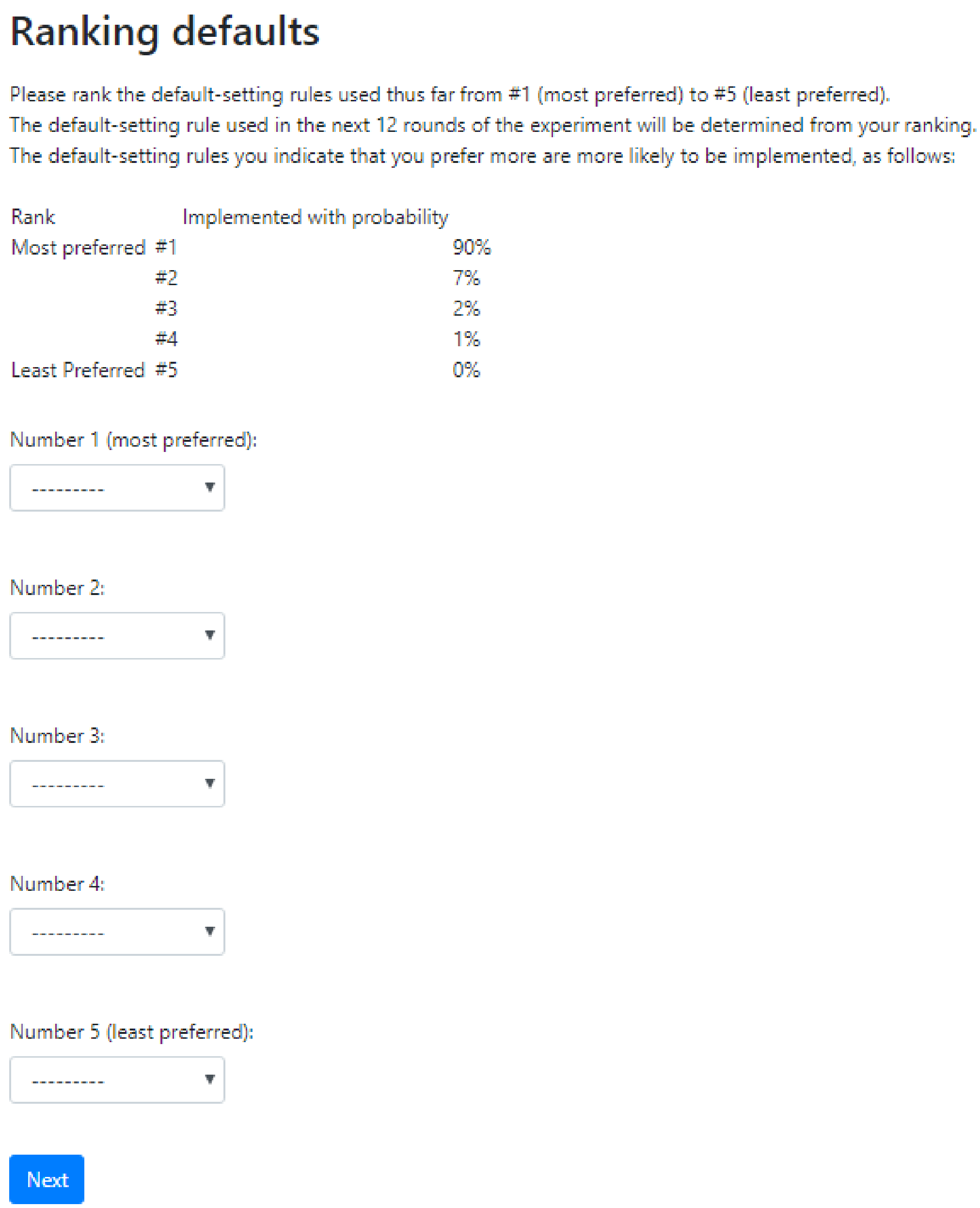
Choice sets

We constructed 24 choice sets of five lotteries each (Appendix C.3). Choice sets qualitatively varied. 6 choice sets consisted of five two-outcome lotteries with the same probabilities of the higher and lower outcome for all lotteries (as in Eckel and Grossman (2002)). 12 choice sets consisted of five one-to-three outcome lotteries with common support. 6 choice sets consisted of five lotteries where all but at most one had support on three or four outcomes. This mix of qualitatively different choice sets that varied in choice complexity precluded construction of simple common heuristics.

Flow

Each subject first faced all 24 choice sets in the No Default treatment. Then they faced each choice set again under two of the four default-setting rules, facing 12 choice sets per rule. Choice sets were arranged so that there were exactly four choice sets common to any two default-setting rules. After these first 72 rounds, each subject was asked to rank the five default-setting rules they faced (including No Default) from most (1) to least (5) preferred (Figure 3.2). They were informed that the default-setting rule for the next 12 rounds would be selected based on their ranking, with rule they ranked #1-4 respectively has a 90%, 7%, 2%, and 1% likelihood of being selected. 12 choice sets were repeated again in the last 12 rounds where the default-setting rule was selected based on the subject's ranking.

Figure 3.2: Preference elicitation of default-setting rules



Procedures

In November 2019 and February-March 2020, we recruited 113 subjects in 22 sessions from the SFU Experimental Economics Lab participation pool.[5] Each session took place in the

[5] We had aimed to recruit 144-196 subjects over 22 sessions. We began to early conduct additional sessions in late February-March 2020 to raise our sample size to that range. However, we only were able to conduct

Lab and lasted approximately 65 minutes. The experiment was conducted using a computerized interface. One round was randomly selected and the subject's chosen lottery in that round was played out to determine their payment. The average payment was a $27.40, including a $7 participation payment. Full details of the experimental procedure including instructions and screenshots are provided in Appendix C.1.

3.4 Results

The frequency with which the defaults assigned by a rule are chosen in the No Default treatment provides us with an aggregate measure of subjective default quality. The default options assigned by the Random, Social, Expert and Custom rules are respectively chosen 21%, 39%, 40%, and 41% of the time in No Default, for an average rate of 35%. That is, on average, subjects tended to choose our treatments' default option more frequently than by chance when the choice was faced in No Default. That is, the quality of default options tends to be better than that of randomly-selected options. Thus, if people tend to choose options that they like, we cannot use the frequency of choosing the default in a given treatment as a measure of default bias unless we control for subjective default quality. Our first result shows this concern is empirically relevant in our setting.

three such sessions before SFU shut down due to the novel coronavirus outbreak. We may conduct additional sessions using the same protocols when SFU reopens to raise our sample size.

Result 1: People are more likely to choose a default option if they chose it without a default.

Figure 3.3: Frequency of choosing the default lottery conditional on whether having chosen the same lottery in No Default

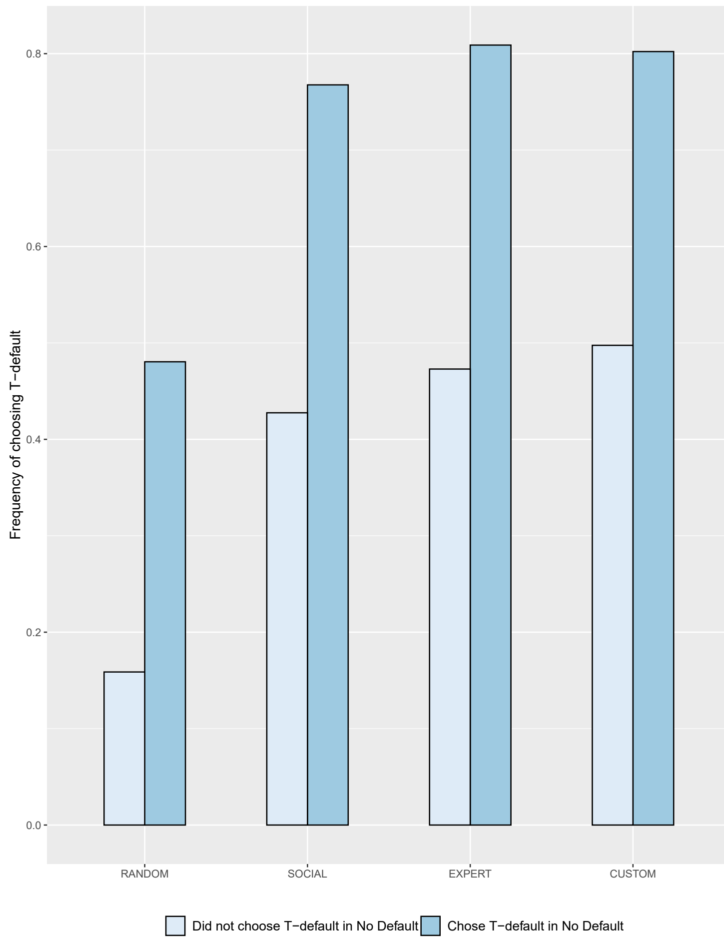


Figure 3.3 presents the frequency of choosing the default lottery conditional on whether the same lottery was previously chosen in No Default, by treatment. Out of all occasions where the subject had chosen a default-setting rule's default when facing the same choice set in No Default, they then chose that option when it was the default option 75% of the time, though this number was only 45% in the Random treatment but 77%, 81%, and 80% in the Social, Expert, and Custom treatments, respectively. In contrast, subjects chose the default lottery in only 37% of occasions when they had not chosen that lottery in No Default – 16%, 43%, 47%, and 50% of the time in Random, Social, Expert, and Custom, respectively.

Choices in experimental settings comparable to our No Default treatment are typically interpreted as noisily revealing a subject's most preferred lottery in each choice set. Viewed this way, Result 1 states that people are more likely to choose the default lotteries that they prefer. This establishes the empirical relevance of controlling for subjective default quality when measuring and comparing default bias across default-setting rules.

Result 2: Default bias is significant in the Social, Expert, and Custom default-setting rules, but not with Random defaults.

Figure 3.4: Frequency of choosing the default lottery in each default-setting rule versus in No Default

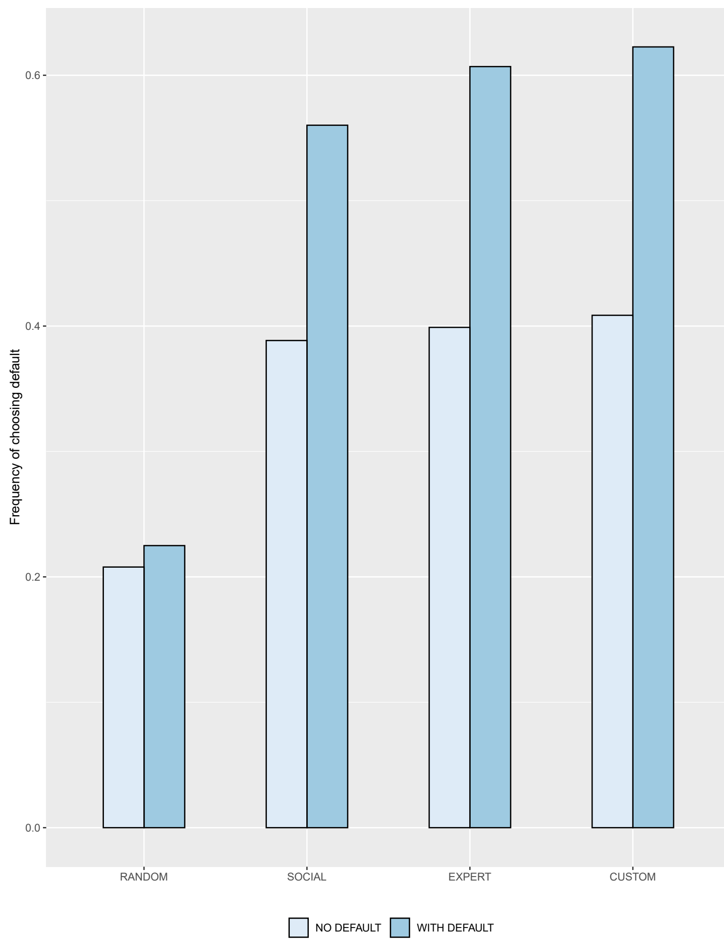


Figure 3.4 compares the frequency at which each default-setting rule's default lottery was chosen in that treatment to the case when the same choice set was faced with in No Default. If there were no default bias under a given default-setting rule, its dark-blue bar would be the same height as its light-blue bar. We only observe this equality for Random. In each of Social, Expert, and Custom, the rule's assigned default lottery was chosen more frequently when it was the default than in No Default – indicating default-bias in each of these default-setting rules.

Table 3.2: Obuchowski tests for default bias

	Random	Social	Expert	Custom
p-value	0.213	< 0.01	< 0.01	< 0.01

p-values for an Obuchowski test for default bias

Each test uses 1356 choices from 113 subjects.

We apply an Obuchowski test of absolute default bias to assess statistical significance while controlling for subjective default quality (Table 3.2). Default bias is insignificant with Random defaults ($p = 0.213$), but significant under the Social, Expert, and Custom default-setting rules ($p < 0.01$ for each of the three tests).

Result 3: Default bias is stronger with intentionally-set defaults than with randomly-set defaults.

While we find evidence of default bias under the Social, Expert, and Custom rules, but not under the Random rule, this does not tell us how the strength of default bias differs across these rules. Even for the comparison between the Random rule and the remaining rules, Result 1 notes significant differences in the quality of defaults between the Random rule and the rules with intentionally-set defaults – Social, Expert, and Custom – thus necessitating the use of our comparative Obuchowski test to compare the strength of default bias while controlling for default quality.

We find that default bias is significantly stronger under each of the Social, Expert, and Custom rules than under Random defaults ($p = 0.003$, 0.03, 0.04, respectively) when applying our test of comparative default bias in each case. When comparing default bias under the Social, Expert, and Custom rules, we find no detectable differences in the strength of default bias between any pair of these rules ($p = 0.07$ for Social vs. Expert, $p = 0.51$ for Social vs. Custom, and $p = 0.83$ for Expert vs. Custom). These results indicate that default bias is stronger when defaults are intentionally set, as in the Social, Expert, and Custom rules, than when defaults are randomly selected.

Result 4: Subjects tended to rank Expert $\succ$ Custom $\succ$ Social $\succ$ No Default $\succ$ Random. A notable minority ranked the No Default rule as most-preferred. The ranking reflects subjects' tendency to prefer rules that set defaults with higher subjective quality.

Table 3.3: Ranking of default-setting rules

Default type	#1	#2	#3	#4	#5
No Default	0.212	0.115	0.115	0.221	0.336
Random	0.044	0.080	0.133	0.345	0.398
Expert	0.425	0.301	0.168	0.044	0.062
Social	0.088	0.230	0.327	0.248	0.106
Custom	0.230	0.274	0.257	0.142	0.097

$n = 113$ subjects

Table 3.3 counts the fraction of subjects who ranked each rule in each position. We construct a Borda count for each default-setting rule and obtain the aggregate ranking Expert $\succ$ Custom $\succ$ Social $\succ$ No Default $\succ$ Random. The overall ranking remains the same for the subject group with a high level of risk aversion and for the subject group with a moderate level of risk aversion, as detailted in Appendix C.5. For the group with a low level of risk aversion, the ranking from a Borda count is Expert $\succ$ Custom $\succ$ No Default $\succ$ Social $\succ$ Random, with a switch in ranking between No Default and Social. We find that 45% of subjects divided their top three ranks among the three informative default-setting rules (Social, Expert, and Custom). Expert was the most preferred default-setting rule for 42.5% of subjects and was the most commonly first-ranked rule; 23% of subjects ranked the Custom rule first, while the Social default-setting rule was most preferred by only 9% of subjects. Random defaults and No Default were the two least favoured rules and were ranked last by 40% and 34% of subjects, respectively. However, a noticeable proportion of our subjects exhibit some preference for choosing without defaults – 21% rank No Default as their first choice, and 44% of subjects put No Default in their top three rules.

To measure how preferences over default-setting rules relate to measures of default quality, we run a rank-ordered logit model with a subject's rank of a default-setting rule on the left hand side and a default quality measure on the right hand side:

$$\text{rank}_{iT} = \beta \text{defaultquality}_{iT} + \epsilon_{iT}$$

where i denotes subject i, $T \in \{\text{Random}, \text{Social}, \text{Expert}, \text{Custom}\}$[6] denotes a default-setting rule, and $\text{defaultquality}_{iT}$ is the number of times (out of 12) the T default option was chosen when the same option was faced with in No Default. We estimate $\beta = 0.208$ (s.e. $= 0.038$, $p < 0.001$). This indicates a significant association between a subject's ranking of a rule and the quality of that rule's defaults for that subject.

Result 5: The strength of subjects' default bias is unaffected by their act of ranking default-setting regimes.

It might be the case that the mere act of choosing a default-setting rule affects a person's degree of default bias – a possibility we now test. After subjects had ranked the five default-setting rules, one was implemented – and their first-choice rule was implemented with a 90% chance. Among subjects whose first choice was implemented (excluding those whose first choice was No Default), we find no significant difference when we compare their default bias in their top-ranked rule before as compared to after they ranked it, using an Obuchowski test of comparative default bias ($p = 0.48$). Thus we conclude that a subject's mere act of choosing to rank a default-setting regime highly does not lead to any detectable strengthening of their default bias.

3.5 Discussion

We study the effectiveness of different default-setting rules in a setting where preferences are heterogeneous, and there were no objective ranking of options. Our experimental design and statistical approach allow us to disentangle the subjective quality of defaults that a rule assigns from the amount of default bias that that same rule induces. We find a significant increase in the probability of choosing default options in the Social, Expert, and Custom treatments, but not in the Random treatment. Our experimental results indicate that the mere presence of a default option is not enough to affect choice, but intentionally set defaults can induce a significant amount of default bias. Our results thus suggest that policymakers need to set defaults that are, on average, good for most people and to communicate this if they wish to harness default bias to nudge decisions in a particular direction.

We find that most participants prefer such intentional default-setting rules (Expert, Custom, and Social) to choosing without a default or to randomly-set defaults – providing the first extant evidence for a preference to be nudged. Our analysis shows that subjects tended to assign rankings of default-setting rules consistent with both the subjective default quality and their strength of default bias.

In our experimental design, the three rules with intentionally-set defaults tended to select options with similar quality, measured by concordance of defaults with choices in

[6] We drop choices under the No Default rule in this specification.

the No Default treatment. The similar strength of default bias across the three rules reflects a similar willingness to follow defaults according to the rules as described – without any evidenced aversion to any of them. However, we suspect that in other settings where decision-makers make a large number of repeated decisions such as in de Haan and de Linde (2018), the actual quality of the defaults prescribed by a rule will eventually matter more than the initial framing of the default.

Bibliography

Agranov, M., A. Caplin, and C. Tergiman (2015): "Naive play and the process of choice in guessing games," *Journal of the Economic Science Association*, 1(2), 146–157.

Alekseev, A., G. Charness, and U. Gneezy (2017): "Experimental methods: When and why contextual instructions are important," *Journal of Economic Behavior & Organization*, 134, 48–59.

Altmann, S., A. Falk, and A. Grunewald (2019): "Incentives and information as driving forces of default effects," Working Paper.

Altmann, S., A. Falk, A. Grunewald, and D. Huffman (2014): "Contractual incompleteness, unemployment, and labour market segmentation," *The Review of Economic Studies*, 81(1), 30–56.

Altmann, S., A. Grunewald, and J. Radbruch (2019): "Cognitive Foundations of Passive Choices," Working Paper.

Anderson, L. R., F. J. DiTraglia, and J. R. Gerlach (2011): "Measuring altruism in a public goods experiment: a comparison of US and Czech subjects," *Experimental Economics*, 14(3), 426–437.

Arad, A., and A. Rubinstein (2012): "The 11-20 money request game: A level-k reasoning study," *American Economic Review*, 102(7), 3561–73.

——— (2018): "The people's perspective on libertarian-paternalistic policies," *Journal of Law and Economics*, 61(2), 311–333.

Aycinena, D., R. Baltaduonis, and L. Rentschler (2014): "Valuation structure in first-price and least-revenue auctions: an experimental investigation," *Experimental Economics*, 17(1), 100–128.

Bardsley, N., J. Mehta, C. Starmer, and R. Sugden (2010): "Explaining focal points: cognitive hierarchy theory versus team reasoning," *Economic Journal*, 120(543), 40–79.

Basu, K. (1994): "The traveler's dilemma: Paradoxes of rationality in game theory," *American Economic Review*, 84(2), 391–395.

Bayer, R.-C., E. Renner, and R. Sausgruber (2013): "Confusion and learning in the voluntary contributions game," *Experimental Economics*, 16(4), 478–496.

Bigoni, M., and D. Dragone (2012): "Effective and efficient experimental instructions," *Economics Letters*, 117(2), 460–463.

Borchers, H. W. (2018): *pracma: Practical Numerical Math Functions*R package version 2.1.8.

Brookins, P., and D. Ryvkin (2014): "An experimental study of bidding in contests of incomplete information," *Experimental Economics*, 17(2), 245–261.

Burchardi, K. B., and S. P. Penczynski (2014): "Out of your mind: Eliciting individual reasoning in one shot games," *Games and Economic Behavior*, 84, 39–57.

Cabrera, S., E. Fatás, J. A. Lacomba, and T. Neugebauer (2013): "Splitting leagues: promotion and demotion in contribution-based regrouping experiments," *Experimental Economics*, 16(3), 426–441.

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004): "A cognitive hierarchy model of games," *Quarterly Journal of Economics*, 119(3), 861–898.

Caplin, A., and D. Martin (2017): "Defaults and attention: the drop out effect," *Revue Économique*, 68(5), 747–755.

Capra, C. M., J. K. Goeree, R. Gomez, and C. A. Holt (1999): "Anomalous behavior in a traveler's dilemma?," *American Economic Review*, 89(3), 678–690.

Cassar, A., and D. Friedman (2004): *Economics lab: an intensive course in experimental economics*. Routledge.

Champely, S. (2018): *pwr: Basic Functions for Power Analysis*R package version 1.2-2.

Chen, D. L., M. Schonger, and C. Wickens (2016): "oTree–An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Chen, R., Y. Chen, and Y. Riyanto (2018): "Public knowledge in coordination games: Learning from non-replication," .

Converse, J., and S. Presser (1986): *Survey Questions: Handcrafting the Standadized Questionnaire*. Sage.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001): "Cognition and behavior in normal-form games: An experimental study," *Econometrica*, 69(5), 1193–1235.

Costa-Gomes, M. A., and V. P. Crawford (2006): "Cognition and behavior in two-person guessing games: An experimental study," *American Economic Review*, 96(5), 1737–1768.

Cox, J. C., and D. James (2012): "Clocks and trees: Isomorphic Dutch auctions and centipede games," *Econometrica*, 80(2), 883–903.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013): "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications," *Journal of Economic Literature*, 51(1), 5–62.

Crawford, V. P., and N. Iriberri (2007): "Fatal attraction: Salience, naivete, and sophistication in experimental "hide-and-seek" games," *American Economic Review*, 97(5), 1731–1750.

Davis, D. D., and C. A. Holt (1993): *Experimental Economics*. Princeton University Press.

de Haan, T., and J. de Linde (2018): "'Good Nudge Lullaby': Choice Architecture and Default Bias Reinforcement," *Economic Journal*, 128(610), 1180–1206.

Eckel, C. C., and P. J. Grossman (2002): "Sex differences and statistical stereotyping in attitudes toward financial risk," *Evolution and Human Behavior*, 23(4), 281–295.

Ericson, K. M. (2014): "Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange," *American Economic Journal: Economic Policy*, 6(1), 38–64.

Ericson, K. M. M., and A. Fuster (2011): "Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments," *Quarterly Journal of Economics*, 126(4), 1879–1907.

Etang, A., D. Fielding, and S. Knowles (2011): "Does trust extend beyond the village? Experimental trust and social distance in Cameroon," *Experimental Economics*, 14(1), 15–35.

Fouraker, L. E., and S. Siegel (1963): *Bargaining Behavior*. McGraw-Hill New York.

Fragiadakis, D. E., D. T. Knoepfle, and M. Niederle (2016): "Who is strategic?," Working Paper.

Fragiadakis, D. E., A. Kovaliukaite, and D. R. Arjona (2019): "Belief-Formation in Games of Initial Play: an Experimental Investigation," .

Friedman, D., and S. Sunder (1994): *Experimental methods: A primer for economists*. Cambridge University Press.

Gill, D., and V. Prowse (2012): "A structural analysis of disappointment aversion in a real effort competition," *American Economic Review*, 102(1), 469–503.

Goeree, J., C. A. Holt, and T. R. Palfrey (2016): *Quantal Response Equilibrium-a Stochastic Theory of Games*. Princeton University Press.

Goeree, J. K., and C. A. Holt (2004): "A model of noisy introspection," *Games and Economic Behavior*, 46(2), 365–382.

Goeree, J. K., P. Louis, and J. Zhang (2018): "Noisy introspection in the 11–20 game," *Economic Journal*, 128(611), 1509–1530.

Gopstein, D. (2018): *clust.bin.pair: Statistical Methods for Analyzing Clustered Matched Pair Data*R package version 0.1.2.

Handel, B. R. (2013): "Adverse selection and inertia in health insurance markets: When nudging hurts," *American Economic Review*, 103(7), 2643–82.

Hargreaves-Heap, S., D. Rojo Arjona, and R. Sugden (2014): "How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games With Non-Neutral Frames," *Econometrica*, 82(3), 1133–1151.

Harris, D., B. Herrmann, A. Kontoleon, and J. Newton (2015): "Is it a norm to favour your own group?," *Experimental Economics*, 18(3), 491–521.

Hart, A., and S. Martínez (2018): *spgs: Statistical Patterns in Genomic Sequences*R package version 1.0-2.

Henningsen, A., and O. Toomet (2011): "maxLik: A package for maximum likelihood estimation in R," *Computational Statistics*, 26(3), 443–458.

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

Hey, J. D. (1995): "Experimental investigations of errors in decision making under risk," *European Economic Review*, 39(3-4), 633–640.

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

Imai, K., L. Keele, and T. Yamamoto (2010): "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25(1), 51–71.

Johnson, E. J., and D. Goldstein (2003): "Do defaults save lives?," *Science*, 302(5649), 1338–1339.

Johnson, E. J., S. B. Shu, B. G. Dellaert, C. Fox, D. G. Goldstein, G. Häubl, R. P. Larrick, J. W. Payne, E. Peters, D. Schkade, B. Wansink, and E. Weber (2012): "Beyond nudges: Tools of a choice architecture," *Marketing Letters*, 23(2), 487–504.

Kahneman, D. (1979): "Tversky a.(1979)," *Prospect theory: an analysis of decision under risk*, pp. 263–292.

Kamei, K., L. Putterman, and J.-R. Tyran (2015): "State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods," *Experimental Economics*, 18(1), 38–65.

Kneeland, T. (2015): "Identifying Higher-Order Rationality," *Econometrica*, 83(5), 2065–2079.

Madrian, B. C., and D. F. Shea (2001): "The power of suggestion: Inertia in 401 (k) participation and savings behavior," *Quarterly Journal of Economics*, 116(4), 1149–1187.

Masatlioglu, Y., and E. A. Ok (2005): "Rational choice with status quo bias," *Journal of Economic Theory*, 121(1), 1–29.

——— (2014): "A canonical model of choice with initial endowments," *Review of Economic Studies*, 81(2), 851–883.

McKelvey, R. D., and T. R. Palfrey (1995): "Quantal response equilibria for normal form games," *Games and Economic Behavior*, 10(1), 6–38.

Merkle, E., and D. You (2020): *nonnest2: Tests of Non-Nested Models*R package version 0.5-3.

Mittone, L., and M. Ploner (2011): "Peer pressure, social spillovers, and reciprocity: an experimental analysis," *Experimental Economics*, 14(2), 203–222.

Nagel, R. (1995): "Unraveling in guessing games: An experimental study," *American Economic Review*, 85(5), 1313–1326.

Noussair, C. N., and J. Stoop (2015): "Time as a medium of reward in three social preference experiments," *Experimental Economics*, 18(3), 442–456.

Obuchowski, N. A. (1998): "On the comparison of correlated proportions for clustered data," *Statistics in Medicine*, 17(13), 1495–1507.

Penczynski, S. P. (2016): "Strategic thinking: The influence of the game," *Journal of Economic Behavior & Organization*, 128, 72–84.

Petersen, L., and A. Winn (2014): "Does money illusion matter? comment," *American Economic Review*, 104(3), 1047–62.

R Core Team (2017): *R: A Language and Environment for Statistical Computing*R Foundation for Statistical Computing, Vienna, Austria.

Ramalingam, A., A. Morales, and J. Walker (2018): "Instruction Length and Content: Effects on Punishment Behaviour in Public Goods Games," *Journal of Behavioral and Experimental Economics*, 73, 66–73.

Rubinstein, A., A. Tversky, and D. Heller (1997): "Naive strategies in competitive games," in *Understanding strategic interaction*, ed. by W. Albers, W. Guth, P. Hammerstein, B. Moldovanu, and E. v. Damme, pp. 394–402. Springer.

Samuelson, W., and R. Zeckhauser (1988): "Status quo bias in decision making," *Journal of Risk and Uncertainty*, 1(1), 7–59.

Signorell, A. (2018): *DescTools: Tools for Descriptive Statistics*R package version 0.99.25.

Smith, N. C., D. G. Goldstein, and E. J. Johnson (2013): "Choice without awareness: Ethical and policy implications of defaults," *Journal of Public Policy & Marketing*, 32(2), 159–172.

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

Stahl, D. O., and P. W. Wilson (1994): "Experimental evidence on players' models of other players," *Journal of Economic Behavior & Organization*, 25(3), 309–327.

——— (1995): "On players' models of other players: Theory and experimental evidence," *Games and Economic Behavior*, 10(1), 218–254.

StataCorp. (2013): *Stata Statistical Software: Release 13*College Station, TX: StataCorp LP.

Stokely, M. (2015): *HistogramTools: Utility Functions for R Histograms*R package version 0.3.2.

Thaler, R. H., and C. R. Sunstein (2003): "Libertarian paternalism," *American Economic Review*, 93(2), 175–179.

Tingley, D., T. Yamamoto, K. Hirose, L. Keele, and K. Imai (2014): "mediation: R Package for Causal Mediation Analysis," *Journal of Statistical Software*, 59(5), 1–38.

Tversky, A., and D. Kahneman (1991): "Loss aversion in riskless choice: A reference-dependent model," *The quarterly journal of economics*, 106(4), 1039–1061.

Venables, W. N., and B. D. Ripley (2002): *Modern Applied Statistics with S*. Springer, New York, fourth edn., ISBN 0-387-95457-0.

Vuong, Q. H. (1989): "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, pp. 307–333.

Wickham, H. (2009): *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wright, J. R., and K. Leyton-Brown (2019): "Level-0 models for predicting human behavior in games," *Journal of Artificial Intelligence Research*, 64, 357–383.

Zeileis, A. (2004): "Econometric Computing with HC and HAC Covariance Matrix Estimators," *Journal of Statistical Software*, 11(10), 1–17.

——— (2006): "Object-Oriented Computation of Sandwich Estimators," *Journal of Statistical Software*, 16(9), 1–16.

Zeileis, A., and T. Hothorn (2002): "Diagnostic Checking in Regression Relationships," *R News*, 2(3), 7–10.

Zhang, Y., S. H. Baik, and J. P. Newhouse (2015): "Use of intelligent assignment to Medicare Part D plans for people with schizophrenia could produce substantial savings," *Health Affairs*, 34(3), 455–460.

Appendix A

Appendix for Chapter 1

A.1 Review of Current Practice

Inclusion/Exclusion criteria

We included experimental papers published between January 2011 and December 2016 in six journals: the American Economic Review, Econometrica, the Quarterly Journal of Economics, the Journal of Political Economy, the Review of Economic Studies, and Experimental Economics. Articles from the AER: Papers and Proceedings were excluded. In order to be included, a paper had to include at least one lab experiment. We excluded field experiments and online experiments that were not conducted in a controlled environment, but we include “lab-in-the-field” experiments that were conducted in a controlled environment.

To classify each included experiment, we reviewed both the text of each paper and supplementary materials available online through the journal’s website, with the exception of uncompiled code (e.g. z-Tree code).

Coding Criteria: Delivery

Delivery methods could include paper instructions or computer instructions. Values in the supplementary table are 1 for yes, 0 for no, 0.5 for uncertain. In some cases, an alternative delivery method was used; for example, Etang et al. (2011) studied subjects in rural Cameroon and used purely verbal instructions because many subjects were illiterate.

We code the study as having paper instructions if it is directly stated or clearly implied that a set of paper instructions were used. Some papers were explicit about their use of printed instructions, while others required us to infer the existence of paper instructions from other details. For instance, Mittone and Ploner (2011, p. 207) write that "after the choices are collected, instructions for the beliefs elicitation phase are distributed." Distribution implies a written set of instructions, though this is not explicitly stated. Sometimes we inferred the form of instructions from the instructions themselves, for instance in Altmann et al.

(2014), the instructions included screenshots, from which we inferred that they must have been printed on paper.

We code the study as having computer instructions if it is directly stated or clearly implied that computerized instructions were used. Sometimes this was explicit, while other times it had to be inferred. For instance, in papers that included copies of their instructions online, some instructions told participants to click on something to proceed to the next screen. This implies that the instructions are computerized, even if it is not explicitly stated in the text of that paper. Cox and James (2012, Supplement p. 2) end their instructions by telling their subjects, "When you have finished reading and have asked any questions you might have, please click Done."

Many papers are unclear on whether the instructions are given on paper or on computers. If there was no explicit statement of the form of instructions in the paper itself, and no clear indication from the instructions where these were available online, the paper was coded as uncertain.

Coding Criteria: Reinforcement

We coded four different forms of reinforcement.

1. Read aloud. We code an experiment as having read aloud its instructions if it is stated or clearly implied that the instructions were presented orally. Most often this meant that the experimenter read the instructions for the participants to hear. Some studies, such as Aycinena et al. (2014, p. 110), included voice recordings of the instructions, which we coded as read aloud as indicated by the following quote "They were provided with instructions and were also shown a video which read these instructions aloud."

2. Demonstration or guided practice. We code a paper as including demonstration or guided practice if we can infer that it used walk-throughs of the experimental interface, examples, or demonstrations of aspects of the experiment during the instructions phase. Walk-throughs involve actively-guided practice by the subject. Examples include hypothetical descriptions of potential actions and consequent outcomes. For instance, Brookins and Ryvkin (2014) give subjects an example of the likelihood of success, conditional on the group members' investment. Demonstrations actively highlight one or more aspects of the experiment, for example, throwing a die to show subjects how uncertainty will be resolved as in Ericson and Fuster (2011). The mere use of graphical or tabular methods to communicate information, or providing screenshots in paper instructions, was considered neither demonstration nor guided practice.

3. Unguided practice. If the experiment included one or more unpaid practice rounds without guidance, we coded this as unguided practice. Sometimes this was explicit in the body of the paper, while other times it was only indicated in the instructions themselves.

4. Quiz. Quizzes or questionnaires were only included if they occurred after the instructions and before the experiment. Many experiments include questionnaires to check participants' understanding ex post, but these are not counted as they do not reinforce participants' understanding of the instructions before the experiment.

When a quiz was given, we checked whether feedback was given after the quiz and before the experiment. If it was clearly stated that subjects were given the correct answers to the quiz, "Feedback" was coded as a 1. If subjects must get 100% to proceed with the experiment, we infer that feedback was given. Many papers give quizzes to "ensure comprehension of instructions" but do not explicitly indicate whether answers were given. For example Cabrera et al. (2013, p. 432) indicate that "subjects completed a quiz to make sure they had fully understood the logic of the game." It is ambiguous whether this implies that feedback was given to promote subject understanding ex-ante or instead quiz performance was used by the experimenters to assess subject comprehension ex-post. Such papers are coded as uncertain with respect to quiz feedback. We also separately code whether subjects were paid for correct quiz answers (Incentivized) and whether participants were required to get all questions correct before continuing (Require 100%).

Coding Criteria: Some main task(s) is (are) one shot

We classified the main task or tasks for each experiment. If at least one of the main tasks is one shot (that is, subject can be viewed as making a single decision) in one or more of the treatments, we coded that paper as having a one shot main task under this column. When researchers use a choice list or the strategy method – where multiple similar decisions are made almost simultaneously, and could in-principle be viewed as one decision – we view this task as a one-shot task. In contrast, when decisions are made in a sequence, even without feedback, we would not consider those to constitute a one-shot task. Anderson et al.'s (2011) study provides an edge case. In their experiment, each subject plays six public goods games with different parameter values, but all six choices are presented at the same time. Since all choices are instances of the same basic task and are presented at once, we coded their experiment as one shot. If these tasks had been presented sequentially on separate screens, we would not have coded this as one shot. An interesting boundary case is a dynamic game with an evolving state variable (e.g. the money supply variable in Petersen and Winn (2014)); subjects in such games make repeated decisions in the same task, but with different incentives depending on the state. We have coded these as repeated (i.e. not one shot) because there is typically feedback between decisions and the state dependence is usually not so severe that subsequent decisions differ fundamentally from those made in initial round. The opportunities for learning from repetition thus usually dominate (though not necessarily always), and we note that we did not explicitly account for this in our coding.

Coding Criteria: Some main task(s) has (have) feedback between decisions

If at least one of the main tasks was repeated with feedback between rounds in one or more of the treatments, we coded that paper as having a repeated main task with feedback under this column (e.g. a repeated public goods game in which subjects learned their payoff after each round (e.g. Bayer et al. 2013). We considered it sufficient for a subsequent round to involve choices in the same basic task as the preceding one for which feedback was given. For example, in Noussair and Stoop (2015), subjects in one treatment completed two dictator games in a row, with different reward media (money and time) with feedback between them – we viewed these as repetitions of the same task with feedback.

Coding Criteria: More than one task

We coded whether an experiment has more than one incentivized task. In some cases, an experiment required subjects to input multiple separate decisions associated with the same broader task – in these instances, we coded this a single task (as discussed above). Sometimes a single task has multiple decisions (e.g. a centipede game as in Cox and James (2012) or a public goods game with punishment as in Harris et al. (2015). Similarly, in an experiment that required subjects to vote on a sanctioning scheme that would then be implemented in a public goods game (Kamei et al. 2015), we viewed the vote and the subsequent game as one task. Many experiments coded as having more than one task would follow up a main task with a secondary preference elicitation.

Cross-Check

Each paper was independently coded by two coders, who read each of the 260 papers in the review along with any instructions available in their online supplementary materials. For each of the 11 categories coded, both coders marked them as true (=1), false (=0), or uncertain (=0.5). Both coders agreed most of the time, only disagreeing (including cases where one coder was uncertain) in 363 out of 11×260 judgments, and only disagreeing fundamentally (i.e. one coder marking a "0" and the other a "1" on a given paper-category judgment) in 200 such judgments. The area with the most disagreement was the presence of demonstration, examples, or guided practice. These are particularly difficult to identify, as they are often buried in lengthy instructions and the difference between explanation and demonstration is somewhat subjective. We note that false negatives are more likely than false positives – it is easy to miss an example or demonstration in instructions but hard to see one where it doesn't actually exist. After each person coded independently, both coders reconciled disagreements to put together the data for Table 1.1. Typically, when only one coder was uncertain, disagreement was resolved in favor of the certain coder. In the case of genuine disagreement coders discussed and settled on the most likely classification.

Correlations amongst practices

One-shot experiments account for about one third of the experiments using computerized instructions (31%) or paper instructions (35%). 57% of experiments that use neither paper nor on-screen instructions are one-shot games; most of these studies are field experiments in which experimenters read instructions aloud or go through the instruction one-on-one with subjects.

We also find that one-shot experiments tend to be less likely to use each of the reinforcement methods (except for reading aloud) – even though such experiments give no feedback,

Table A.1: Correlation between experiment type and delivery and reinforcement

	One-shot	*p*-value	Feedback between decisions	*p*-value
Paper only	.048	.437	.008	.899
Computer only	-.011	.863	-.082	.189
Both	.018	.770	.022	.722
Neither	.157	.011	-.180	.004
Read aloud	.112	.072	-.092	.141
Practice/Demonstration	-.191	.002	.190	.002
Quiz	-.146	.019	.159	.010
Table reports pairwise correlations between delivery/reinforcement category (rows) and experiment type (columns) and their *p*-values.				

Table A.2: Instruction practices by feedback

	One-shot	Feedback between decisions
Total	84	152
Read aloud	52	76
Practice/Demonstration	36	98
Quiz	24	69

making each subject's initial understanding of the instructions crucial. We suspect that this is because one-shot experiments tend to be simpler and therefore easier to explain. Instructions are read aloud more often in one-shot game experiments (62%) than in experiments with feedback between decisions (50%). Other reinforcement methods are used less often in one-shot experiments than in experiments with feedback between decisions (respectively, 43% versus 65% use some form of practice or demonstration, while 29% versus 45% use a quiz). These differences result in a significant negative association between one-shot experiments and use of practice/demonstration ($\rho = -.191$, $p = .002$) and quizzes ($\rho = -.146$, $p = .019$) in the instructions.

A.2 Experimental Instructions

The experimental sessions all followed the script in Figure A.1.

We include copies of all instructions pages as seen by each subject in all treatments. First, we show the screenshots that apply for all except for the ENHANCED treatment. Note that the printed instructions for the paper treatment did not include the screenshots shown in Figure A.5 and Figure A.7, since they completed practice periods for Tasks 1 and 2 as part of the on-screen instructions, like all other subjects.

Next, we include screenshots from the instructions from the ENHANCED treatment. Note that, unlike in the other treatments, the final summary screen remained displayed in the ENHANCED while subjects wrote the quiz.

Our quiz, which was included after the instructions and before the main experiment in all treatments except for NO QUIZ, featured the following six questions:

In our follow-up experimental sessions, we slightly re-worded some of the quiz questions to make them more clear. This new quiz was administered to all subjects in the ENHANCED treatment and some of the subjects in the QUIZ treatment.

While scores in the QUIZ treatment did increase slightly under the new quiz, from an average of 3.9 to 4.4, this difference is not statistically significant ($p = .11$, rank-sum test), and thus we pool data from all QUIZ sessions. We also did not observe any significant differences in NMB ($p = .50$, Fisher's exact test).

Figure A.1: Experimenter's script for running a session

How to Run a Session

1. Log in to computer 24 with your SFU email
2. Log in to students' computers using username "econ subject" and password "economics" (computers 11 and 12 sometimes freeze!)
3. Open ESILauncher on computer 24
4. Highlight the machine numbers students are using
5. Check the Auto Connect box
6. Select the file "C:\Experiments\PoodleJump\Client\Client.exe"
    a. Replace leading dots with "C:\Experiments"
7. Open "C:\Experiments\PoodleJump\Server\Server.exe" on computer 24
8. Hit "Load Settings" button and select "C:\Experiments\PoodleJump\Server\ExperimentSettings\Low.txt"
9. As participants arrive, mark them as "participated" on http://experiments.econ.sfu.ca/
10. Set the number of participants in both ESI and Server
11. Give consent forms and receipts and instruct participants to fill out everything except the payment amount
12. Take in consent forms
13. Give the pre-experiment speech
    a. Eyes on own screen
    b. Don't communicate with other participants
    c. Raise hand to ask question
    d. No food
    e. Keep drinks in closed containers
    f. Cell phones away
    g. If doing paid quiz, explain about the paid quiz
14. Click the big green check mark in ESI to launch the program
15. Instruct subjects to click "Run"
16. Tell participants to sit quietly once they have finished instructions
17. (if doing quiz) Tell them about quiz (and incentives if quiz is incentivized)
18. Click "Begin Instructions"
19. Allow them to go through the instructions
20. (if doing quiz) Hand out quiz
21. (if doing quiz) Take in quiz
22. (if doing quiz + answers) Read quiz answers
23. Click start button
24. (if doing quiz) Grade quiz during the experiment
25. Mark experiment as "Finished" on http://experiments.econ.sfu.ca/
26. When experiment is complete, ask students to wait at their computers and have their receipts ready
27. Call students by computer number and pay them $7+their experiment payoff, filling out dollar amounts in each receipt
28. Move data files from "..\PoodleJump\Server\Server_Data\" into "Dropbox\PoodleJump\data\[appropriate folder]\"

Figure A.2: Instructions page 1: introduction to the experiment

Introduction

This is an experiment on economic decision-making. If you read the instructions carefully and make good decisions, you can earn a considerable amount of money, which will be paid to you in CASH at the end of the experiment.

During the experiment you are not allowed to communicate with any other participant. If you have any questions, raise your hand, and the experimenter(s) will answer them privately. You must also put away all materials unrelated to the experiment, including cell-phones, tablets, and pen-and-paper.

If you do not follow these instructions you will be **excluded** from the experiment and deprived of all payments aside from the show-up payment of 7 CAD.

Overview

In this experiment, you have the opportunity to complete two tasks for money. The experiment will last 30:00 minutes, and this is divided into 30 Periods of 1:00 each.

Figure A.3: Instructions page 2: description of Task 1

Task 1

Poodle Jump - this task can be performed continuously throughout the experiment. You control a poodle who climbs a series of platforms. The poodle will automatically jump when it touches a platform.

Use the Left and Right mouse buttons to move the poodle around the screen, and make sure you land on a platform, or the poodle will fall down, and you'll start climbing again.

While you are performing this task, the total height that your poodle has climbed in the current period will be recorded in the corner of the screen. A timer will tell you how much time has passed. Practice on the next screen to learn how Poodle Jump works.

Figure A.4: Instructions page 3: Task 1 practice

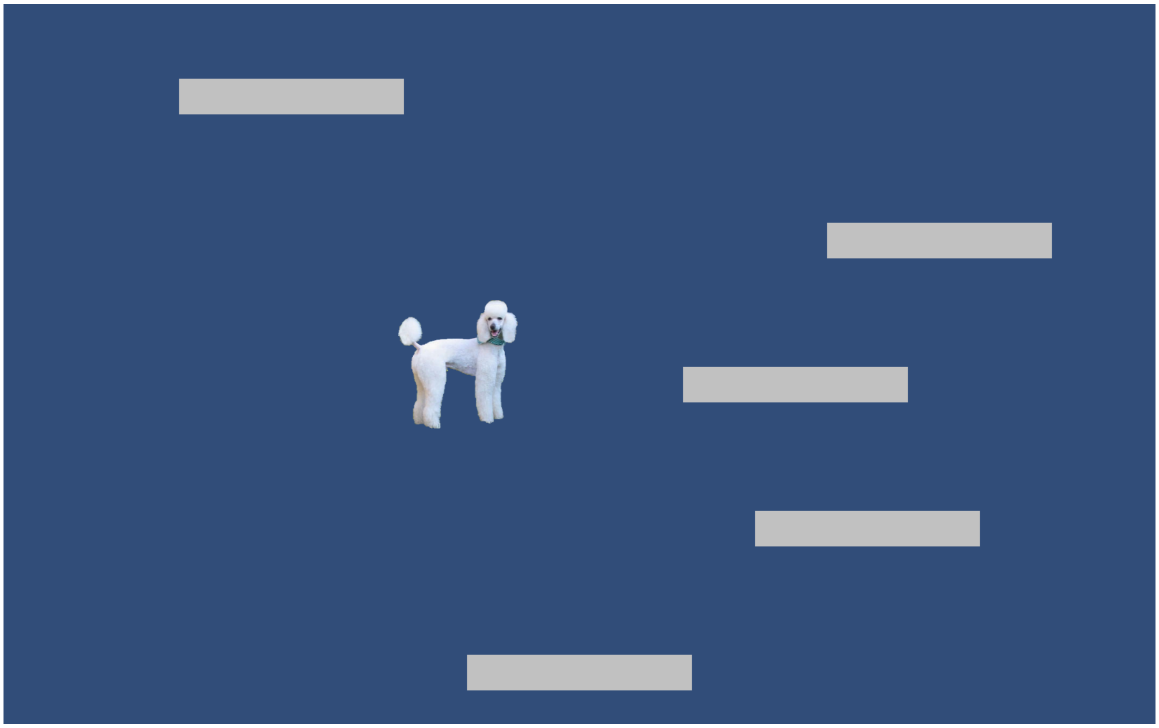


Figure A.5: Instructions page 4: description of Task 2

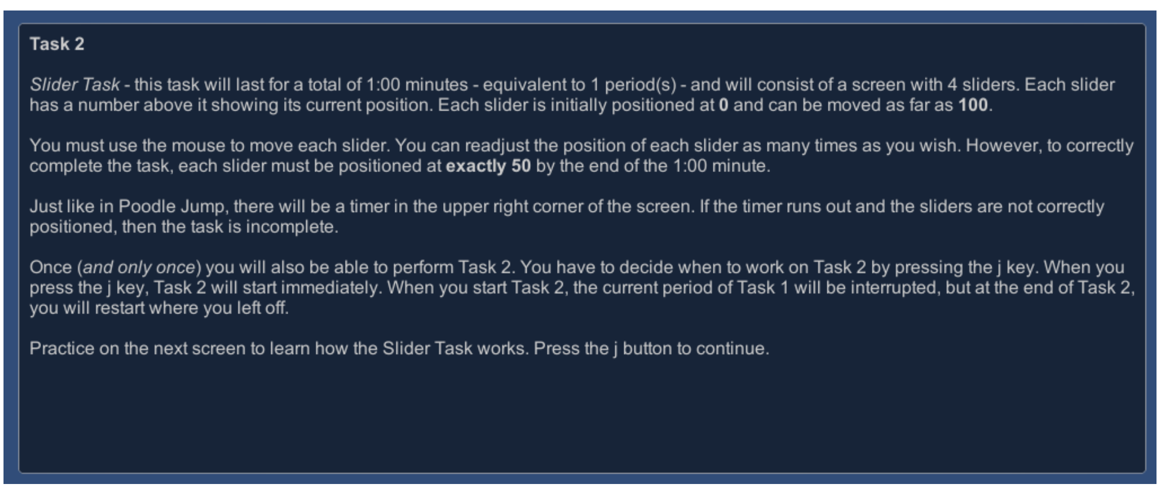


Figure A.6: Instructions page 5: Task 2 practice

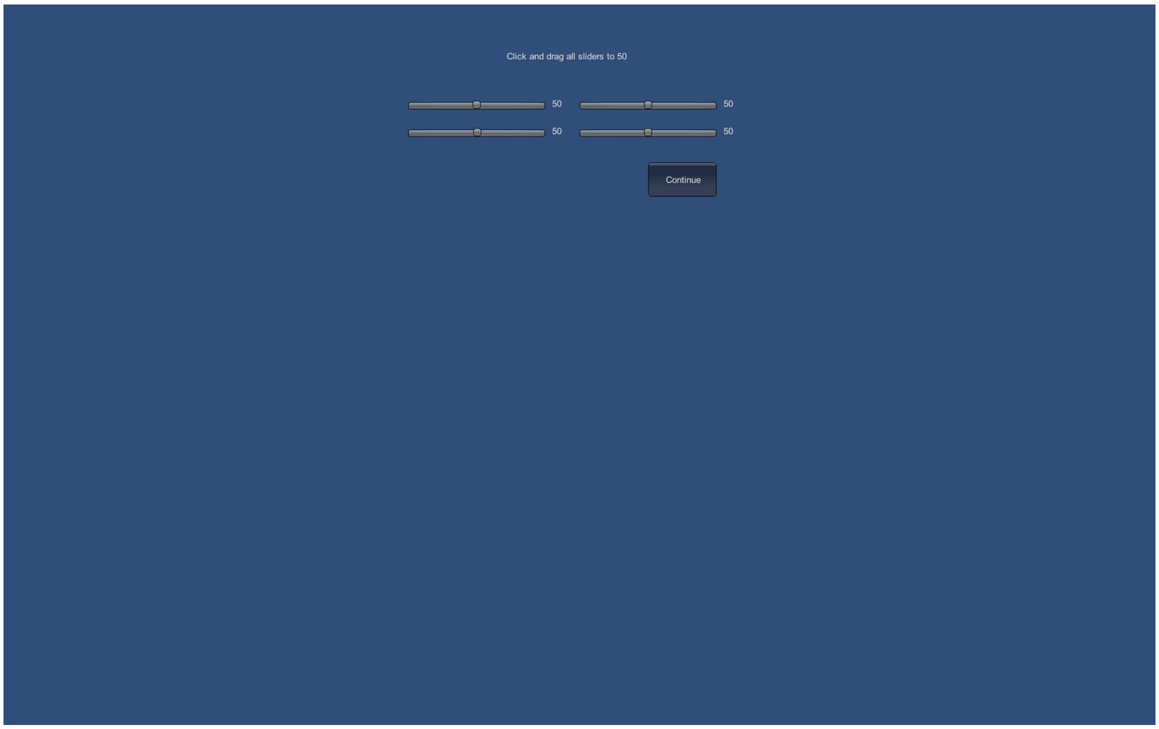


Figure A.7: Instructions page 6: payment schedule description

Your Payment

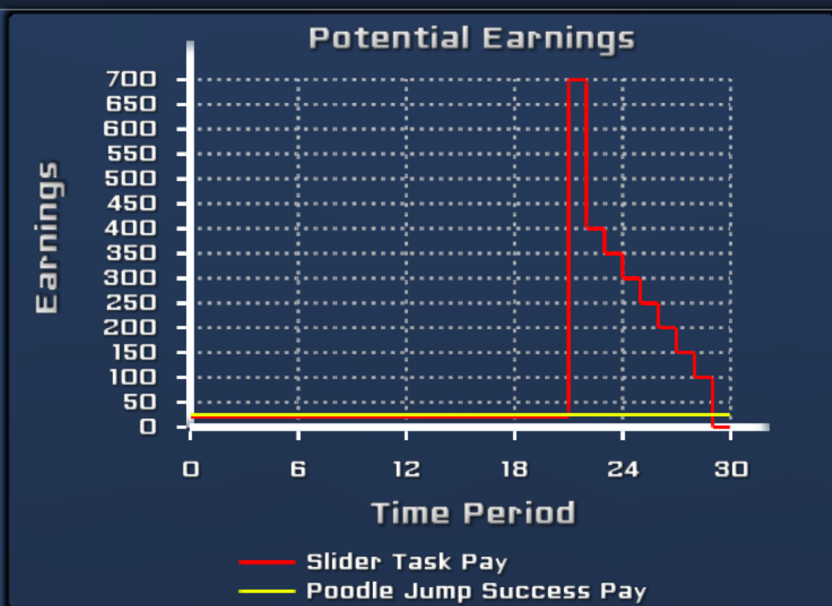
Throughout the experiment, you will perform Task 1. In this task, you get paid 25 Cents each Period only if your poodle is able to jump a total height of 75 Units.

The height you've jumped is displayed in the upper right portion of the screen. The total height is cumulative in a period, so if you fall down and start jumping again, you will **not** have to restart the count. Once you reach the required height of 75, the payment of 25 cents will be added to your earnings. At the start of the next period, the counter will reset.

Your payment for correctly completing Task 2 will depend on when you decide to start Task 2. Look carefully at the figure below. On the horizontal axis is time. On the vertical axis is the payment you would receive if you start Task 2 at that time and correctly complete it.

For the first 21 Periods, the payment for Task 2 is 20 cents. In period 22, the payment is 700 cents. In period 23, the payment is 400 cents. In each subsequent period, the payment declines by 50 cents.

This is the **only time** we will show you this information.



Continue

Figure A.8: Instructions page 7: summary

Summary

The experiment will start, and you will perform Task 1. During the experiment, you have to decide when (and whether) to perform Task 2.

Note that you cannot perform Task 1 and Task 2 at the same time. This means that any time spent on Task 2 will cause you to forgo your earnings from Task 1. In total, the experiment lasts 30 periods. If you never perform Task 2, you will have 30 periods to perform Task 1. By performing Task 2, you will give up 1 periods of Task 1 in order to earn your payment for Task 2.

You earn 25 cents in each period of Task 1, only if your poodle jumps at least 75 pixels in total height. You earn **0** cents otherwise.

Your payment for correctly completing Task 2 depends on when you start Task 2. Note that you are not required to perform Task 2. It is optional.

Are there any questions?

Figure A.9: ENHANCED Instructions page 1: introduction to the experiment

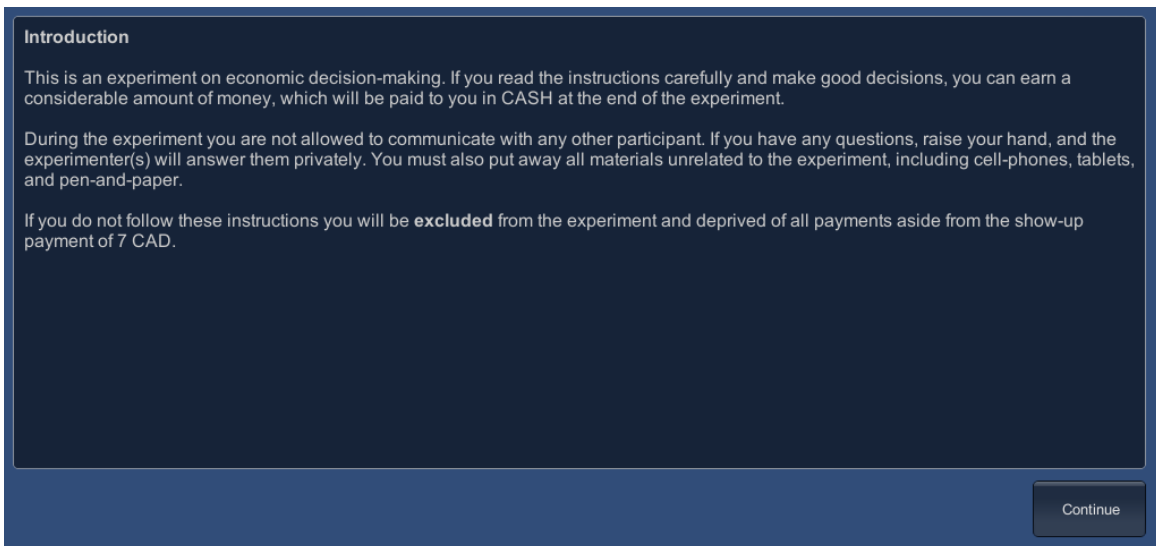


Figure A.10: ENHANCED Instructions page 2: overview and payment

Overview

In this experiment, you have the opportunity to complete two tasks for money. The experiment will last 30:00 minutes, and this is divided into 30 Periods of 1:00 each.

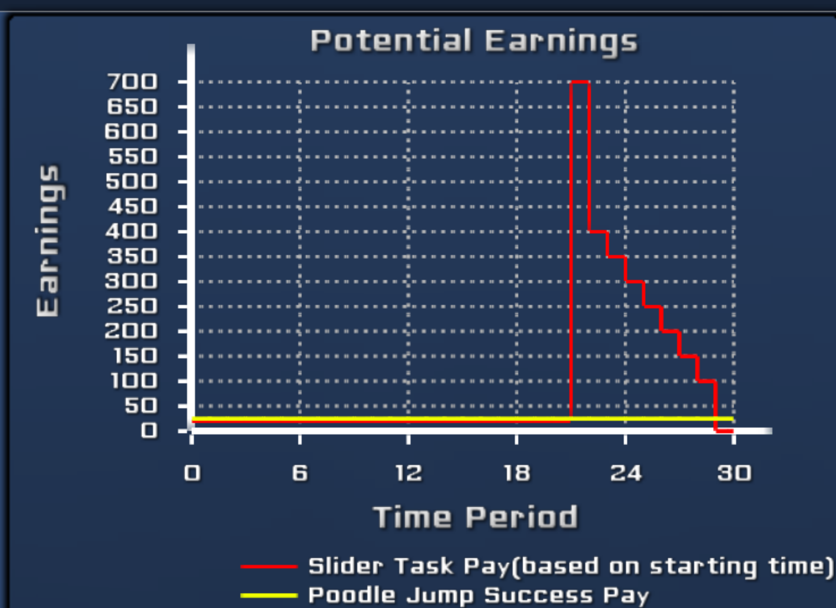
Your Payment

Throughout the experiment, you will perform Task 1, called 'Poodle Jump'. In this task, you get paid 25 Cents each Period only if your poodle is able to jump a total height of 75 Units.

The height you've jumped is displayed in the upper right portion of the screen. The total height is cumulative in a period, so if you fall down and start jumping again, you will **not** have to restart the count. Once you reach the required height of 75, the payment of 25 cents will be added to your earnings. At the start of the next period, the counter will reset.

You can complete Task 2, called the 'Slider Task', at most ***once*** *during the experiment. Your payment for correctly completing Task 2 will depend on when you decide to start Task 2.* Look carefully at the figure below. On the horizontal axis is time. On the vertical axis is the payment you would receive if you start Task 2 at that time and correctly complete it.

For the first 21 Periods, the payment for Task 2 is 20 cents. In period 22, the payment is 700 cents. In period 23, the payment is 400 cents. In each subsequent period, the payment declines by 50 cents.



Continue

Figure A.11: ENHANCED Instructions page 3: payment examples

Figure A.12: ENHANCED Instructions page 4: description of Task 1

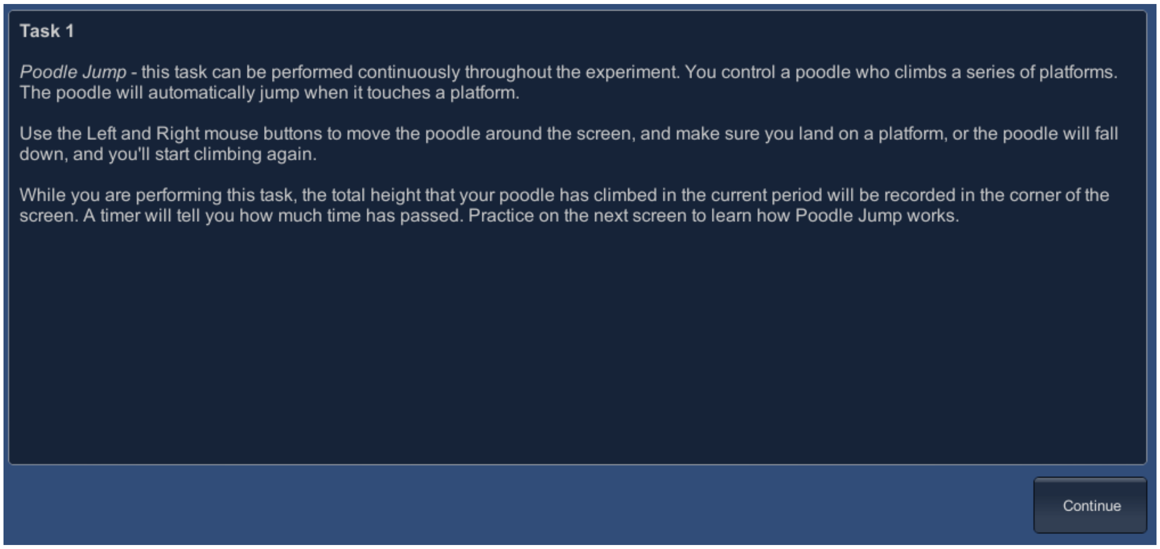


Figure A.13: ENHANCED Instructions page 5: Task 1 practice

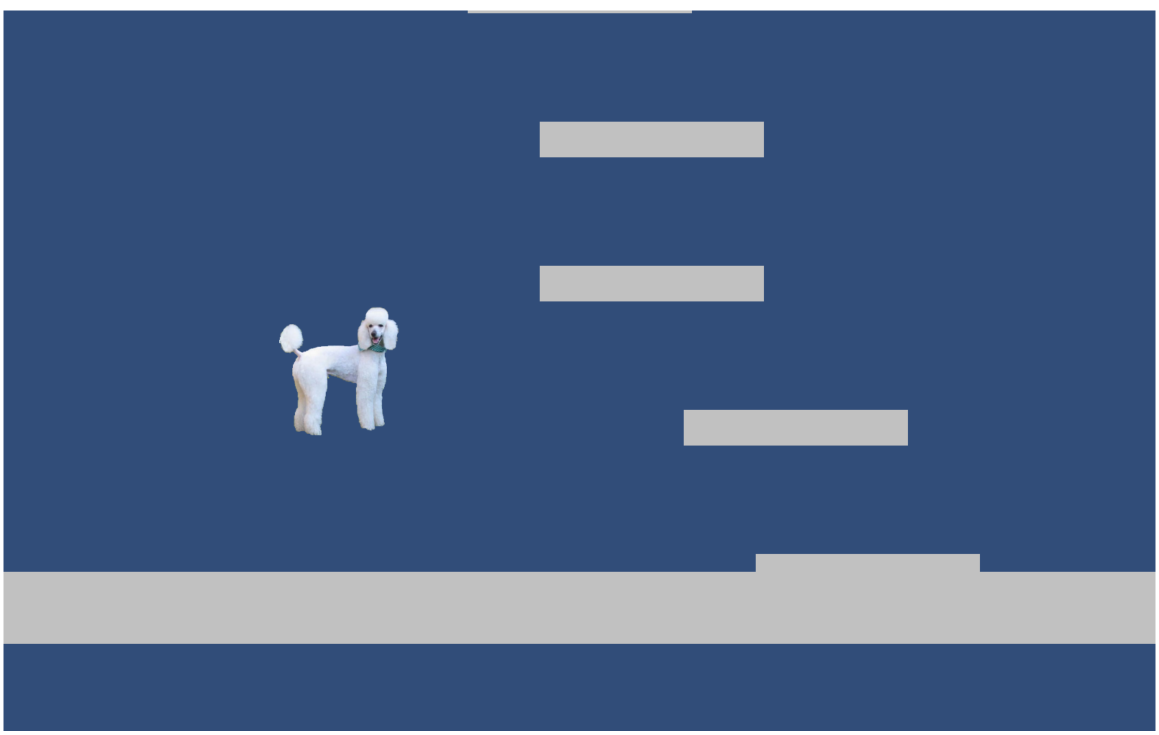
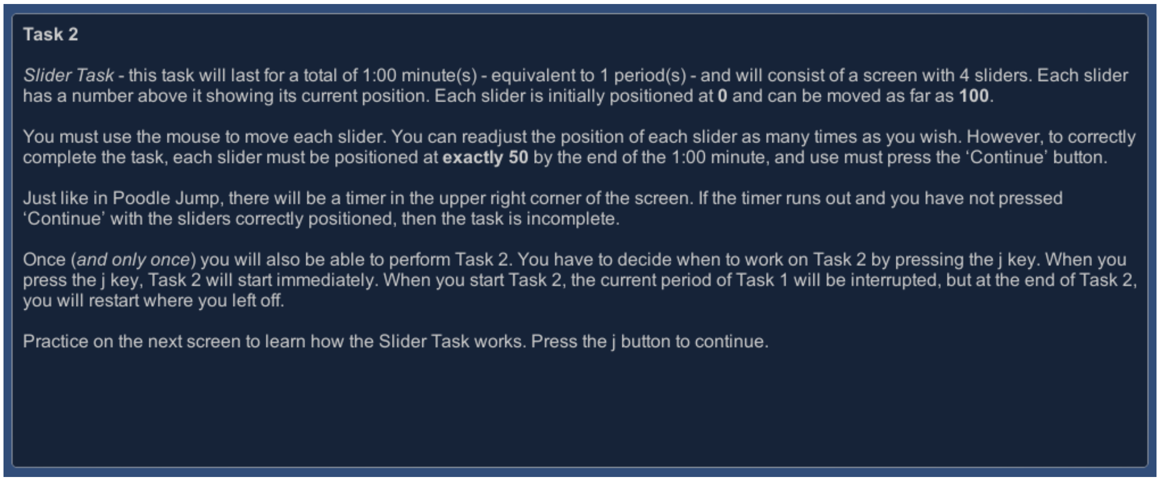


Figure A.14: ENHANCED Instructions page 6: description of Task 2



Task 2

Slider Task - this task will last for a total of 1:00 minute(s) - equivalent to 1 period(s) - and will consist of a screen with 4 sliders. Each slider has a number above it showing its current position. Each slider is initially positioned at **0** and can be moved as far as **100**.

You must use the mouse to move each slider. You can readjust the position of each slider as many times as you wish. However, to correctly complete the task, each slider must be positioned at **exactly 50** by the end of the 1:00 minute, and use must press the ‘Continue’ button.

Just like in Poodle Jump, there will be a timer in the upper right corner of the screen. If the timer runs out and you have not pressed ‘Continue’ with the sliders correctly positioned, then the task is incomplete.

Once (*and only once*) you will also be able to perform Task 2. You have to decide when to work on Task 2 by pressing the j key. When you press the j key, Task 2 will start immediately. When you start Task 2, the current period of Task 1 will be interrupted, but at the end of Task 2, you will restart where you left off.

Practice on the next screen to learn how the Slider Task works. Press the j button to continue.

Figure A.15: ENHANCED Instructions page 7: Task 2 practice

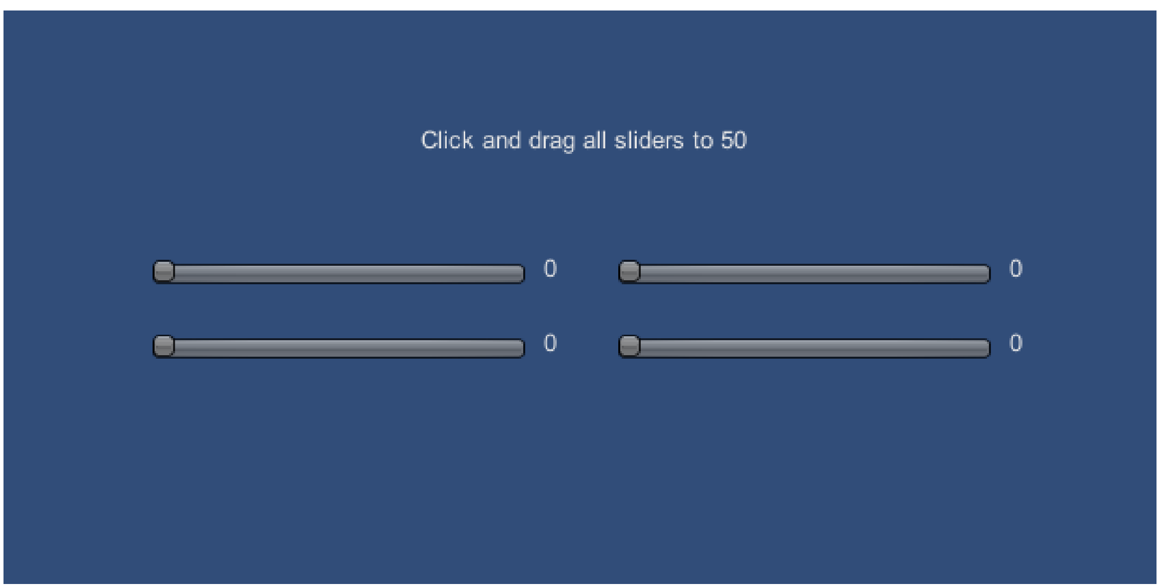


Figure A.16: ENHANCED Instructions page 8: payment recap

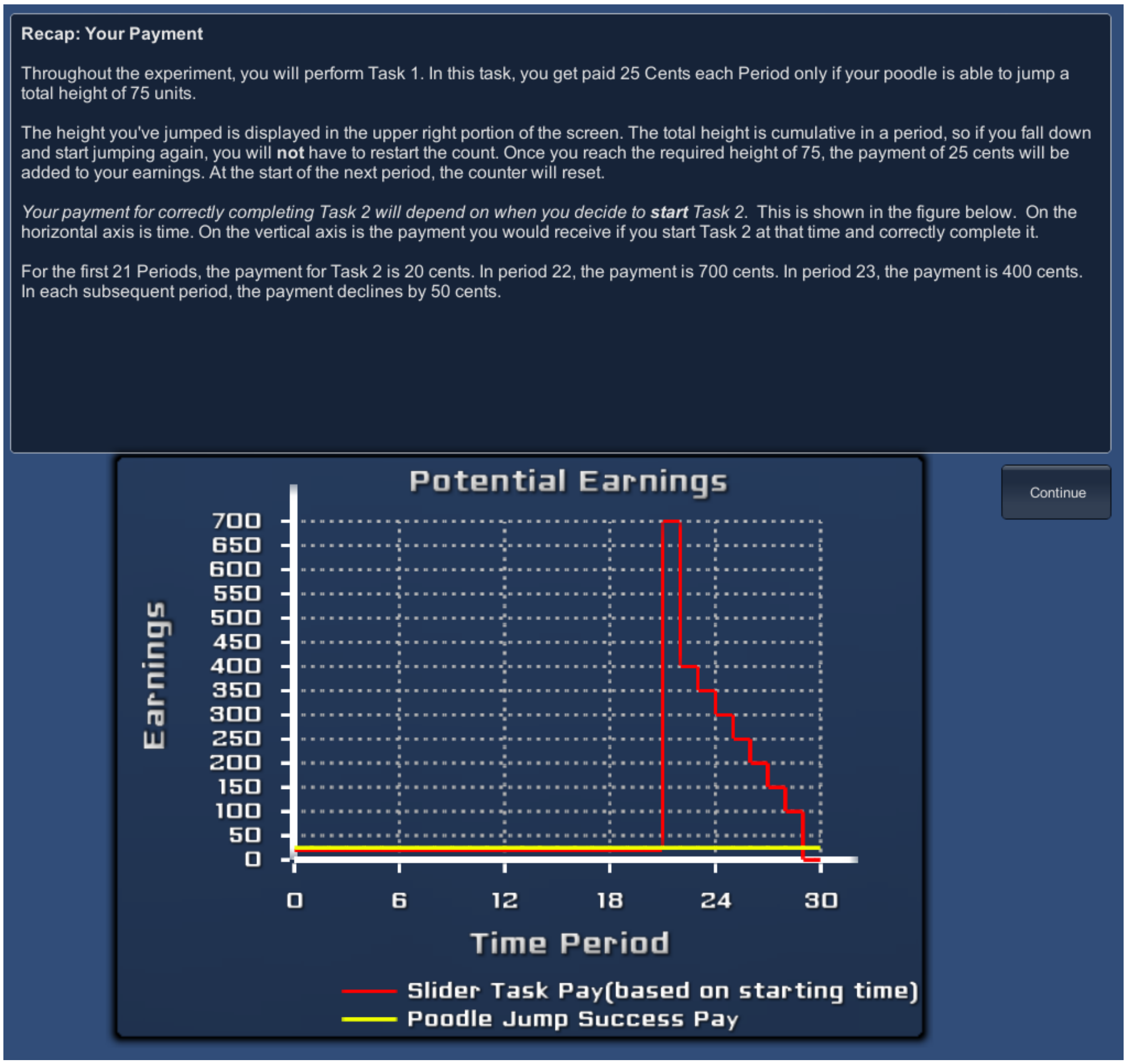


Figure A.17: ENHANCED Instructions page 7: summary

Figure A.18: Post-instructions quiz

Q1. At what period is the payment to completing Task 2 the highest? A: _______

Q2. What is the payment for completing Task 2 at a time indicated in your answer to Q1? A: _______

Q3. What is the payment for completing Task 2 at a time before your answer to Q1? A: _______

Q4. What is the payment for completing each period of Task 1? A: _______

Q5. What key do you need to press to switch from Task 1 to Task 2? A: _______

Q6. How many times may you complete Task 2? A: _______

Figure A.19: Revised post-instructions quiz

Q1. At what period is the payment to starting Task 2 the highest, assuming that you complete it? A: _______

Q2. What is the payment for Task 2 at the time indicated in your answer to Q1? A: _______

Q3. What is the payment for starting Task 2 at any time before your answer to Q1? A: _______

Q4. What is the payment for completing each period of Task 1? A: _______

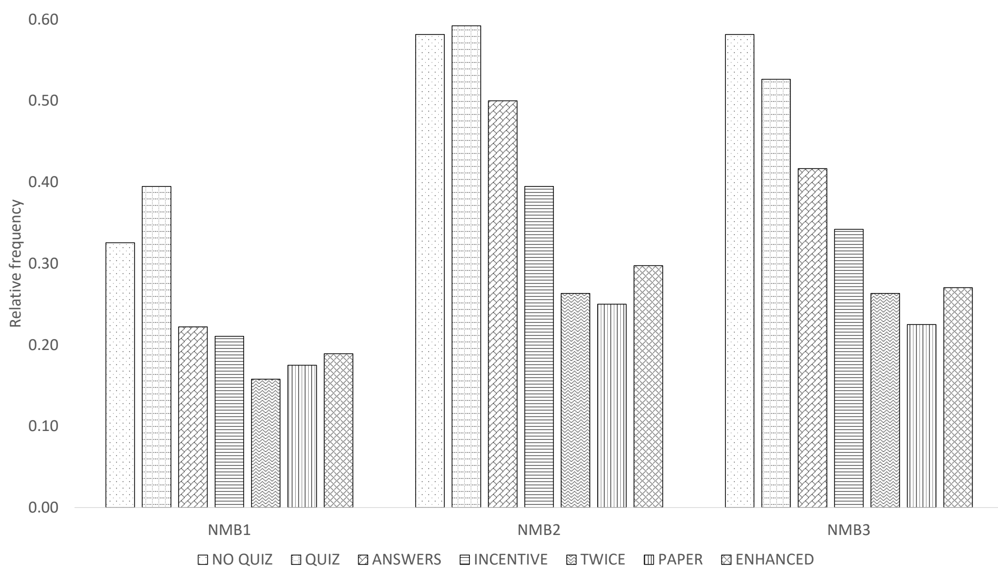
Q5. What key do you need to press to switch from Task 1 to Task 2? A: _______

Q6. How many times may you complete Task 2? A: _______

A.3 Robustness Checks

We redo our analysis with three alternative measures of NMB to check the robustness our results. The specifications reported in Table A.3.3-5 are all analogous to the specifications in Table 1.4, but with alternative definitions of NMB. The dependent variable "NMB1" is equal to one if the subject did Task 2 before period 21 and equal to zero otherwise; this measure of NMB allows for trembles. The "NMB2" variable defines any behavioral deviation from optimality as NMB. That is, it classifies a subject as exhibiting NMB unless they did Task 2 exactly in period 22. Finally, the "NMB3" variable classifies those who did Task 2 before period 22 or never at all as NMB. The results of these alternative specifications are broadly consistent with those reported in Table 1.4. Figure A.20 plots the share of subjects with NMB in each treatment, by each of these alternative measures. To check the robustness of our logit regressions, Table A.6 reports estimated linear probability models with (OLS analogues to columns 1 and 2 of Table 1.4); for comparison purposes note that we do not report marginal effects in Table 1.4 since the mediation analysis in column 4 provides the economically meaningful estimates of interest.

Figure A.20: Percentage of subjects revealing NMB, under three alternative definitions of NMB, by treatment.



We note that our statistical tests find significant differences between our main QUIZ treatment and each of our INCENTIVE, TWICE, PAPER, and ENHANCED treatments, but do not detect significant differences among the latter four treatments, and also detects no significant difference between the ANSWERS treatment and other treatments (see Table 1.2 in the main text). This raises the question of statistical power. We note that the comparisons between the QUIZ treatment and each of the INCENTIVE, TWICE, PAPER, and ENHANCED treatments appear to be appropriately powered. Across the latter four treatments, 21.6% of subject misunderstand (a fraction which ranges between 18.4-23.7% across

Table A.3: Treatment effects on NMB1 and Quiz Scores

	Dependent variable			Mediation analysis	
	NMB1		Quiz Score	NMB1	
	(1)	(2)	(3)	(4)	
NO QUIZ	-0.301				n
	(-1.096, 0.495)				
ANSWERS	-0.825*	0.207	-0.050	-0.169*	112
	(-1.746, 0.096)	(-2.648, 3.061)	(-0.586, 0.487)	(-0.329, 0.008)	
ANSWERS × Quiz Score		-0.324		0.005	
		(-1.062, 0.413)		(-0.056, 0.070)	
INCENTIVE	-0.894*	-1.380	0.211	-0.164*	114
	(-1.810, 0.022)	(-4.202, 1.422)	(-0.354, 0.775)	(-0.331, 0.021)	
INCENTIVE × Quiz Score		0.127		-0.022	
		(-0.508, 0.762)		(-0.091, 0.039)	
TWICE	-1.247**	-0.677	0.421	-0.199**	114
	(-2.244, -0.249)	(-3.940, 2.586)	(-0.156, 0.998)	(-0.367, -0.010)	
TWICE × Quiz Score		-0.135		-0.044	
		(-0.847, 0.578)		(-0.119, 0.016)	
PAPER	-1.123**	7.787**	1.320***	0.163	116
	(-2.070, -0.176)	(1.053, 14.521)	(0.922, 1.718)	(-0.118, 0.375)	
PAPER × Quiz Score		-1.632**		-0.133***	
		(-2.901, -0.363)		(-0.223, -0.046)	
ENHANCED	-1.028**	0.249	0.489*	-0.139*	113
	(-1.981, -0.074)	(-3.675, 4.174)	(-0.030, 1.008)	(-0.325, 0.060)	
ENHANCED × Quiz Score		-0.273		-0.051*	
		(-1.144, 0.598)		(-0.123, 0.003)	
Quiz Score		-0.519***			
		(-0.875, -0.164)			
Intercept	-0.427*	1.662**	4.105***		
	(-0.893, 0.038)	(0.121, 3.202)	(3.789, 4.422)		
Observations	308	265	265		

QUIZ is the omitted category. *, **, and *** respectively denote $p < .1$, $p < .05$, $p < .01$. Robust (HC1) 95% confidence intervals are in parentheses in Columns (1)-(4). Mediation column reports estimated "direct effects" in the row of a treatment dummy, and mediated effects in the row of the interaction term between Quiz Score and that treatment dummy, both evaluated relative to the QUIZ baseline. That is, the direct effect of a treatment corresponds to $\mathbb{E}[\text{NMB}|\text{Treatment}, \text{Quiz Score} = 4.1] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$, while the mediated effect corresponds to $\mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = \mathbb{E}[\text{Quiz Score}|\text{Treatment}]] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$.

Table A.4: Treatment effects on NMB2 and Quiz Scores

	Dependent variable			Mediation analysis	
	NMB2		Quiz Score	NMB2	
	(1)	(2)	(3)	(4)	
NO QUIZ	-0.044				n
	(-0.812, 0.724)				
ANSWERS	-0.373	-1.477	-0.050	-0.103	112
	(-1.179, 0.434)	(-5.163, 2.209)	(-0.586, 0.487)	(-0.268, 0.070)	
ANSWERS × Quiz Score		0.192		0.009	
		(-0.624, 1.009)		(-0.098, 0.116)	
INCENTIVE	-0.800*	-2.840	0.211	-0.164*	114
	(-1.605, 0.004)	(-6.591, 0.911)	(-0.354, 0.775)	(-0.344, 0.013)	
INCENTIVE × Quiz Score		0.443		-0.042	
		(-0.353, 1.239)		(-0.157, 0.069)	
TWICE	-1.402***	-2.201	0.421	-0.254***	114
	(-2.267, -0.538)	(-6.525, 2.122)	(-0.156, 0.998)	(-0.424, -0.078)	
TWICE × Quiz Score		0.130		-0.084	
		(-0.879, 1.138)		(-0.203, 0.031)	
PAPER	-1.471***	8.269	1.320***	0.056	116
	(-2.331, -0.612)	(-4.351, 20.889)	(0.922, 1.718)	(-0.288, 0.236)	
PAPER × Quiz Score		-1.652		-0.284***	
		(-4.001, 0.698)		(-0.389, -0.182)	
ENHANCED	-1.233***	-2.724	0.489*	-0.216**	113
	(-2.083, -0.383)	(-6.883, 1.434)	(-0.030, 1.008)	(-0.402, -0.033)	
ENHANCED × Quiz Score		0.345		-0.101*	
		(-0.560, 1.249)		(-0.212, 0.007)	
Quiz Score		-1.344***			
		(-1.872, -0.816)			
Intercept	0.373	6.236***	4.105***		
	(-0.090, 0.835)	(3.637, 8.836)	(3.789, 4.422)		
Observations	308	265	265		

QUIZ is the omitted category. *, **, and *** respectively denote $p < .1$, $p < .05$, $p < .01$. Robust (HC1) 95% confidence intervals are in parentheses in Columns (1)-(4). Mediation column reports estimated "direct effects" in the row of a treatment dummy, and mediated effects in the row of the interaction term between Quiz Score and that treatment dummy, both evaluated relative to the QUIZ baseline. That is, the direct effect of a treatment corresponds to $\mathbb{E}[\text{NMB}|\text{Treatment}, \text{Quiz Score} = 4.1] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$, while the mediated effect corresponds to $\mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = \mathbb{E}[\text{Quiz Score}|\text{Treatment}]] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$.

Table A.5: Treatment effects on NMB3 and Quiz Scores

	Dependent variable			Mediation analysis	
	NMB3		Quiz Score	NMB3	
	(1)	(2)	(3)	(4)	
NO QUIZ	0.223				n
	(-0.540, 0.987)				
ANSWERS	-0.442	-0.360	-0.050	-0.112	112
	(-1.252, 0.369)	(-3.559, 2.839)	(-0.586, 0.487)	(-0.278, 0.070)	
ANSWERS × Quiz Score		-0.075		0.009	
		(-0.851, 0.702)		(-0.089, 0.106)	
INCENTIVE	-0.759*	-2.207	0.211	-0.157*	114
	(-1.810, 0.022)	(-5.334, 0.921)	(-0.354, 0.775)	(-0.336, 0.028)	
INCENTIVE × Quiz Score		0.127		-0.037	
		(-0.508, 0.762)		(-0.141, 0.063)	
TWICE	-1.135***	-0.365	0.421	-0.183**	114
	(-1.996, -0.274)	(-4.401, 3.670)	(-0.156, 0.998)	(-0.356, -0.004)	
TWICE × Quiz Score		-0.193		-0.074	
		(-1.163, 0.776)		(-0.181, 0.026)	
PAPER	-1.342***	5.617	1.320***	0.074	116
	(-2.220, -0.464)	(-2.618, 13.852)	(0.922, 1.718)	(-0.252, 0.278)	
PAPER × Quiz Score		-1.143		-0.240***	
		(-2.649, -0.363)		(-0.340, -0.143)	
ENHANCED	-1.099**	0.321	0.489*	-0.158*	113
	(-1.962, -0.235)	(-3.790, 4.431)	(-0.030, 1.008)	(-0.349, 0.028)	
ENHANCED × Quiz Score		-0.314		-0.088*	
		(-1.201, 0.574)		(-0.189, 0.006)	
Quiz Score		-1.021***			
		(-1.471, -0.571)			
Intercept	0.105	4.400***	4.105***		
	(-0.350, 0.561)	(2.314, 6.486)	(3.789, 4.422)		
Observations	308	265	265		

QUIZ is the omitted category. *, **, and *** respectively denote $p < .1$, $p < .05$, $p < .01$. Robust (HC1) 95% confidence intervals are in parentheses in Columns (1)-(4). Mediation column reports estimated "direct effects" in the row of a treatment dummy, and mediated effects in the row of the interaction term between Quiz Score and that treatment dummy, both evaluated relative to the QUIZ baseline. That is, the direct effect of a treatment corresponds to $\mathbb{E}[\text{NMB}|\text{Treatment}, \text{Quiz Score} = 4.1] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$, while the mediated effect corresponds to $\mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = \mathbb{E}[\text{Quiz Score}|\text{Treatment}]] - \mathbb{E}[\text{NMB}|\text{QUIZ}, \text{Quiz Score} = 4.1]$.

Table A.6: Treatment effects on NMB – linear probability model robustness checks

	Dependent variable					
	NMB1		NMB2		NMB3	
NO QUIZ	-0.069		-0.011		0.055	
	(0.092)		(0.095)		(0.096)	
ANSWERS	-0.173*	-0.098	-0.092	-0.098	-0.110	-0.056
	(0.090)	(0.289)	(0.102)	(0.156)	(0.101)	(0.183)
INCENTIVE	-0.184**	-0.366	-0.197**	-0.316	-0.184*	-0.374
	(0.088)	(0.294)	(0.098)	(0.203)	(0.097)	(0.234)
TWICE	-0.237***	-0.283	-0.329 ***	-0.378*	-0.263***	-0.242
	(0.082)	(0.303)	(0.092)	(0.194)	(0.093)	(0.205)
PAPER	-0.220 ***	0.893*	-0.342***	0.911**	-0.301***	0.697
	(0.083)	(0.454)	(0.090)	(0.409)	(0.088)	(0.460)
ENHANCED	-0.206**	-0.126	-0.295***	-0.330	-0.256***	-0.111
	(0.086)	(0.347)	(0.095)	(0.254)	(0.094)	(0.250)
Quiz Score		-0.117***		-0.209***		-0.192***
		(0.035)		(0.022)		(0.026)
ANSWERS × Quiz Score		-0.020		-0.001		-0.016
		(0.060)		(0.040)		(0.043)
INCENTIVE × Quiz Score		0.048		0.038		0.053
		(0.059)		(0.042)		(0.048)
TWICE × Quiz Score		0.021		0.030		0.013
		(0.058)		(0.039)		(0.041)
PAPER × Quiz Score		-0.177**		-0.180**		-0.137*
		(0.079)		(0.069)		(0.078)
ENHANCED × Quiz Score		-0.005		0.030		-0.011
		(0.067)		(0.050)		(0.046)
Intercept	0.395***	0.873***	0.592***	1.450***	0.526***	1.313***
	(0.057)	(0.166)	(0.057)	(0.090)	(0.058)	(0.113)
Observations	308	265	308	265	308	265
R^2	0.044	0.194	0.082	0.387	0.072	0.340

QUIZ is the omitted category. *, **, and *** respectively denote $p < 0.1$, $p < .05$, $p < .01$.

Robust (HC1) standard errors are in parentheses.

these treatments),[1] while 47.4% of subjects in the QUIZ treatment misunderstand. A simple ex-post power calculation indicates that if we recruited $n_1 = 76$ and $n_2 = 38$ subjects to two treatments in which each subject misunderstands with probability $p_1 = .474$ and $p_2 = .216$ (respectively), then we have a 79.4% chance of detecting a statistically significant difference between treatments (at the 5% significance level). This suggests a reasonable level of power in our comparisons between the four aformentioned treatments and QUIZ. However, 33.3% of subject misunderstand in the ANSWERS treatment – an intermediate case between QUIZ and these other four treatments. If we recruited $n_1 = 76$ and $n_2 = 36$ subjects to two treatments in which each subject misunderstands with probability $p_1 = .474$ and $p_2 = .333$ (respectively), then we have only a 33.2% chance of detecting a statistically significant difference between treatments. If instead we recruited $n_1 = 38$ and $n_2 = 36$ subjects to two treatments in which each subject misunderstands with probability $p_1 = .216$ and $p_2 = .333$ (respectively), then we have only a 18.2% chance of detecting a statistically significant difference between treatments. These calculations indicate that our sample sizes are too small to reliably detect a statistically significant difference between our ANSWERS treatment and the QUIZ treatment, or between the ANSWERS treatment and any of the INCENTIVE, TWICE, PAPER, and ENHANCED treatments. If we instead view the NO QUIZ and QUIZ, pooled, as baseline instructions treatments without reinforcement, and the remaining treatments as enhanced instructions or reinforcement treatments, then our samples have $n_1 = 119$, $n_2 = 189$, $p_1 = .462$, and $p_2 = .238$; under these samples sizes and NMB probabilities, we had a 98.3% chance of detecting a significant difference in NMB.

Our statistical analysis was conducted in R (R Core Team 2017). The regressions in Table 1.4 (and above) used the 'lm' and 'glm' command in the base 'stats' package, with robust standard errors calculated using the 'sandwich' package (Zeileis 2004; 2006). Mediation analysis used the 'mediation' package (Tingley et al. 2014). Goodman-Kruskal gamma tests use the 'DescTools' package (Signorell, 2018). We used the 'pwr' package (Champely 2018) for the power analysis reported above. Figures made in 'ggplot2' (Wickham 2009).

A.4 Post-experiment Questionnaire

At suggestion of a referee and the editor, we added a post-experiment questionnaire to our ENHANCED treatment, and ran additional sessions of the QUIZ treatment followed by this questionnaire to paint a more complete picture of subjects' decisionmaking processes as they went though the experiment. We asked nine questions in total.

Our first observation is that there is no statistical difference between QUIZ and ENHANCED on any of the first six quantitative questions.

[1] These numbers are relatively close to each other, so we use the 21.6% for our illustrative calculations below.

Figure A.21: Post-experiment questionnaire (Page 1)

Post-Experiment Questionnaire

Q1. Please think back to when you read the instructions and rate how much you agree with the following three statements on a scale of 1 to 7:

i. The instructions were clear.

1	2	3	4	5	6	7
Strongly Disagree			Neither Agree nor Disagree			Strongly Agree

ii. I understood the best time to switch to task 2 (the slider task) – that is, when to switch in order to get the highest payment.

1	2	3	4	5	6	7
Strongly Disagree			Neither Agree nor Disagree			Strongly Agree

iii. I understood that I could only complete task 2 once.

1	2	3	4	5	6	7
Strongly Disagree			Neither Agree nor Disagree			Strongly Agree

Figure A.22: Post-experiment questionnaire (Page 2)

Q2. Please think back to when the experiment was underway and rate how much you agree with the following three statements on a scale of 1 to 7:

i. My main goal in the experiment was to maximize my earnings.

1	2	3	4	5	6	7
Strongly Disagree			Neither Agree nor Disagree			Strongly Agree

ii. I remembered the best time to switch to task 2.

1	2	3	4	5	6	7
Strongly Disagree			Neither Agree nor Disagree			Strongly Agree

iii. I remembered that I could only complete task 2 once.

1	2	3	4	5	6	7
Strongly Disagree			Neither Agree nor Disagree			Strongly Agree

Figure A.23: Post-experiment questionnaire (Page 3)

Q3. Describe, in your own words, the rules of the experiment.

Q4. Describe, in your own words, how you decided whether and when to switch to task 2.

Q5. What advice would you give to a future participant in this experiment?

Table A.7: Subject evaluation

	QUIZ	ENHANCED	p-value
Comprehension			
Q1i (Clarity)	5.7 (6)	5.4 (6)	0.31
Q1ii (Understood Optimum)	5.7 (7)	5.6 (7)	0.41
Q1iii (Understood Once)	5.4 (7)	5.9 (7)	0.55
Retention			
Q2i (Maximized Earnings)	6.4 (7)	6.3 (7)	0.43
Q2ii (Remembered Optimum)	5.8 (7)	5.6 (6)	0.57
Q2iii (Remembered Once)	5.6 (7)	6.0 (7)	0.38

Mean (median) reported; p-values for rank-sum tests of equality of distributions.

Table A.8: Correlation between subjects' evaluation and misunderstanding and quiz score

	misunderstanding	p.value_misunderstanding	quiz score	p.value_score
Q1i	-0.168	0.159	0.281	0.017
Q1ii	-0.267	0.024	0.202	0.089
Q1iii	-0.406	0.0004	0.202	0.088
Q2i	0.039	0.744	0.046	0.700
Q2ii	-0.371	0.001	0.383	0.001
Q2iii	-0.356	0.002	0.196	0.100

Table A.7 shows that our post-experimental questionnaire results indicate that subjects largely felt that they both understood and retained the key pieces of information from the instructions – with the median subject indicating that they agreed or strongly agreed that they understood and remembered when they should switch (Q1ii, Q2ii), and how many times they could switch (Q1iii, Q2iii). In addition, most subjects agreed with the statement "The instructions were clear", with the median subject rating the statement a 6 out of 7. We find no significant differences between the distribution of answers to any of these questions between the QUIZ and ENHANCED treatments ($p > .3$ in all pairwise comparisons, rank-sum tests). Since we do observe a difference in NMB revealed in the experiment, our post-experimental questionnaire inadvertently reveals its limits at diagnosing reasons for NMB and the potential for improvements. That being said, Table A.8 indicates that subjects' post-experiment answers strongly correlate with both NMB in the experiment and quiz scores. Post-experiment reports of understanding (Q1ii,iii) and retention (Q2ii,iii) were each negatively correlated with NMB ($p < .03$ in all cases). In addition, the subject's post-experimental agreement with the statement "The instructions were clear" was positively correlated with their post-instructions quiz score ($\rho = .281$, $p = .017$).

22 of the 72 subjects who wrote the questionnaire mentioned the instructions in their written answers. Nearly all of these were in Q5: "What advice would you give to a future participant in this experiment?" For instance, the first three subjects to mention the instructions answered Q5 as follows: "Pay attention to the instructions." "Do the experiment with patience and read instructions very carefully." "Read the instructions and follow them for more $." These are typical answers; many subjects recognized, ex post, that paying close attention to the instructions was important for achieving the maximum payoff.

21 of the 72 subjects who wrote the questionnaire showed some kind of mistaken understanding of the experiment, even after having completed it. Many of these misunderstandings were orthogonal to our variable of interest (the time to do task 2). For instance, although our instructions clearly stated that one could get a $0.25 payoff for each period of task 1 if a certain threshold was reached, many seemed to believe that one could earn more than $0.25 by doubling or tripling the threshold. For instance, one subject wrote, "You have a poodle that jumps on to platforms, each 75 units, you get paid 25c." Another one wrote, "Roughly, I would only get 50c at most doing poodle jump for the whole period." The payoff is fixed at 25 cents, so 50 would be impossible. Many subjects appear to believe that they could earn for both tasks 1 and 2 if they completed the minimum height before switching. This is a minor misunderstanding, though it is stated in the instructions that one must forego earnings from one period of task 1 in order to perform task 2.

However, the majority of subjects do not show explicit misunderstandings in their answers, and some even demonstrate learning. One subject who did not perform task 2 at the correct time wrote, "I wasn't aware I can only switch to task 2 only once. So I switched to task 2 in the first period." Another wrote, "I thought it didn't mention number of times we could do the bonus so I did it very early on." These subjects clearly realized their mistakes after they had made them, which suggests that repeated decisions (with feedback of some form) can be a substitute for reinforcing understanding. On the other hand, some subjects failed to understand our instructions and still didn't understand them afterwards. One such subject wrote, "If you taking task 1, you can change game into task 2, but you cannot turn back to task 1."

Appendix B

Appendix for Chapter 2

B.1 Derivation

This section shows the derivation of $L1$'s best response to a Uniform $L0$. Consider a traveler's dilemma game with a discrete choice range $[\underline{x}, \overline{x}]$ and a reward/penalty parameter R. Let $L0$ be a uniform distribution over possible choices.

If player i holds the belief that $L0$ plays a mixed strategy, specifically, uniformly randomly $x_{-i} \sim U[\underline{x}, \overline{x}]$, then i chooses x_i to maximize expected payoff π_i^e:

$\text{Prob}(x_i > x_{-i})[E(x_{-i}|x_{-i} < x_i) - R] + \text{Prob}(x_i < x_{-i})(x_i + R) + \text{Prob}(x_i = x_{-i})x_i$

$= \frac{x_i - \underline{x}}{\overline{x} - \underline{x} + 1} \left(\frac{x_i - 1 + \underline{x}}{2} - R \right) + \frac{(\overline{x} - x_i)(x_i + R)}{\overline{x} - \underline{x} + 1} + \frac{x_i}{\overline{x} - \underline{x} + 1}$

$= \frac{-1}{2(\overline{x} - \underline{x} + 1)} x_i^2 + \frac{\overline{x} + \frac{1}{2} - 2R}{\overline{x} - \underline{x} + 1} x_i - \frac{\underline{x}^2 + \underline{x}}{2(\overline{x} - \underline{x} + 1)} + \frac{R(\underline{x} + \overline{x})}{\overline{x} - \underline{x} + 1}$

F.O.C. $-x_i + (\overline{x} + \frac{1}{2} - 2R) = 0$

$\Rightarrow x_i^* = \overline{x} + \frac{1}{2} - 2R$

When R is an integer, this yields a non-integer; since the original objective function is quadratic and concave, $\overline{x} - 2R$ and $\bar{x} + 1 - 2R$ will both be best replies.

For $R = 0$, we obtain a corner solution at $x_i^* = \overline{x}$.

This solves the problem for $L1$. Since Lk best-replies to $L(k-1)$, we can solve for best replies for $L2$, $L3$, and so on, and obtain that $\overline{x} - 2R - k + 1$ and $\bar{x} + 2 - 2R - k$ will be best replies for Lk.

B.2 Experimental Instructions

Experimental Implementation

At the beginning of each session, paper instructions were distributed and read aloud; subjects followed along and could use a pen or pencil to write notes on the instructions. Also, an electronic summary of the instructions appeared on-screen in each round. Copies or screenshots of all materials used in the experiment are provided in the supplementary material. After the instructions, subjects completed a comprehension quiz. The experimenter checked the answers privately, and when they encountered incorrect answers, the experimenter pointed the subject to the relevant part of the instructions and gave the subject the opportunity to revise their answers. After all subjects had answered all questions correctly, the experiment commenced. The experiment was conducted through a web interface based on oTree (Chen et al. 2016), with a 10 second forced delay before a subject could submit their choice in each round.

Instructions

Instructions

Overview

- You are going to take part in an experimental study of decision making.
- During the experiment, you are not allowed to talk or communicate with other participants.
- If at any time you have any questions, please raise your hand and the experimenter will come to your desk to answer it.
- Your earnings in the experiment will depend on your choices, the choices of other participants, and an element of chance. By following the instructions and making decisions carefully, you may earn a considerable amount of money.
- At the end of the experiment, the number of Experimental Currency Units (ECU) that you earn will be converted to Canadian dollars at the exchange rate 1 ECU = $0.10.
- You will be given a $7 show up payment in addition to your earnings for the experiment.
- Your total earnings will be paid to you in cash today at the end of this experiment.

General Description

- You will be randomly and anonymously matched with another participant in the room for the duration of the experiment. Your identity will never be revealed to your opponent, and their identity will never be revealed to you.
- The experiment consists of a number of rounds. At the end of the experiment, one round will be randomly selected, and the decisions that you and that other participant made in that round will determine the amount earned by each of you.
    - Since you do not know which round will be selected to determine your payment, you should treat each round as if it will determine your final payment.
- The choices that you and the other subject will make, and the corresponding results, will not be communicated to you at the end of each round.
- At the end of the whole experiment, you will be informed of your choice, the other person's choice and the result of only the payment round.

Task and Earnings

- In each round, at the same time, you and the person you are matched with will each choose a number or "claim" between a specified minimum claim and a specified maximum claim (inclusive).
    - The minimum and maximum claim <u>will be different in each round</u>.
- If the claims are equal, then you and the other person each receive the amount claimed.
- If the claims are not equal, then each of you receives the lower of the two claims. In addition, the person who makes the lower claim earns a reward, and the person with the higher claim pays a penalty of the same amount as the reward.
    - The amount of the reward/penalty <u>will change twice</u> during the experiment, and you will be notified in advance of these changes.
- Thus you will earn an amount that equals the lower of the two claims, plus a reward if you are the person making the lower claim, or minus a penalty if you are the person making the higher claim. There is no penalty or reward if the two claims are exactly equal, in which case each person receives what they claimed.

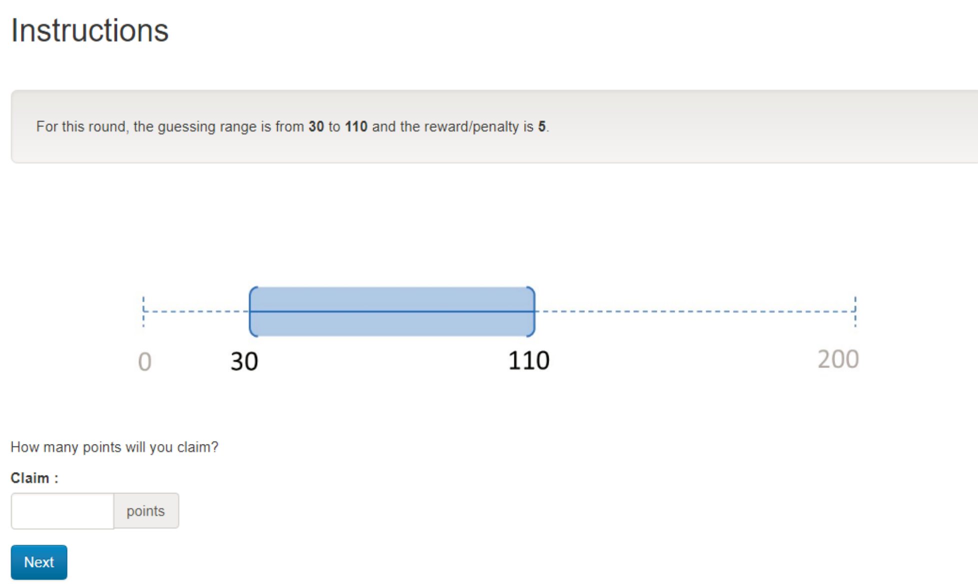
Example

Suppose that your claim is X and the other's claim is Y, and the reward/penalty is T.

- If X = Y (both claim the same amount), you get X, and the other gets Y.
- If X > Y (your claim is higher), you get Y minus T, and the other gets Y plus T.
- If X < Y (your claim is lower), you get X plus T, and the other gets X minus T.

Interface

The following screenshot shows how the minimum and maximum claims, the penalty, and your claim input box will be displayed on each decision screen.



Summary

- You have been randomly and anonymously paired with another participant.
- In each round, you and the other participant each simultaneously and independently choose a number between the round's minimum and maximum claim (inclusive).
- If both choose the same number, then this amount will be paid to both.
- If you choose different amounts, then the lower amount will be paid to both. Additionally, the one with the lower claim will receive a reward; the one with the higher claim will receive a penalty.
- The reward and penalty are the same magnitude, and this will be specified each round.
- You will be paid your earnings in cash at the end of the experiment based on one randomly selected round. Your actual decisions and those of the participant with whom you are paired in that round will determine your earnings for the experiment.

During the experiment, you are not permitted to speak or communicate with the other participants. If you have a question while the experiment is going on, please raise your hand and the experimenter will come to your desk to answer it.

Comprehension Quiz

To verify your comprehension of the instructions, please complete the following comprehension quiz. Your answers will not affect your earnings in any way. We just want to ensure that you understand how the experiment works and how your earnings will be calculated. We will come around and check your responses.

For questions Q1-Q5 below, suppose that the minimum claim is 20, the maximum claim is 200, and the reward/penalty is 5.

Q1. What is the *highest number* that you can claim? _____________

Q2. What is the *lowest number* that you can claim? _____________

Below, write any claim for your opponent in the first blank, and any claim of your own in the second blank, such that your claim is *higher than* your opponent's claim. You will use these numbers to answer Q3, Q4, and Q5.

Suppose that your opponent claims ____ and you claim ____.

Q3. The lower of the two claims is: _____________

Q4. Your opponent will earn: _____________

Q5. You will earn: _____________

For questions Q6-Q10 below, suppose that the minimum claim is 80, the maximum claim is 120, and the reward/penalty is 10.

Q6. What is the *highest number* that you can claim? _____________

Q7. What is the *lowest number* that you can claim? _____________

Below, write any claim for your opponent in the first blank, and any claim of your own in the second blank, such that your claim is *equal to* your opponent's claim. You will use these numbers to answer Q8, Q9, and Q10.

Suppose that your opponent claims ____ and you claim ____.

Q8. The lower of the two claims is: _____________

Q9. Your opponent will earn: _____________

Q10. You will earn: _____________

For questions Q11-Q15 below, suppose that the minimum claim is 70, the maximum claim is 150, and the reward/penalty is 2.

Q11. What is the *highest number* that you can claim? _____________

Q12. What is the *lowest number* that you can claim? _____________

Below, write any claim for your opponent in the first blank, and any claim of your own in the second blank, such that your claim is *lower than* your opponent's claim. You will use these numbers to answer Q13, Q14, and Q15.

Suppose that your opponent claims ____ and you claim ____.

Q13. The lower of the two claims is: _____________

Q14. Your opponent will earn: _____________

Q15. You will earn: _____________

B.3 Classification Procedure

We apply the following steps for each individual subject. We classify each subject to the best fitting model whenever we can reject the null hypothesis of random or naive behavior according to the procedure we describe below. All tests use the 5% significance level.

1. Fit all 15 models (12 strategic and 3 non-strategic) using the maximum likelihood estimation.
2. Find the model with the highest log-likelihood (labeled m herein).
   (a) In case of a tie with uniform randomization model, select uniform randomization model.
   (b) In case of a tie with one of the other two non-strategic models, select the non-strategic model.
3. Compare the best-fitting model m with the non-strategic model of uniform randomization using a likelihood ratio test.
4. If we do not reject H_0 in step 3, we assign this subject to uniform randomization.
5. When we reject H_0 in step 3:
   (a) If best-fitting model m is a non-strategic model, assign that subject to m.
   (b) If best-fitting model m is Nash, assign that subject to m.
   (c) If best-fitting model m is a strategic model Lk, NIk, or QRE. Compare m to the corresponding non-strategic model.
      i. Use a likelihood ratio test[1] when the non-strategic model is nested within a strategic model, e.g. non-strategic model of playing the top with action trembles versus a best-fitting strategic model level k ($k \geq 0$) with $L0$ as the top and action trembles.
      ii. Use Vuong tests when a non-strategic model is not nested within a strategic model, e.g. non-strategic model of playing the top with action trembles versus a best-fitting strategic model level k ($k \geq 0$) with $L0$ as the top and payoff trembles.
      iii. If we cannot reject the null hypothesis in likelihood ratio tests or in Vuong tests, classify that subject to the best fitting non-strategic model that is not rejected. If we reject the null in each of these tests, classify the subject to m.
6. Among subjects classified to a strategic model Lk or NIk, further classify subjects into strategic ($k \geq 1$) and non-strategic based ($k = 0$). A subject's level distribution is modeled as a truncated Poisson distribution with parameter τ[2]. The probability of a player being type k is $f(k;\tau) = \frac{(\frac{e^{-\tau} * \tau^k}{k!})}{\sum_l (\frac{e^{-\tau} * \tau^l}{l!})}$. The lower τ is, the higher the probability

[1]This is equivalent to using a Vuong test when the two models are nested.

[2]Our structural estimation approach is similar to that of Goeree et al. (2018).

that a subject is $L0$ is. $\tau < 0.5$ assigns more than 50% weight on a subject being $L0$. For a subject with $\tau < 0.5$, we classify that subject as non-strategic under the corresponding Lk or NIk and for a subject with $\tau \geq 0.5$,we classify that subject as strategic under the corresponding Lk or NIk.

7. If there is a tie in the likelihood ratios between two strategic models, classify the subject to the model with the fewest parameters.

Our statistical analysis was conducted in R (R Core Team 2017). Maximum likelihood estimations used the 'maxLik' command in the 'maxLik' package (Henningsen and Toomet 2011) and the 'fsolve' command in the 'pracma' package (Borchers 2018). Vuong tests used the 'vuongtest' command in the 'nonnest2' package (Merkle and You 2020). Likelihood ratio tests used the 'lrtest' command in the 'lmtest' package (Zeileis and Hothorn 2002). Kolmogorov-Smirnov test used the 'ks.unif.test' command in the 'spgs' package (Hart and Martínez 2018). Paired samples Wilcoxon test used the 'wilcox.test' command in the 'MASS' package (Venables and Ripley 2002). Figures made using the 'PlotRelativeFrequency' command in the 'HistogramTools' package (Stokely, 2015) and using 'ggplot2' (Wickham, 2009).

B.4 Estimation Results

We estimate 15 models for each subject. The estimation results are recorded below. Note that a model of uniform randomization play yields a log-likelihood of -145 for all subjects, we therefore omit a column for uniform randomization.

"S" denotes subject, going from 1 to 60 for 60 subjects in our data set. "Para" denotes estimated parameters: ϵ for action tremble specification (uniform error), γ for payoff tremble specification (logit error), and τ for level distribution. "LL" stands for log-likelihood.

For 14 models, we list the abbreviations and parameters of interest below.

Table B.1: Abbreviations

Abbreviation	Model	Parameters
T	Non-strategic play at the top	ϵ
M	Non-strategic play at the middle	ϵ
N_at	Nash equilibrium with action trembles	ϵ
N_pt	Nash equilibrium with payoff trembles	γ
QRE	Quantal response equilibrium	γ
LK_at_T	Level k with action trembles, $L0$ at the top	ϵ, τ
LK_pt_T	Level k with payoff trembles, $L0$ at the top	γ, τ
LK_at_M	Level k with action trembles, $L0$ at the middle	ϵ, τ
LK_pt_M	Level k with payoff trembles, $L0$ at the middle	γ, τ
LK_at_U	Level k with action trembles, $L0$ as uniform randomization	ϵ, τ
LK_pt_U	Level k with payoff trembles, $L0$ as uniform randomization	γ, τ
NI_T	Noisy introspection $L0$ at the top	γ, τ
NI_M	Noisy introspection $L0$ at the middle	γ, τ
NI_U	Noisy introspection $L0$ as uniform randomization	γ, τ

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
1	ε or γ	0.672	1	0.975	0.046	0.076	0.592	0.07	1	0.015	0	0.057	0.078	0.076	0.116
		(0.087)	(0.183)	(0.033)	(0.110)	(0.016)	(0.095)	(0.016)	(0.182)	(0.014)	(2.885)	(0.025)	(0.019)	(Inf)	(0.027)
	τ						0.312	1.205	1	5.797	0.059	2.372	0.983	20	3.104
							(0.199)	(0.297)	(NA)	(3.493)	(0.176)	(1.218)	(0.228)	(Inf)	(1.258)
	LL	-115.261	-145.085	-144.433	-145.017	-129.03	-113.243	-92.706	-145.085	-144.522	-142.407	-131.854	-95.815	-129.03	-126.847
2	ε or γ	1	0.841	1	0	0.006	1	0	0.841	0.027	1	0.002	0.006	0.011	0.006
		(0.183)	(0.069)	(0.183)	(0.167)	(0.011)	(0.182)	(0.005)	(0.067)	(0.015)	(0.034)	(0.006)	(0.000)	(0.012)	(0.000)
	τ						1	12.181	0	1.896	1.458	9.262	10.547	1.788	4.058
							(NA)	(6.113)	(NA)	(0.455)	(NA)	(NA)	(NA)	(0.427)	(NA)
	LL	-145.085	-134.932	-145.085	-145.085	-144.957	-145.085	-145.085	-134.933	-133.052	-145.085	-145.01	-144.959	-134.593	-144.958
3	ε or γ	0	1	1	0	0.1	0	0	1	0.024	1	0.049	0	0.108	NA
		(0.184)	(0.183)	(0.183)	(0.167)	(0.023)	(0.156)	(0.039)	(0.182)	(0.014)	(NA)	(0.012)	(0.039)	(0.025)	(NA)
	τ						0	0	1	5.343	1.458	3.991	0	13.785	NA
							(NA)	(0.002)	(NA)	(2.442)	(NA)	(1.655)	(0.001)	(8.394)	(NA)
	LL	0	-145.085	-145.085	-145.085	-123.985	0	0	-145.085	-143.641	-145.085	-128.966	0	-123.941	NA
4	ε or γ	0	1	1	0	0.1	0	0	1	0.024	1	0.049	0	0.108	NA
		(0.184)	(0.183)	(0.183)	(0.167)	(0.023)	(0.156)	(0.039)	(0.182)	(0.014)	(NA)	(0.012)	(0.039)	(0.025)	(NA)
	τ						0	0	1	5.343	1.458	3.991	0	13.785	NA
							(NA)	(0.002)	(NA)	(2.442)	(NA)	(1.655)	(0.001)	(8.394)	(NA)
	LL	0	-145.085	-145.085	-145.085	-123.985	0	0	-145.085	-143.641	-145.085	-128.966	0	-123.941	NA
5	ε or γ	1	1	0	3.737	0	1	0	1	0	1	0	0	0	0
		(0.183)	(0.183)	(0.184)	(23.727)	(0.006)	(0.182)	(0.005)	(0.182)	(0.000)	(NA)	(NA)	(0.006)	(0.006)	(0.000)
	τ						1	6	1	7.536	1.458	3	5.812	8.622	1.001
							(NA)	(3.445)	(NA)	(NA)	(NA)	(NA)	(3.568)	(5.932)	(NA)
	LL	-145.085	-145.085	0	0	-145.085	-145.085	-145.159	-145.085	-145.101	-145.085	-145.085	-145.175	-145.09	-145.085
6	ε or γ	0.975	0.907	0.37	0.925	0	0.975	0	0.907	0	1	0	0	0	0
		(0.033)	(0.055)	(0.089)	(0.090)	(0.007)	(0.028)	(0.005)	(0.053)	(0.012)	(NA)	(0.000)	(0.008)	(0.008)	(0.000)
	τ						0	3.672	0	4.864	1.458	3.001	3.676	2.378	1.015
							(NA)	(1.332)	(NA)	(NA)	(NA)	(NA)	(1.475)	(0.598)	(NA)
	LL	-144.433	-140.125	-72.999	-66.951	-145.085	-144.433	-144.433	-140.125	-143.262	-145.085	-145.085	-144.433	-140.125	-145.085
7	ε or γ	0.672	1	1	0	0.1	0.672	0.08	1	0.024	0	0.168	0.19	0.128	0.127
		(0.087)	(0.183)	(0.183)	(0.167)	(0.021)	(0.086)	(0.018)	(0.182)	(0.014)	(0.631)	(0.080)	(0.102)	(0.000)	(0.000)
	τ						0	1.265	1	5.343	0.259	1.359	1.517	18.918	11.962
							(NA)	(0.320)	(NA)	(2.679)	(0.182)	(0.412)	(0.525)	(NA)	(NA)
	LL	-115.657	-145.085	-145.085	-145.085	-119.416	-115.657	-90.086	-145.085	-143.641	-125.142	-127.922	-88.715	-118.852	-118.859
8	ε or γ	1	0.976	0.706	0.313	0	1	0	0.976	0	1	0	0	0	0
		(0.183)	(0.033)	(0.084)	(0.040)	(0.008)	(0.182)	(0.005)	(0.028)	(0.012)	(0.000)	(0.000)	(0.008)	(0.009)	(NA)
	τ						1	6	0	3.727	1.458	6	6.813	5.875	1.057
							(NA)	(4.231)	(NA)	(1.445)	(NA)	(NA)	(6.037)	(NA)	(NA)
	LL	-145.085	-144.565	-120.133	-110.446	-145.085	-145.085	-145.159	-144.566	-144.565	-145.085	-145.085	-145.118	-144.919	-145.085

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
9	ε or γ	0.336	0.773	1	0.000	0.100	0.336	0.000	0.745	0.110	1	0.084	0.000	0.249	0.252
		(0.087)	(0.078)	(0.183)	(0.167)	(0.023)	(0.086)	(0.009)	(0.082)	(0.034)	(0.000)	(0.019)	(0.010)	(0.060)	(0.084)
	τ						0.000	0.399	0.109	1.410	1.458	10.235	0.460	2.007	3.625
							(NA)	(0.127)	(0.129)	(0.318)	(NA)	(3.255)	(0.147)	(0.367)	(0.685)
	LL	-67.376	-127.544	-145.085	-145.085	-117.763	-67.376	-67.380	-126.550	-117.082	-145.085	-118.412	-67.444	-103.232	-113.285
10	ε or γ	0.907	0.941	0.740	0.254	0.016	0.906	0.002	0.941	0.000	0.001	0.015	0.006	0.012	0.153
		(0.055)	(0.046)	(0.081)	(0.032)	(0.011)	(0.053)	(0.005)	(0.043)	(0.012)	(6.108)	(NA)	(0.010)	(0.014)	(0.128)
	τ						0.000	2.382	0.000	2.832	0.023	2.249	2.393	2.985	0.465
							(NA)	(0.600)	(NA)	(0.779)	(0.146)	(NA)	(0.605)	(0.852)	(0.301)
	LL	-139.817	-142.609	-124.208	-124.512	-144.286	-139.817	-139.717	-142.609	-142.609	-144.574	-143.435	-139.666	-142.291	-142.829
11	ε or γ	0.908	1	0.101	1.315	0.000	0.908	0.000	1.000	0.000	1	0.000	0.000	0.000	0.000
		(0.055)	(0.183)	(0.055)	(0.126)	(0.007)	(0.053)	(0.005)	(0.182)	(0.000)	(NA)	(0.000)	(0.007)	(0.007)	(0.001)
	τ						0.000	2.271	1.000	8.799	1.458	2.000	2.378	10.206	1.002
							(NA)	(0.558)	(NA)	(NA)	(NA)	(NA)	(0.688)	(5.932)	(NA)
	LL	-140.291	-145.085	-24.040	-22.980	-145.085	-140.291	-140.310	-145.085	-145.089	-145.085	-145.085	-140.291	-145.086	-145.085
12	ε or γ	0.336	1	1	0.000	0.100	0.336	0.138	1.000	0.024	1	0.050	0.178	0.111	NA
		(0.087)	(0.183)	(0.183)	(0.167)	(0.022)	(0.086)	(0.042)	(0.182)	(0.014)	(NA)	(0.013)	(0.087)	(0.025)	(NA)
	τ						0.000	0.453	1.000	5.347	1.458	3.699	0.423	17.196	NA
							(NA)	(0.147)	(NA)	(2.442)	(NA)	(1.682)	(0.144)	(NA)	(NA)
	LL	-67.376	-145.085	-145.085	-145.085	-122.120	-67.376	-48.758	-145.085	-143.641	-145.085	-128.826	-49.632	-122.012	NA
13	ε or γ	0.906	1	1	0.000	0.100	0.341	0.118	1.000	0.024	1	0.055	0.419	0.141	NA
		(0.055)	(0.183)	(0.183)	(0.167)	(0.021)	(0.098)	(0.025)	(0.182)	(0.014)	(NA)	(0.017)	(0.047)	(0.030)	(NA)
	τ						4.628	2.802	1.000	5.347	1.330	3.109	2.548	25.348	NA
							(0.578)	(0.737)	(NA)	(2.442)	(NA)	(1.446)	(0.356)	(8.402)	(NA)
	LL	-139.752	-145.085	-145.085	-145.085	-118.305	-112.211	-97.283	-145.085	-143.641	-145.085	-128.342	-88.446	-117.281	NA
14	ε or γ	0.940	0.975	0.672	0.280	0.000	0.940	0.000	0.975	0.000	1	0.000	0.000	0.000	0.000
		(0.046)	(0.033)	(0.087)	(0.034)	(0.008)	(0.043)	(0.005)	(0.028)	(0.008)	(NA)	(0.001)	(0.008)	(0.009)	(0.000)
	τ						0.000	2.000	0.000	13.310	1.458	2.000	2.797	3.648	1.021
							(NA)	(0.449)	(NA)	(NA)	(NA)	(NA)	(0.853)	(1.266)	(NA)
	LL	-142.352	-144.433	-115.526	-117.843	-145.085	-142.352	-143.188	-144.433	-145.085	-145.085	-145.085	-142.353	-144.433	-145.085
15	ε or γ	0.975	1	1	0.000	0.031	0.918	0.066	0.682	0.091	1	0.044	0.031	0.153	NA
		(0.033)	(0.183)	(0.183)	(0.167)	(0.009)	(0.056)	(0.026)	(0.088)	(0.023)	(NA)	(0.015)	(0.009)	(0.031)	(NA)
	τ						0.655	57.318	1.003	2.431	1.458	34.889	12.793	1.778	NA
							(0.531)	(7.428)	(0.328)	(0.586)	(NA)	(4.346)	(5.932)	(0.306)	(NA)
	LL	-144.433	-145.085	-145.085	-145.085	-140.650	-142.499	-136.140	-125.901	-136.049	-145.085	-137.797	-140.651	-135.969	NA
16	ε or γ	1	0.973	1	0.000	0.048	0.931	0.064	0.973	0.062	0.000	0.052	0.048	0.082	0.048
		(0.183)	(0.033)	(0.183)	(0.167)	(0.011)	(0.000)	(0.020)	(0.030)	(0.018)	(1.890)	(0.015)	(0.011)	(0.026)	(0.000)
	τ						8.785	61.477	0.000	4.231	0.092	25.374	13.152	2.978	6.844
							(NA)	(6.567)	(NA)	(2.028)	(0.180)	(4.486)	(NA)	(0.691)	(NA)
	LL	-145.085	-144.207	-145.085	-145.085	-135.629	-144.467	-135.912	-144.207	-136.103	-140.257	-133.594	-135.630	-132.649	-135.624

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
17	ε or γ	0.840	0.975	0.942	0.140	0.017	0.840	0.000	0.975	0.064	1	0.059	0.000	0.020	NA
		(0.069)	(0.033)	(0.046)	(0.049)	(0.010)	(0.067)	(0.006)	(0.029)	(0.022)	(0.000)	(0.020)	(0.009)	(0.012)	(NA)
	τ						0.000	2.000	0.000	11.124	1.469	58.174	1.829	3.396	NA
							(NA)	(0.483)	(NA)	(3.297)	(NA)	(3.549)	(0.435)	(1.176)	(NA)
	LL	-133.762	-144.433	-142.766	-142.919	-144.058	-133.762	-133.835	-144.433	-138.867	-145.085	-139.410	-133.762	-143.173	NA
18	ε or γ	1	0.940	0.973	0.046	0.000	1.000	0.000	0.940	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.046)	(0.033)	(0.110)	(0.009)	(0.182)	(0.005)	(0.043)	(0.000)	(0.000)	(0.000)	(0.009)	(0.010)	(NA)
	τ						1.000	6.000	0.000	14.955	1.458	5.992	6.229	2.810	1.106
							(NA)	(3.445)	(NA)	(NA)	(NA)	(NA)	(3.474)	(0.757)	(NA)
	LL	-145.085	-142.216	-144.259	-145.017	-145.085	-145.085	-145.159	-142.216	-145.085	-145.085	-145.085	-145.144	-142.216	-145.085
19	ε or γ	1	0.842	1	0.000	0.004	1.000	0.000	0.814	0.140	1	0.000	0.004	0.019	0.004
		(0.183)	(0.069)	(0.183)	(0.167)	(0.011)	(0.182)	(0.000)	(0.074)	(0.034)	(0.000)	(0.006)	(0.011)	(0.014)	(0.011)
	τ						1.000	8.987	0.146	15.998	1.458	5.995	10.463	1.625	3.147
							(NA)	(NA)	(0.175)	(2.163)	(NA)	(6.160)	(NA)	(0.421)	(5.939)
	LL	-145.085	-135.277	-145.085	-145.085	-145.033	-145.085	-145.089	-134.372	-126.566	-145.085	-145.085	-145.034	-134.316	-145.033
20	ε or γ	1	1	0.942	0.169	0.000	1.000	0.000	1.000	0.109	1	0.000	0.000	0.000	0.000
		(0.183)	(0.183)	(0.046)	(0.040)	(0.009)	(0.183)	(0.005)	(0.000)	(0.033)	(NA)	(0.000)	(0.009)	(0.009)	(0.000)
	τ						0.970	7.000	3.802	34.117	1.458	5.994	7.444	8.100	1.191
							(NA)	(4.218)	(NA)	(3.641)	(NA)	(NA)	(5.968)	(5.932)	(NA)
	LL	-145.085	-145.085	-142.766	-140.592	-145.085	-145.085	-145.112	-145.085	-137.110	-145.085	-145.085	-145.102	-145.094	-145.085
21	ε or γ	0.975	1	0.235	1.090	0.000	0.975	0.000	1.000	0.000	1	0.000	0.000	0.000	0.000
		(0.033)	(0.183)	(0.078)	(0.094)	(0.007)	(0.028)	(0.005)	(0.182)	(NA)	(NA)	(0.000)	(0.006)	(0.007)	(0.001)
	τ						0.000	6.000	1.000	6.372	1.458	6.001	5.813	12.073	1.003
							(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(5.932)	(NA)
	LL	-144.433	-145.085	-50.291	-46.819	-145.085	-144.433	-144.896	-145.085	-145.136	-145.085	-145.085	-144.865	-145.085	-145.085
22	ε or γ	0.538	1	1	0.000	0.100	0.538	0.071	1.000	0.024	0.000	0.055	0.108	0.122	0.122
		(0.092)	(0.183)	(0.183)	(0.167)	(0.021)	(0.091)	(0.018)	(0.182)	(0.014)	(0.848)	(0.017)	(0.032)	(0.000)	(0.000)
	τ						0.000	0.819	1.000	5.335	0.192	3.015	0.901	15.713	12.761
							(NA)	(0.214)	(NA)	(2.442)	(0.177)	(1.420)	(0.237)	(NA)	(NA)
	LL	-97.813	-145.085	-145.085	-145.085	-120.465	-97.813	-79.275	-145.085	-143.641	-131.300	-128.534	-78.004	-120.085	-120.083
23	ε or γ	1	1	0.101	1.315	0.000	1.000	0.000	1.000		1	0.000	0.000	0.000	0.000
		(0.183)	(0.183)	(0.055)	(0.126)	(0.006)	(0.182)	(0.005)	(0.182)	()	(NA)	(0.000)	(0.007)	(0.006)	(0.001)
	τ						1.000	6.000	1.000		1.458	6.000	6.252	16.835	1.003
							(NA)	(4.231)	(NA)	()	(NA)	(NA)	(6.318)	(NA)	(NA)
	LL	-145.085	-145.085	-24.622	-22.980	-145.085	-145.085	-145.159	-145.085		-145.085	-145.085	-145.143	-145.085	-145.085
24	ε or γ	1	1	1	0.000	0.000	1.000	0.000	1.000	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.183)	(0.183)	(0.167)	(0.007)	(0.182)	(0.005)	(0.183)	(NA)	(NA)	(0.000)	(0.007)	(0.007)	(0.001)
	τ						1.000	6.000	1.000	6.848	1.458	6.000	5.668	7.761	1.052
							(NA)	(2.979)	(4.195)	(NA)	(NA)	(NA)	(3.603)	(5.932)	(NA)
	LL	-145.085	-145.085	-145.085	-145.085	-145.085	-145.085	-145.159	-145.085	-145.117	-145.085	-145.085	-145.189	-145.098	-145.085

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
25	ε or γ	1	1	1	0.000	0.000	1.000	0.000	1.000	0.093	1	0.000	0.000	0.000	0.000
		(0.183)	(0.183)	(0.183)	(0.167)	(0.009)	(0.182)	(0.005)	(NA)	(0.035)	(NA)	(NA)	(0.009)	(0.009)	(0.000)
	τ						1.000	10.000	3.368	29.527	1.458	6.000	12.317	9.000	1.114
							(NA)	(NA)	(NA)	(4.841)	(NA)	(NA)	(NA)	(5.932)	(NA)
	LL	-145.085	-145.085	-145.085	-145.085	-145.085	-145.085	-145.086	-145.085	-139.527	-145.085	-145.085	-145.085	-145.088	-145.085
26	ε or γ	0.470	1	1	0.000	0.076	0.470	0.052	1.000	0.024	1	0.040	0.059	0.076	NA
		(0.092)	(0.183)	(0.183)	(0.167)	(0.015)	(0.091)	(0.015)	(0.182)	(0.014)	(NA)	(0.011)	(0.017)	(0.000)	(NA)
	τ						0.000	0.659	1.000	5.335	1.458	3.971	0.645	12.499	NA
							(NA)	(0.181)	(NA)	(2.442)	(NA)	(2.101)	(0.176)	(NA)	(NA)
	LL	-88.264	-145.085	-145.085	-145.085	-128.056	-88.265	-75.994	-145.085	-143.641	-145.085	-132.624	-76.198	-128.058	NA
27	ε or γ	0.973	0.773	0.808	0.213	0.000	0.973	0.000	0.745	0.000	1	0.001	0.000	0.000	0.000
		(0.033)	(0.078)	(0.074)	(0.033)	(0.013)	(0.030)	(0.005)	(0.082)	(0.013)	(0.000)	(NA)	(0.011)	(0.015)	(0.019)
	τ						0.000	3.443	0.109	1.285	1.458	9.381	3.880	1.482	1.200
							(NA)	(1.105)	(0.128)	(0.289)	(NA)	(NA)	(1.624)	(0.343)	(6.240)
	LL	-144.207	-127.414	-131.819	-133.859	-145.085	-144.207	-144.216	-126.420	-127.602	-145.085	-145.056	-144.230	-127.414	-145.085
28	ε or γ	0.773	0.907	0.973	0.127	0.035	0.773	0.011	0.907	0.009	0.045	0.087	0.016	0.036	0.235
		(0.078)	(0.055)	(0.033)	(0.054)	(0.010)	(0.076)	(0.006)	(0.053)	(0.013)	(4.149)	(0.062)	(0.008)	(0.014)	(0.173)
	τ						0.000	1.507	0.000	2.347	0.063	1.234	1.498	2.756	1.034
							(NA)	(0.353)	(NA)	(0.582)	(0.280)	(0.627)	(0.349)	(0.684)	(0.501)
	LL	-127.503	-140.125	-144.207	-143.586	-140.084	-127.504	-125.662	-140.125	-139.862	-142.410	-137.153	-125.725	-136.634	-135.788
29	ε or γ	0.000	1	1	0.000	0.100	0.000	0.000	1.000	0.024	1	0.049	0.000	0.108	NA
		(0.184)	(0.183)	(0.183)	(0.167)	(0.023)	(0.156)	(0.039)	(0.182)	(0.014)	(NA)	(0.012)	(0.039)	(0.025)	(NA)
	τ						0.000	0.000	1.000	5.343	1.458	3.991	0.000	13.785	NA
							(NA)	(0.002)	(NA)	(2.442)	(NA)	(1.655)	(0.001)	(8.394)	(NA)
	LL	0.000	-145.085	-145.085	-145.085	-123.985	0.000	0.000	-145.085	-143.641	-145.085	-128.966	0.000	-123.941	NA
30	ε or γ	0.773	0.975	1	0.000	0.100	0.546	0.054	0.975	0.025	0.000	0.066	0.110	0.116	NA
		(0.078)	(0.033)	(0.183)	(0.167)	(0.019)	(0.094)	(0.012)	(0.028)	(0.014)	(1.859)	(0.035)	(0.028)	(0.022)	(NA)
	τ						0.570	1.507	0.000	3.515	0.093	2.515	1.752	6.497	NA
							(0.219)	(0.362)	(NA)	(1.164)	(0.181)	(1.524)	(0.384)	(1.585)	(NA)
	LL	-127.153	-144.433	-145.085	-145.085	-120.359	-111.064	-106.715	-144.433	-142.713	-139.806	-128.028	-102.942	-119.852	NA
31	ε or γ	0.974	1	1	0.000	0.029	0.974	0.009	0.870	0.039	0.001	0.018	0.028	0.029	NA
		(0.033)	(0.183)	(0.183)	(0.167)	(0.009)	(0.029)	(0.005)	(0.079)	(0.016)	(NA)	(0.007)	(0.000)	(NA)	(NA)
	τ						0.000	3.945	4.178	5.434	0.025	5.479	5.674	19.297	NA
							(NA)	(1.670)	(1.344)	(2.653)	(NA)	(5.933)	(NA)	(NA)	(NA)
	LL	-144.314	-145.085	-145.085	-145.085	-141.124	-144.314	-142.941	-142.659	-141.484	-144.442	-141.315	-141.129	-141.124	NA
32	ε or γ	1	0.370	0.874	0.219	0.000	1.000	0.000	0.305	0.000	1	0.000	0.000	0.000	0.001
		(0.183)	(0.089)	(0.063)	(0.032)	(0.012)	(0.182)	(0.005)	(0.085)	(0.019)	(0.000)	(0.006)	(0.012)	(0.020)	(0.017)
	τ						1.000	8.996	0.093	0.461	1.458	9.999	9.944	0.463	1.103
							(NA)	(6.071)	(0.068)	(0.141)	(NA)	(NA)	(5.932)	(0.141)	(NA)
	LL	-145.085	-73.224	-137.389	-132.780	-145.084	-145.085	-145.088	-68.454	-73.224	-145.085	-145.085	-145.085	-73.224	-145.084

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
33	ε or γ	0.672	0.976	0.941	0.169	0.095	0.643	0.018	0.976	0.024	0.000	0.080	0.028	0.210	0.229
		(0.087)	(0.033)	(0.046)	(0.040)	(0.021)	(0.089)	(0.007)	(0.028)	(0.014)	(0.737)	(0.026)	(0.010)	(0.010)	(0.065)
	τ						0.081	1.137	0.000	3.554	0.225	2.535	1.203	11.029	2.697
							(0.092)	(0.273)	(NA)	(1.188)	(0.182)	(0.830)	(0.296)	(1.505)	(0.655)
	LL	-116.069	-144.565	-142.569	-140.592	-126.501	-114.884	-112.066	-144.566	-142.970	-127.970	-126.031	-111.331	-118.338	-120.846
34	ε or γ	1	1	0.000	3.737	0.000	1.000	0.000	1.000	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.183)	(0.184)	(23.727)	(0.006)	(0.182)	(0.005)	(0.182)	(0.000)	(NA)	(NA)	(0.006)	(0.006)	(0.000)
	τ						1.000	6.000	1.000	7.536	1.458	3.000	5.812	8.622	1.001
							(NA)	(3.445)	(NA)	(NA)	(NA)	(NA)	(3.568)	(5.932)	(NA)
	LL	-145.085	-145.085	0.000	0.000	-145.085	-145.085	-145.159	-145.085	-145.101	-145.085	-145.085	-145.175	-145.090	-145.085
35	ε or γ	0.168	1	1	0.000	NA	0.000	5.965	1.000	0.024	1	0.049	5.965	0.108	0.108
		(0.069)	(0.183)	(0.183)	(0.167)	(NA)	(0.186)	(Inf)	(0.182)	(0.014)	(NA)	(0.012)	(NA)	(0.000)	(0.026)
	τ						0.167	0.167	1.000	5.334	1.458	3.980	0.167	13.797	11.571
							(0.075)	(Inf)	(NA)	(2.679)	(NA)	(1.658)	(0.072)	(NA)	(NA)
	LL	-38.115	-145.085	-145.085	-145.085	NA	-13.959	-13.972	-145.085	-143.641	-145.085	-128.959	-13.972	-123.825	-123.824
36	ε or γ	0.907	1	0.101	0.884	0.000	0.907	0.000	1.000	0.000	1	0.000	0.000	0.000	0.000
		(0.055)	(0.183)	(0.055)	(0.092)	(0.006)	(0.053)	(0.005)	(0.182)	(0.000)	(NA)	(0.000)	(Inf)	(0.000)	(0.000)
	τ						0.000	2.374	1.000	8.336	1.458	3.000	13.692	10.266	1.001
							(NA)	(0.593)	(NA)	(NA)	(NA)	(NA)	(Inf)	(NA)	(NA)
	LL	-139.959	-145.085	-24.401	-71.474	-145.085	-139.959	-139.959	-145.085	-145.092	-145.085	-145.085	-145.084	-145.086	-145.085
37	ε or γ	1	1	1	0.000	0.045	1.000	0.017	1.000	0.075	0.645	0.033	0.045	0.179	0.045
		(0.183)	(0.183)	(0.183)	(0.167)	(0.011)	(Inf)	(0.007)	(0.183)	(0.023)	(0.111)	(0.000)	(Inf)	(0.013)	(0.000)
	τ						1.000	17.907	1.000	18.638	10.687	9.106	18.846	3.874	12.224
							(Inf)	(6.091)	(5.932)	(4.570)	(1.791)	(NA)	(Inf)	(0.580)	(NA)
	LL	-145.085	-145.085	-145.085	-145.085	-137.429	-145.085	-141.098	-145.085	-137.790	-136.294	-136.006	-137.429	-129.010	-137.431
38	ε or γ	0.975	0.873	0.806	0.196	0.000	0.975	0.000	0.873	0.000	0.000	0.000	0.000	0.000	0.000
		(0.033)	(0.063)	(0.074)	(0.035)	(0.010)	(0.028)	(0.005)	(0.061)	(0.013)	(1.970)	(0.005)	(0.009)	(0.011)	(0.004)
	τ						0.000	3.605	0.000	2.072	0.092	3.000	3.681	2.065	1.064
							(NA)	(1.238)	(NA)	(0.497)	(0.188)	(NA)	(1.335)	(0.495)	(NA)
	LL	-144.433	-137.049	-130.250	-136.932	-145.085	-144.433	-144.434	-137.049	-137.049	-140.257	-145.085	-144.433	-137.049	-145.085
39	ε or γ	1	1	1	0.000	0.000	1.000	0.000	1.000	0.242	1	0.000	0.000	0.000	0.000
		(0.183)	(0.183)	(0.183)	(0.167)	(0.010)	(0.182)	(0.005)	(0.183)	(0.099)	(NA)	(0.000)	(0.010)	(0.010)	(0.003)
	τ						1.000	8.220	1.000	24.727	1.458	3.000	6.336	7.992	1.052
							(NA)	(6.061)	(NA)	(1.915)	(NA)	(NA)	(6.220)	(4.195)	(NA)
	LL	-145.085	-145.085	-145.085	-145.085	-145.085	-145.085	-145.093	-145.085	-124.169	-145.085	-145.085	-145.138	-145.095	-145.085
40	ε or γ	0.269	1	0.943	0.169	NA	0.102	0.026	1.000	0.022	1	0.048	0.029	0.176	0.180
		(0.081)	(0.183)	(0.046)	(0.040)	(NA)	(0.056)	(0.012)	(0.182)	(0.014)	(NA)	(0.012)	(0.013)	(0.013)	(0.009)
	τ						0.185	0.309	1.000	5.799	1.458	4.185	0.316	8.151	10.711
							(0.083)	(0.110)	(NA)	(2.977)	(NA)	(2.086)	(0.112)	(1.387)	(2.410)
	LL	-56.039	-145.085	-142.924	-140.592	NA	-37.312	-53.214	-145.085	-143.885	-145.085	-129.494	-53.025	-119.934	-120.844

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
41	ε or γ	0.908	1	1	0.000	NA	0.000	5.965	1.000	0.024	1	0.049	1.242	0.114	0.114
		(0.055)	(0.183)	(0.183)	(0.167)	(NA)	(0.188)	(NA)	(0.182)	(0.014)	(NA)	(0.012)	(0.331)	(0.000)	(0.027)
	τ						1.433	1.431	1.000	5.334	1.458	3.916	0.979	16.500	13.216
							(0.219)	(0.219)	(NA)	(2.679)	(NA)	(1.673)	(0.188)	(NA)	(NA)
	LL	-140.392	-145.085	-145.085	-145.085	NA	-38.610	-38.622	-145.085	-143.641	-145.085	-128.929	-56.119	-122.828	-122.829
42	ε or γ	0.538	1	0.976	0.127	NA	0.204	0.047	1.000	0.033	1	0.045	0.055	0.172	0.174
		(0.092)	(0.183)	(0.033)	(0.054)	(NA)	(0.074)	(0.013)	(0.182)	(0.015)	(NA)	(0.012)	(0.015)	(0.009)	(0.008)
	τ						0.416	0.752	1.000	6.034	1.458	3.934	0.757	10.524	9.107
							(0.132)	(0.194)	(NA)	(3.426)	(NA)	(2.038)	(0.194)	(8.536)	(1.924)
	LL	-98.096	-145.085	-144.565	-143.586	NA	-62.787	-85.482	-145.085	-142.385	-145.085	-130.740	-85.267	-121.992	-121.337
43	ε or γ	0.235	1	1	0.000	NA	0.000	5.965	1.000	0.024	1	0.049	5.965	0.109	NA
		(0.078)	(0.183)	(0.183)	(0.167)	(NA)	(0.186)	(6.849)	(0.182)	(0.014)	(NA)	(0.012)	(6.849)	(0.026)	(NA)
	τ						0.233	0.233	1.000	5.334	1.458	3.987	0.233	13.922	NA
							(0.088)	(0.088)	(NA)	(2.679)	(NA)	(1.802)	(0.088)	(8.394)	(NA)
	LL	-50.175	-145.085	-145.085	-145.085	NA	-17.187	-17.205	-145.085	-143.641	-145.085	-128.962	-17.205	-123.757	NA
44	ε or γ	0.976	1	1	0.000	NA	0.845	0.100	1.000	0.024	0.000	0.174	0.153	0.104	NA
		(0.033)	(0.183)	(0.183)	(0.167)	(NA)	(0.077)	(0.022)	(0.182)	(0.014)	(0.554)	(0.077)	(0.044)	(0.021)	(NA)
	τ						1.235	7.816	1.000	5.334	0.292	1.326	2.493	8.811	NA
							(0.873)	(1.565)	(NA)	(2.679)	(0.182)	(0.375)	(0.487)	(4.902)	(NA)
	LL	-144.565	-145.085	-145.085	-145.085	NA	-139.352	-106.071	-145.085	-143.641	-121.608	-128.139	-111.607	-121.916	NA
45	ε or γ	1	1	0.000	3.737	0.000	1.000	0.000	1.000	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.183)	(0.184)	(23.727)	(0.006)	(0.182)	(0.005)	(0.182)	(0.000)	(NA)	(NA)	(0.006)	(0.006)	(0.000)
	τ						1.000	6.000	1.000	7.536	1.458	3.000	5.812	8.622	1.001
							(NA)	(3.445)	(NA)	(NA)	(NA)	(NA)	(3.568)	(5.932)	(NA)
	LL	-145.085	-145.085	0.000	0.000	-145.085	-145.085	-145.159	-145.085	-145.101	-145.085	-145.085	-145.175	-145.090	-145.085
46	ε or γ	0.908	1	1	0.000	0.042	0.684	0.012	1.000	0.071	0.000	0.040	0.024	0.204	0.692
		(0.055)	(0.183)	(0.183)	(0.167)	(0.011)	(0.097)	(0.006)	(0.183)	(0.024)	(1.057)	(0.035)	(0.009)	(0.021)	(0.774)
	τ						7.124	2.455	1.000	19.974	0.159	1.936	2.603	4.377	0.586
							(0.989)	(0.637)	(NA)	(6.812)	(0.181)	(1.909)	(0.721)	(0.607)	(0.296)
	LL	-140.249	-145.085	-145.085	-145.085	-138.637	-135.271	-137.788	-145.085	-138.921	-134.168	-138.192	-136.944	-121.160	-137.377
47	ε or γ	1	1	1	0.000	NA	0.000	5.965	1.000	0.024	1	0.049	5.965	0.113	NA
		(0.183)	(0.183)	(0.183)	(0.167)	(NA)	(0.187)	(3.752)	(0.182)	(0.014)	(NA	(0.012)	(3.661)	(0.027)	(NA)
	τ						1.000	1.000	1.000	5.334	1.458	3.953	1.001	16.475	NA
							(0.183)	(0.183)	(NA)	(2.679)	(NA)	(1.602)	(0.183)	(NA)	(NA)
	LL	-145.085	-145.085	-145.085	-145.085	NA	-30.000	-30.077	-145.085	-143.641	-145.085	-128.946	-30.077	-123.137	NA
48	ε or γ	1	0.976	0.067	1.408	0.000	1.000	0.000	0.976	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.033)	(0.046)	(0.149)	(0.006)	(0.182)	(0.000)	(0.028)	(0.007)	(NA)	(0.001)	(0.006)	(0.007)	(0.000)
	τ						1.000	17.969	0.000	12.152	1.458	3.000	7.863	3.712	1.002
							(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(1.333)	(NA)
	LL	-145.085	-144.565	-16.562	-16.187	-145.085	-145.085	-145.085	-144.566	-145.084	-145.085	-145.085	-145.096	-144.566	-145.085

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
49	ε or γ	1	1	1	0.000	0.064	1.000	0.082	0.819	0.084	1	0.067	0.064	0.093	0.066
		(0.183)	(0.183)	(0.183)	(0.167)	(0.013)	(Inf)	(0.000)	(0.075)	(0.022)	(0.000)	(0.017)	(0.013)	(0.025)	(0.000)
	τ						1.000	44.987	1.000	2.433	1.312	14.988	14.634	2.913	6.112
							(Inf)	(NA)	(0.430)	(0.567)	(NA)	(2.841)	(NA)	(0.868)	(NA)
	LL	-145.085	-145.085	-145.085	-145.085	-130.549	-145.085	-129.168	-137.002	-137.727	-145.085	-126.636	-130.549	-129.770	-130.513
50	ε or γ	0.940	1	1	0.000	NA	0.000	5.965	1.000	0.024	1	0.049	5.965	0.113	NA
		(0.046)	(0.183)	(0.183)	(0.167)	(NA)	(0.187)	(Inf)	(0.182)	(0.014)	(NA)	(0.012)	(3.954)	(0.000)	(NA)
	τ						0.933	0.937	1.000	5.334	1.458	3.957	0.933	16.477	NA
							(0.176)	(Inf)	(NA)	(2.679)	(NA)	(1.601)	(0.176)	(NA)	(NA)
	LL	-142.411	-145.085	-145.085	-145.085	NA	-29.932	-30.004	-145.085	-143.641	-145.085	-128.948	-30.004	-123.189	NA
51	ε or γ	1	0.976	0.941	0.087	0.000	1.000	0.000	0.976	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.033)	(0.046)	(0.076)	(0.008)	(0.182)	(0.005)	(0.028)	(0.012)	(0.000)	(NA)	(0.000)	(0.008)	(0.000)
	τ						1.000	7.685	0.000	3.750	1.458	3.001	9.056	3.727	1.063
							(NA)	(6.056)	(NA)	(1.442)	(NA)	(NA)	(NA)	(1.361)	(NA)
	LL	-145.085	-144.565	-142.568	-144.674	-145.085	-145.085	-145.099	-144.566	-144.565	-145.085	-145.085	-145.088	-144.565	-145.085
52	ε or γ	0.336	1	1	0.000	NA	0.034	0.412	1.000	0.024	1	0.050	0.708	0.109	0.109
		(0.087)	(0.183)	(0.183)	(0.167)	(NA)	(0.033)	(0.131)	(0.182)	(0.014)	(NA)	(0.012)	(0.145)	(0.025)	(0.026)
	τ						0.310	0.358	1.000	5.334	1.458	3.889	0.452	13.756	11.863
							(0.103)	(0.114)	(NA)	(2.679)	(NA)	(1.618)	(0.129)	(NA)	(8.395)
	LL	-67.376	-145.085	-145.085	-145.085	NA	-28.512	-38.972	-145.085	-143.641	-145.085	-128.916	-35.933	-123.277	-123.273
53	ε or γ	1	0.840	0.639	0.254	0.000	1.000	0.000	0.840	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.069)	(0.089)	(0.032)	(0.008)	(0.182)	(0.005)	(0.067)	(0.013)	(NA)	(0.000)	(0.008)	(0.010)	(0.000)
	τ						1.000	5.908	0.000	1.679	1.458	3.001	8.798	1.790	1.064
							(NA)	(3.444)	(NA)	(0.383)	(NA)	(NA)	(NA)	(0.416)	(NA)
	LL	-145.085	-133.890	-112.058	-124.512	-145.085	-145.085	-145.166	-133.890	-133.958	-145.085	-145.085	-145.089	-133.895	-145.085
54	ε or γ	1	0.773	1	0.000	0.019	1.000	0.000	0.074	0.647	1	0.034	0.019	0.844	0.019
		(0.183)	(0.078)	(0.183)	(0.167)	(0.008)	(0.182)	(0.005)	(0.050)	(0.101)	(NA)	(0.015)	(0.000)	(0.063)	(0.009)
	τ						1.000	7.315	1.591	1.140	1.458	45.059	11.963	1.013	5.260
							(NA)	(6.057)	(0.244)	(0.233)	(NA)	(4.271)	(NA)	(0.193)	(NA)
	LL	-145.085	-127.719	-145.085	-145.085	-143.339	-145.085	-145.103	-66.695	-65.231	-145.085	-141.390	-143.340	-65.173	-143.334
55	ε or γ	1	0.437	0.672	0.415	0.000	1.000	0.000	0.372	0.000	1	0.000	0.000	0.000	0.000
		(0.183)	(0.091)	(0.087)	(0.083)	(0.010)	(0.182)	(0.000)	(0.089)	(0.018)	(NA)	(0.004)	(0.000)	(0.018)	(0.000)
	τ						1.000	7.973	0.102	0.574	1.458	3.000	10.381	0.574	1.098
							(NA)	(NA)	(0.075)	(0.162)	(NA)	(NA)	(NA)	(0.162)	(NA)
	LL	-145.085	-83.432	-115.657	-98.126	-145.085	-145.085	-145.095	-79.023	-83.432	-145.085	-145.085	-145.086	-83.432	-145.085
56	ε or γ	0.302	0.975	0.807	0.229	NA	0.272	0.000	0.975	0.000	1	0.057	0.000	0.027	0.251
		(0.084)	(0.033)	(0.074)	(0.032)	(NA)	(0.082)	(0.010)	(0.028)	(0.013)	(NA)	(0.068)	(0.010)	(0.012)	(0.230)
	τ						0.042	0.360	0.000	8.090	1.458	1.052	0.360	4.321	0.611
							(0.046)	(0.121)	(NA)	(NA)	(NA)	(0.956)	(0.122)	(1.791)	(0.351)
	LL	-61.202	-144.433	-131.430	-130.544	NA	-59.673	-61.202	-144.433	-145.058	-145.085	-141.023	-61.202	-142.329	-140.102

S	Para	T	M	N_at	N_pt	QRE	LK_at_T	LK_pt_T	LK_at_M	LK_pt_M	LK_at_U	LK_pt_U	NI_T	NI_M	NI_U
57	ε or γ	0.973	0.976	1	0.000	NA	0.973	0.065	0.976	0.037	0.670	0.078	0.118	0.123	0.126
		(0.033)	(0.033)	(0.183)	(0.167)	(NA)	(0.030)	(0.026)	(0.028)	(0.015)	(0.155)	(0.017)	(0.020)	(0.000)	(0.030)
	τ						0.000	25.929	0.000	3.472	1.097	6.637	6.769	17.777	5.474
							(NA)	(14.554)	(NA)	(1.084)	(0.709)	(3.520)	(2.223)	(NA)	(4.880)
	LL	-144.207	-144.565	-145.085	-145.085	NA	-144.207	-124.797	-144.566	-141.047	-133.970	-119.318	-118.652	-118.631	-118.673
58	ε or γ	0.875	0.976	0.538	0.331	0.000	0.875	0.000	0.976	0.000	1	0.000	0.000	0.000	0.000
		(0.063)	(0.033)	(0.092)	(0.044)	(0.007)	(0.060)	(0.005)	(0.028)	(0.012)	(NA)	(NA)	(0.007)	(0.007)	(NA)
	τ						0.000	2.076	0.000	3.799	1.458	3.000	2.103	3.746	1.006
							(NA)	(0.498)	(NA)	(1.491)	(NA)	(NA)	(0.560)	(1.445)	(NA)
	LL	-137.704	-144.565	-98.495	-107.230	-145.085	-137.704	-137.704	-144.566	-144.566	-145.085	-145.085	-137.705	-144.565	-145.085
59	ε or γ	0.873	0.975	1	0.000	NA	0.873	0.083	0.975	0.027	0.000	0.285	0.359	NA	2.136
		(0.063)	(0.033)	(0.183)	(0.167)	(NA)	(0.061)	(0.018)	(0.028)	(0.014)	(0.398)	(0.114)	(0.027)	(NA)	(0.290)
	τ						0.000	11.246	0.000	3.425	0.392	1.374	5.102	NA	1.222
							(NA)	(2.834)	(NA)	(1.068)	(0.181)	(0.366)	(0.741)	(NA)	(0.288)
	LL	-137.153	-144.433	-145.085	-145.085	NA	-137.153	-112.547	-144.433	-142.410	-109.800	-121.652	-101.078	NA	-114.253
60	ε or γ	1	0.941	1	0.000	0.011	1.000	0.000	0.941	0.051	1	0.012	0.011	0.017	NA
		(0.183)	(0.046)	(0.183)	(0.167)	(0.010)	(0.182)	(0.005)	(0.043)	(0.017)	(0.034)	(0.008)	(0.000)	(0.012)	(NA)
	τ						1.000	11.807	0.000	3.383	1.458	23.768	11.046	2.603	NA
							(NA)	(6.106)	(NA)	(1.241)	(NA)	(5.978)	(NA)	(0.738)	(NA)
	LL	-145.085	-142.510	-145.085	-145.085	-144.594	-145.085	-145.085	-142.510	-136.908	-145.085	-143.845	-144.595	-141.606	NA

Appendix C

Appendix for Chapter 3

C.1 Experimental Instructions

At the beginning of each session, paper instructions (Figure C.1) were distributed and read aloud; subjects followed along and could use a pen or pencil to write notes on the instructions, however, they were not allowed to use a pen or pencil once the experiment started. After the instructions, subjects completed a comprehension quiz (Figure C.2). The experimenter checked the answers privately, and when they encountered incorrect answers, the experimenter pointed the subject to the relevant part of the instructions and gave the subject the opportunity to revise their answers. After all subjects had answered all questions correctly, the experiment commenced. The experiment was programmed using the oTree web-based platform (Chen et al. 2016) and completed by subjects in their web browser.

Figure C.1: Instructions

Instructions

Overview

- You are going to take part in an experimental study of decision making.
- During the experiment, you are not allowed to talk or communicate with other participants.
- Also, please turn off your smart phones and put them away for the duration of the experiment.
- If at any time you have any questions, please raise your hand and the experimenter will come to your desk to answer it.
- Your earnings in the experiment will depend on your choices and an element of chance. By following the instructions and making decisions carefully, you may earn a considerable amount of money.
- The lottery you chose in one round of the experiment will be played out for real to determine your earnings from the experiment: thus you should make each choice as though it will be played out "for real" to determine your payment.
- Your earnings and a $7 participation payment will be paid to you in cash at the end of this experiment.

Choice task

- In each round you will choose one from among five lotteries.
- The decision screen (see the following page) will list the five lotteries, where each lottery consists of a set of possible payoffs and a probability of attaining each payoff.
- You should choose the lottery that you most prefer.
- There are no right or wrong answers and your responses may differ from other participants.
- You will have 60 seconds to make each decision.
- You finalize your choice by clicking the “Next” button or by allowing the time to run out.

Interface

The following screenshot shows how each choice task will be displayed on each decision screen.

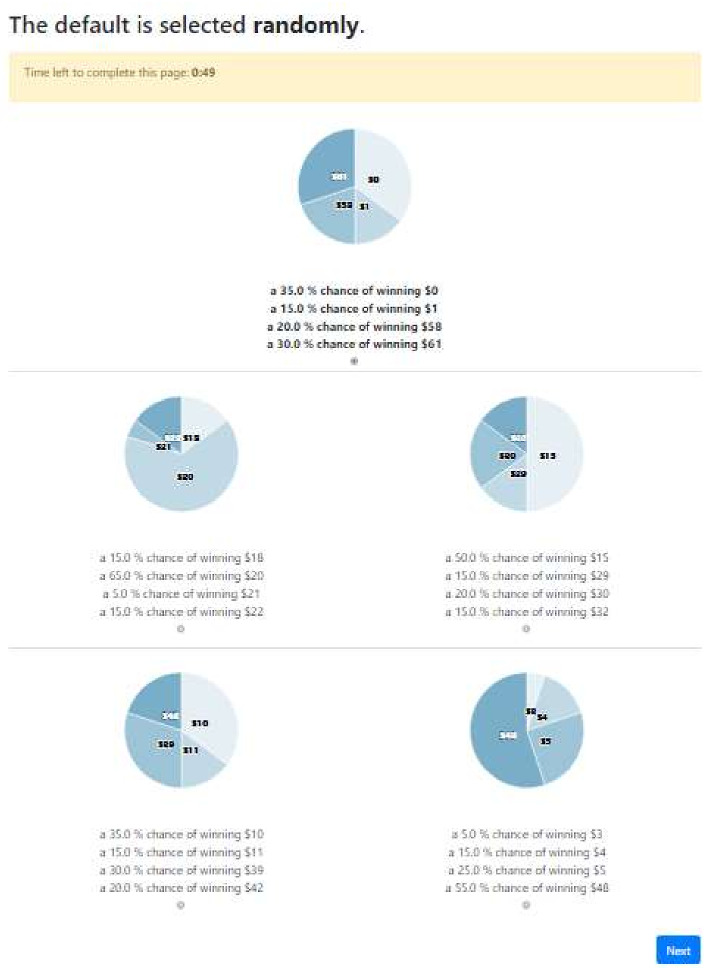
Below is an example of how the information about the set of possible payoffs and the corresponding probabilities are presented to you.

- The payoffs are ordered in increasing order.
- The size of a portion of the pie represents the probability of attaining the corresponding payoff in that portion.

Default options

- In some rounds, one lottery is selected as a “default” option.
- In such a round, the procedure used to select the default will be described at the top of the decision screen.
- When present, a default option is initially selected for you and will be presented at the top of the screen and in bold.
- For each task, you will have 60 seconds to make a decision.
- If you do not select another option before the time runs out, then the default option will automatically become your choice.
- You are always free to select the default or to choose another option: it’s your decision.

The following screenshot is an example of a round with default option.



Payment

- You will complete a total of 84 rounds in the experiment.
- One (and only one) round will be randomly selected to be the round that counts to determine your payment.
- Since any round could be the round that counts, you should treat your choice in each round of the experiment as though it will determine your earnings for the experiment.

Figure C.2: Quiz

Quiz

In each round, I can choose any lottery I wish.

True / False

In rounds with “default” option, if I don’t choose another option before the time runs out, the default option will be chosen for me.

True / False

C.2 Experimental Flow

Once the experiment started, each subject went through the experiment in the following order.

1. No Default treatment (24 rounds)
   (a) Instruction for the treatment (shown once at the beginning of the treatment) as in Figure C.3
   (b) Reminder/ waiting page between choices: Empty for 2 seconds for No Default treatment
   (c) Choice page as in Figure C.4
2. Random treatment (12 rounds)
   (a) Instruction for the treatment (shown once at the beginning of the treatment) as in Figure C.5
   (b) Reminder/ waiting page between choices for 2 seconds as in Figure C.6
   (c) Choice page as in Figure C.7
3. Depending on the session, subjects went through one of the six possible orders for Expert, Social, and Custom treatments – this was varied across subjects. Below is an example where the order is Social, Expert, and Custom.
   (a) Social treatment (12 rounds)
      i. Instruction for the treatment (shown once at the beginning of the treatment) as in Figure C.8
      ii. Reminder/ waiting page between choices for 2 seconds as in Figure C.9
      iii. Choice page as in Figure C.10
   (b) Expert treatment (12 rounds)
      i. Instruction for the treatment (shown once at the beginning of the treatment) as in Figure C.11
      ii. Reminder/ waiting page between choices for 2 seconds as in Figure C.12
      iii. Choice page as in Figure C.13
   (c) Custom treatment (12 rounds)
      i. Instruction for the treatment (shown once at the beginning of the treatment) as in Figure C.14
      ii. Reminder/ waiting page between choices for 2 seconds as in Figure C.15
      iii. Choice page as in Figure C.16
4. Ranking of default regimes at the beginning of round 73 as in Figure C.17
5. One treatment is implemented for the last 12 rounds.

Figure C.3: Instruction for No Default treatment

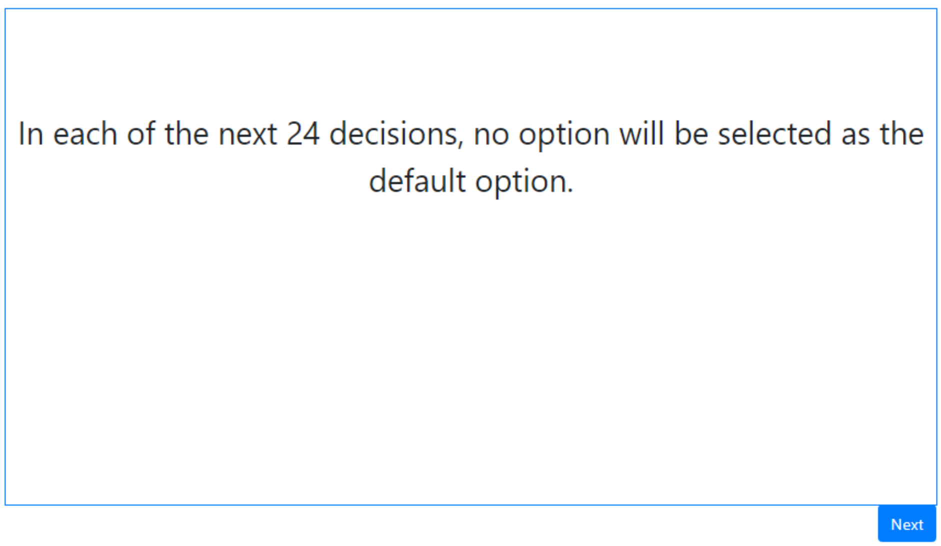


Figure C.4: Choice screen for No Default treatment

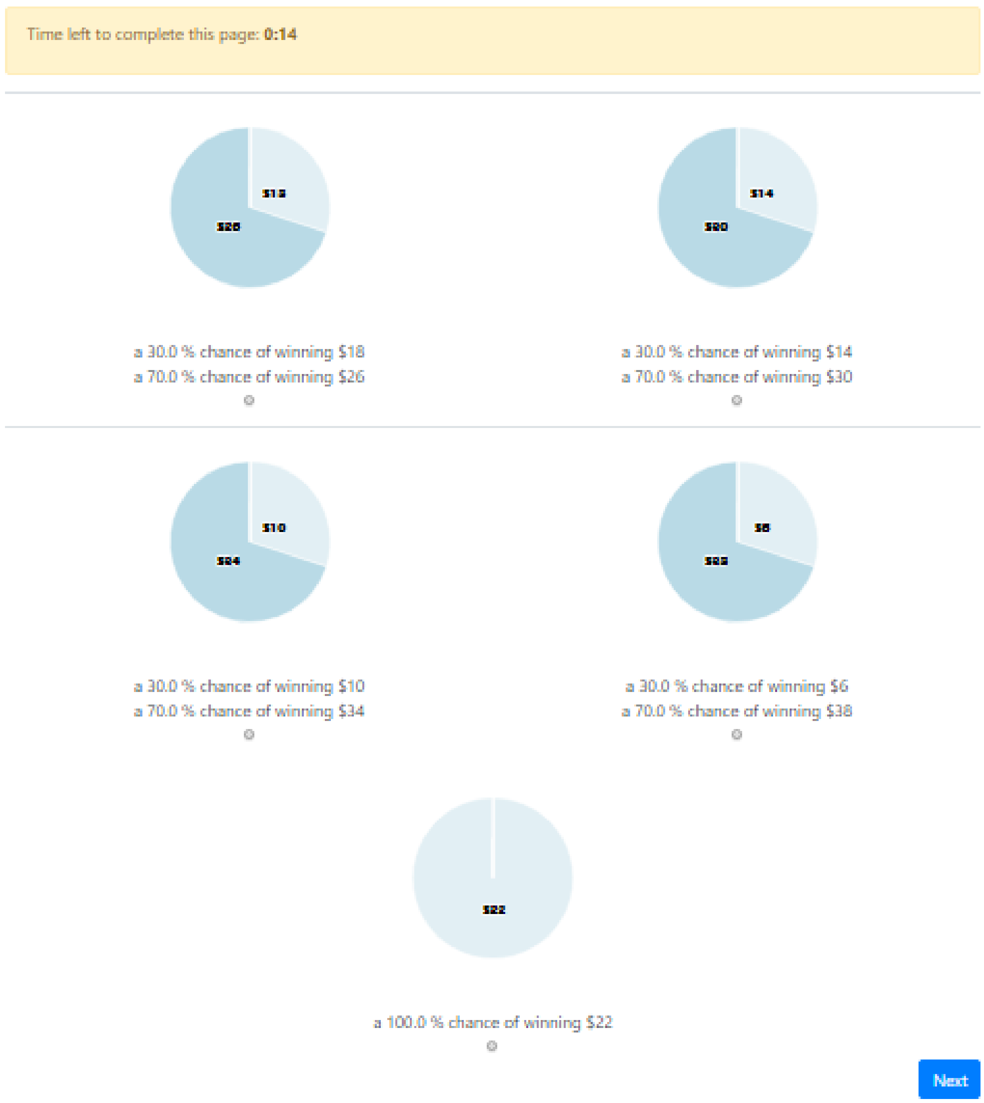


Figure C.5: Instruction for Random treatment

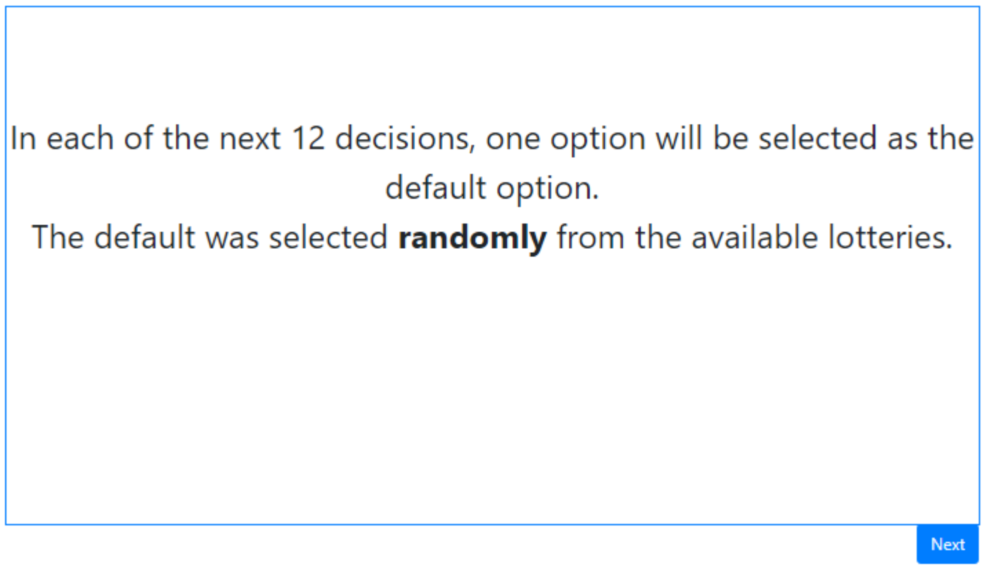


Figure C.6: Waiting page for Random treatment

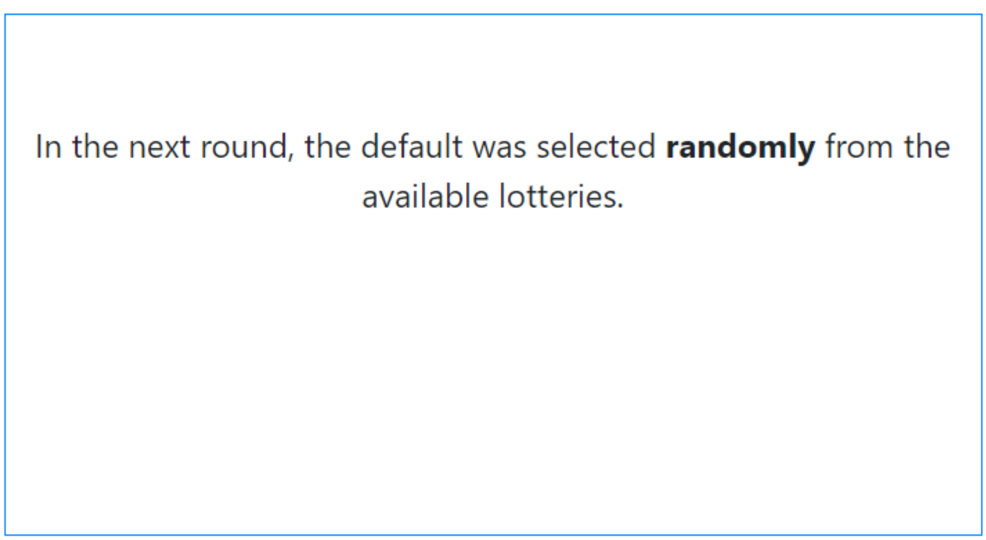


Figure C.7: Choice screen for Random treatment

Figure C.8: Instruction for Social treatment

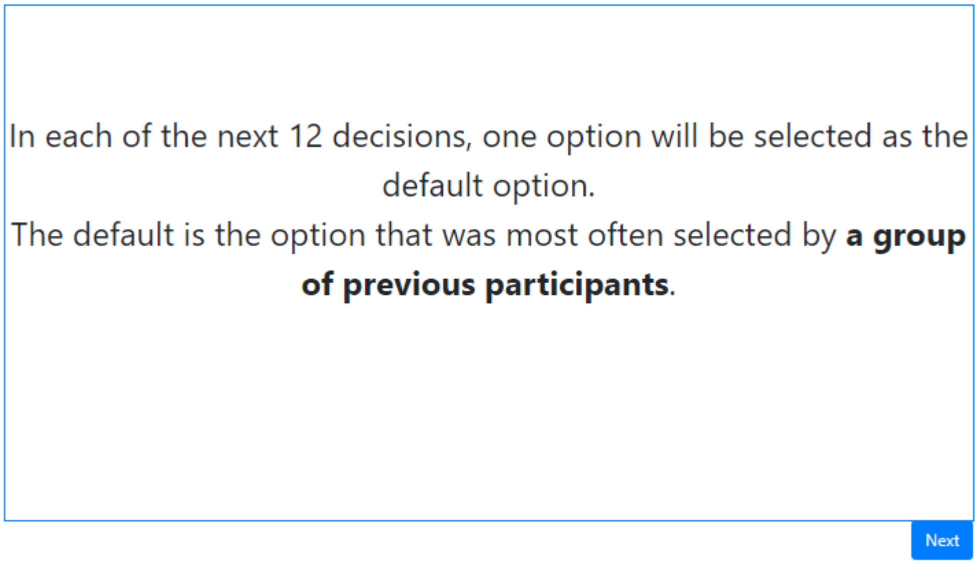


Figure C.9: Waiting page for Social treatment

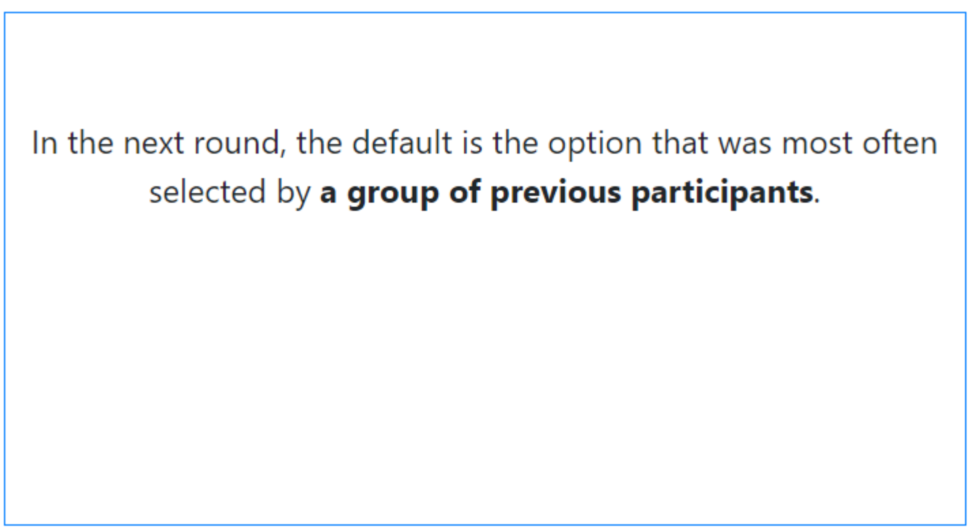


Figure C.10: Choice screen for Social treatment

Figure C.11: Instruction for Expert treatment

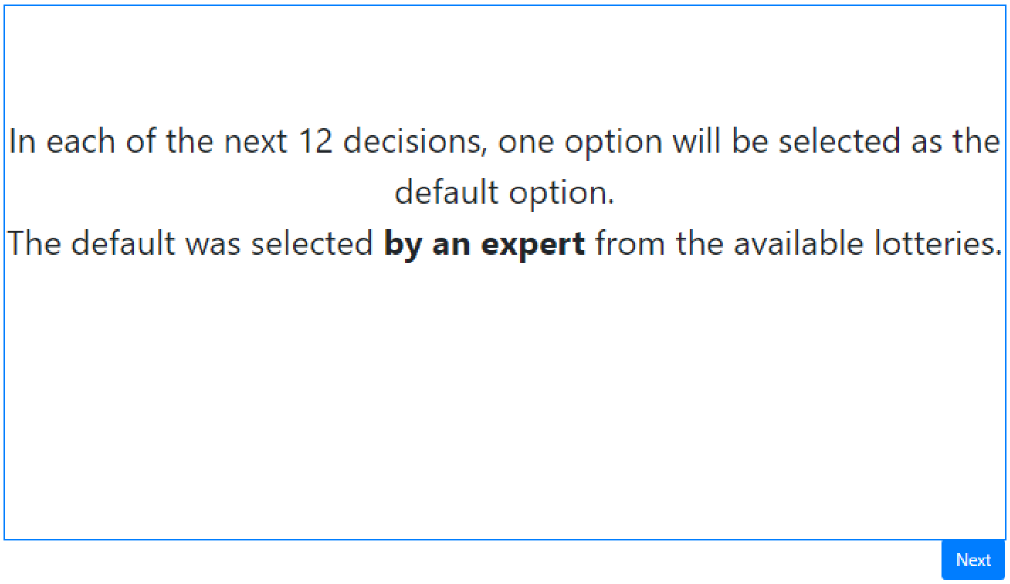


Figure C.12: Waiting page for Expert treatment

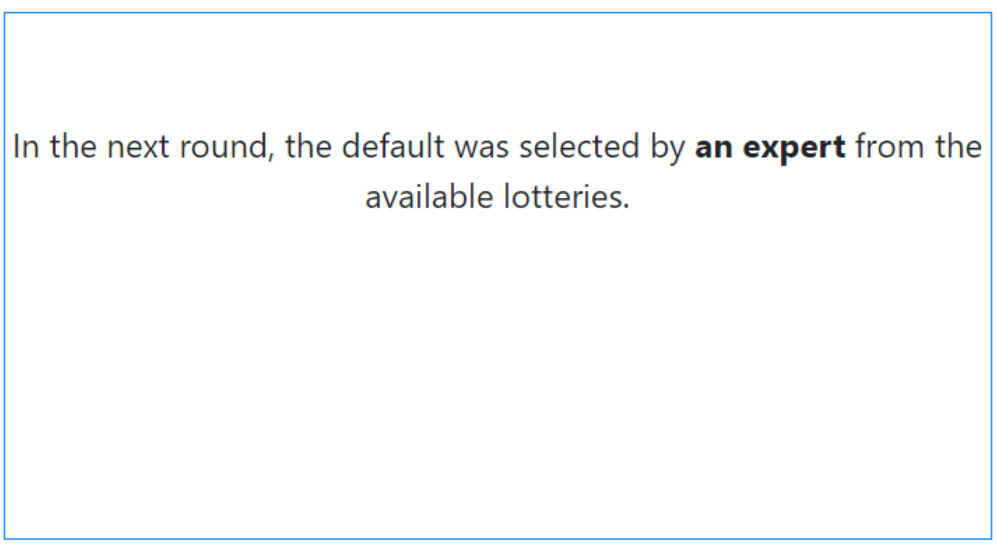


Figure C.13: Choice screen for Expert treatment

Figure C.14: Instruction for Custom treatment

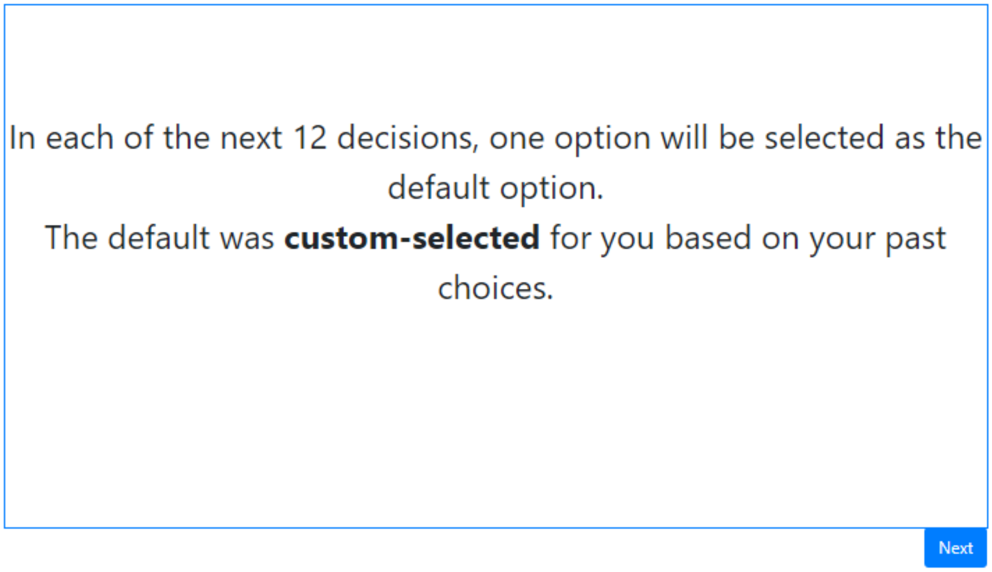


Figure C.15: Waiting page for Custom treatment

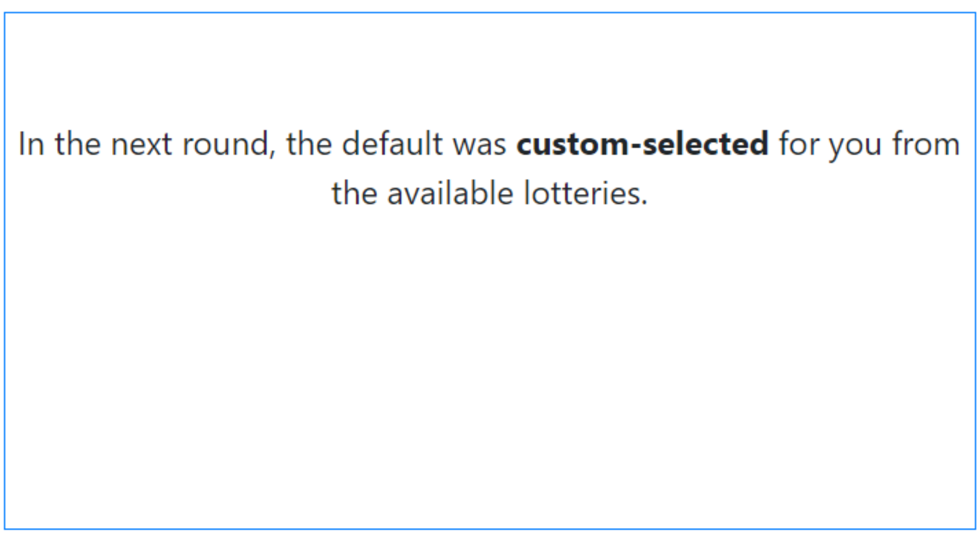


Figure C.16: Choice screen for Custom treatment

Figure C.17: Ranking of default regimes

Ranking defaults

Please rank the default-setting rules used thus far from #1 (most preferred) to #5 (least preferred).
The default-setting rule used in the next 12 rounds of the experiment will be determined from your ranking.
The default-setting rules you indicate that you prefer more are more likely to be implemented, as follows:

Rank		Implemented with probability
Most preferred	#1	90%
	#2	7%
	#3	2%
	#4	1%
Least Preferred	#5	0%

Number 1 (most preferred):

Number 2:

Number 3:

Number 4:

Number 5 (least preferred):

Next

C.3 Choice Sets

The following table lists 24 unique choice sets[1] which we organized into six blocks of four lotteries (A to F). Each choice set comprises of five lotteries (1 to 5). Each lottery has at most four outcomes. We use pa, pb, pc, pd and xa, xb, xc, xd to denote probabilities and payoffs of outcomes a, b, c, and d, respectively.

We use * to mark the default option for a choice set, under a default-setting regime. Default-setting rules are explained in the Experimental Design section. For the Random treatment, a lottery was randomly selected from each choice set, for each subject, so we drop a column for this treatment. For Expert and Custom treatment, we used an expected utility model with constant relative risk averse utility-for-income function $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$. For the Expert treatment, we used $\gamma = \frac{3}{4}$ to select a default lottery from each choice set. For the Custom treatment, we assigned each participant into one of three groups based on their choices in three Eckel-Grossman style choice sets in No Default treatment. We then used the parameters $\gamma = 2$,1.25, 0.5 to select a default lottery for groups C_H, C_M, and C_L, respectively.

[1] We constructed 84 rounds from 24 choice sets using an overlapping structure outlined in the next section.

Set	Name	Lottery	pa	pb	pc	pd	xa	xb	xc	xd	Social	Expert	C_H	C_M	C_L
1	A1	1	1	0	0	0	18	0	0	0	*				
1	A1	2	0.5	0.5	0	0	16	24	0	0			*	*	
1	A1	3	0.5	0.5	0	0	13	30	0	0		*			*
1	A1	4	0.5	0.5	0	0	8	36	0	0					
1	A1	5	0.5	0.5	0	0	0	42	0	0					
2	A2	1	1	0	0	0	12	0	0	0	*	*	*	*	*
2	A2	2	0.25	0.375	0.375	0	0	12	24	0					
2	A2	3	0.375	0.625	0	0	0	24	0	0					
2	A2	4	0.375	0.25	0.375	0	0	12	36	0					
2	A2	5	0.5	0.5	0	0	0	36	0	0					
3	A3	1	0.5	0.5	0	0	0	60	0	0					
3	A3	2	0.2	0.8	0	0	0	40	0	0					
3	A3	3	0.3	0.6	0.1	0	0	40	60	0					
3	A3	4	0.2	0.2	0.6	0	0	20	40	0					
3	A3	5	0.6	0.4	0	0	20	40	0	0	*	*	*	*	*
4	A4	1	1	0	0	0	20	0	0	0					
4	A4	2	0.25	0.25	0.25	0.25	11	13	31	33					
4	A4	3	0.5	0.25	0.25	0	12	24	42	0					
4	A4	4	0.25	0.25	0.5	0	4	22	32	0					
4	A4	5	0.25	0.5	0.25	0	12	23	42	0	*	*	*	*	*
5	B1	1	1	0	0	0	18	0	0	0	*				
5	B1	2	0.1	0.9	0	0	15	19	0	0					
5	B1	3	0.1	0.9	0	0	12	20	0	0			*		
5	B1	4	0.1	0.9	0	0	9	21	0	0				*	
5	B1	5	0.1	0.9	0	0	6	22	0	0		*			*
6	B2	1	1	0	0	0	20	0	0	0					
6	B2	2	0.25	0.5	0.25	0	10	20	30	0					
6	B2	3	0.5	0.5	0	0	10	30	0	0					
6	B2	4	0.25	0.375	0.375	0	0	20	30	0					
6	B2	5	0.875	0.125	0	0	20	30	0	0	*	*	*	*	*

Set	Name	Lottery	pa	pb	pc	pd	xa	xb	xc	xd	Social	Expert	C_H	C_M	C_L
7	B3	1	0.6	0.4	0	0	0	60	0	0					
7	B3	2	0.5	0.5	0	0	0	40	0	0					
7	B3	3	0.2	0.6	0.2	0	0	20	40	0	*				
7	B3	4	0.3	0.5	0.2	0	0	20	60	0					
7	B3	5	0.1	0.9	0	0	0	20	0	0		*	*	*	*
8	B4	1	1	0	0	0	18	0	0	0			*		
8	B4	2	0.2	0.3	0.05	0.45	2	20	34	36					
8	B4	3	0.25	0.35	0.15	0.25	2	21	35	37					
8	B4	4	0.05	0.45	0.2	0.3	9	11	25	45		*		*	*
8	B4	5	0.35	0.15	0.3	0.2	9	12	25	46	*				
9	C1	1	1	0	0	0	24	0	0	0					
9	C1	2	0.5	0.5	0	0	23	30	0	0					
9	C1	3	0.5	0.5	0	0	21	35	0	0	*	*	*	*	*
9	C1	4	0.5	0.5	0	0	18	39	0	0					
9	C1	5	0.5	0.5	0	0	14	43	0	0					
10	C2	1	0.5	0.5	0	0	0	60	0	0					
10	C2	2	0.3	0.7	0	0	0	40	0	0					
10	C2	3	0.1	0.4	0.5	0	0	20	40	0	*				*
10	C2	4	0.2	0.6	0.2	0	0	20	60	0					
10	C2	5	1	0	0	0	20	0	0	0		*	*	*	
11	C3	1	0.5	0.125	0.375	0	0	20	30	0					
11	C3	2	0.125	0.5	0.375	0	0	10	30	0					*
11	C3	3	0.375	0.125	0.5	0	0	10	20	0					
11	C3	4	0.125	0.75	0.125	0	0	10	20	0					
11	C3	5	0.75	0.125	0.125	0	10	20	30	0	*	*	*	*	
12	C4	1	0.15	0.45	0.15	0.25	22	24	25	26			*	*	
12	C4	2	0.05	0.45	0.4	0.1	17	18	33	34		*			*
12	C4	3	0.15	0.35	0.05	0.45	10	13	41	42	*				
12	C4	4	0.5	0.05	0.1	0.35	6	50	51	52					
12	C4	5	0.45	0.05	0.1	0.4	0	1	51	62					

Set	Name	Lottery	pa	pb	pc	pd	xa	xb	xc	xd	Social	Expert	C_H	C_M	C_L
13	D1	1	1	0	0	0	20	0	0	0	*				
13	D1	2	0.2	0.8	0	0	16	22	0	0			*		
13	D1	3	0.2	0.8	0	0	12	24	0	0		*		*	
13	D1	4	0.2	0.8	0	0	8	26	0	0					*
13	D1	5	0.2	0.8	0	0	4	28	0	0					
14	D2	1	0.5	0.5	0	0	0	60	0	0					
14	D2	2	0.3	0.7	0	0	0	40	0	0					
14	D2	3	0.1	0.4	0.5	0	0	20	40	0					*
14	D2	4	0.2	0.6	0.2	0	0	20	60	0					
14	D2	5	1	0	0	0	20	0	0	0	*	*	*	*	
15	D3	1	0.25	0.375	0.375	0	0	10	30	0					
15	D3	2	0.125	0.625	0.25	0	0	10	30	0					
15	D3	3	0.125	0.25	0.625	0	0	10	20	0					
15	D3	4	0.5	0.5	0	0	10	20	0	0	*	*	*	*	*
15	D3	5	0.375	0.5	0.125	0	0	20	30	0					
16	D4	1	0.35	0.25	0.15	0.25	17	18	19	20	*				
16	D4	2	0.3	0.2	0.35	0.15	15	17	25	26		*	*	*	*
16	D4	3	0.4	0.2	0.15	0.25	13	14	29	31					
16	D4	4	0.5	0.15	0.25	0.1	8	35	36	37					
16	D4	5	0.35	0.05	0.1	0.5	0	1	2	42					
17	E1	1	1	0	0	0	15	0	0	0					
17	E1	2	0.5	0.5	0	0	12	21	0	0	*	*	*	*	
17	E1	3	0.5	0.5	0	0	9	26	0	0					*
17	E1	4	0.5	0.5	0	0	6	30	0	0					
17	E1	5	0.5	0.5	0	0	3	33	0	0					
18	E2	1	1	0	0	0	20	0	0	0	*		*	*	
18	E2	2	0.125	0.5	0.375	0	0	20	30	0					
18	E2	3	0.375	0.625	0	0	10	30	0	0		*			*
18	E2	4	0.25	0.375	0.375	0	10	20	30	0					
18	E2	5	0.25	0.75	0	0	0	30	0	0					

Set	Name	Lottery	pa	pb	pc	pd	xa	xb	xc	xd	Social	Expert	C_H	C_M	C_L
19	E3	1	0.5	0.2	0.3	0	0	20	40	0	*				*
19	E3	2	0.6	0.1	0.3	0	0	20	60	0					
19	E3	3	0.6	0.4	0	0	0	40	0	0					
19	E3	4	0.7	0.1	0.2	0	0	40	60	0					
19	E3	5	0.4	0.6	0	0	0	20	0	0		*	*	*	
20	E4	1	0.15	0.45	0.25	0.15	15	16	17	18					
20	E4	2	0.5	0.15	0.1	0.25	12	20	21	24					
20	E4	3	0.2	0.25	0.1	0.45	12	13	15	25	*	*	*	*	*
20	E4	4	0.35	0.1	0.25	0.3	2	10	33	35					
20	E4	5	0.45	0.1	0.2	0.25	2	32	34	35					
21	F1	1	1	0	0	0	22	0	0	0	*				
21	F1	2	0.3	0.7	0	0	18	26	0	0			*		
21	F1	3	0.3	0.7	0	0	14	30	0	0				*	
21	F1	4	0.3	0.7	0	0	10	34	0	0		*			
21	F1	5	0.3	0.7	0	0	6	38	0	0					*
22	F2	1	0.375	0.375	0.25	0	0	10	20	0					
22	F2	2	0.5	0.125	0.375	0	0	10	20	0					
22	F2	3	0.5	0.375	0.125	0	0	10	30	0					
22	F2	4	0.75	0.25	0	0	0	30	0	0					
22	F2	5	0.25	0.75	0	0	0	10	0	0	*	*	*	*	*
23	F3	1	0.2	0.8	0	0	0	40	0	0					
23	F3	2	0.3	0.6	0.1	0	0	40	60	0					
23	F3	3	0.2	0.2	0.6	0	0	20	40	0					
23	F3	4	0.6	0.4	0	0	20	40	0	0	*	*	*	*	*
23	F3	5	1	0	0	0	20	0	0	0					
24	F4	1	0.35	0.45	0.15	0.05	22	24	25	27					
24	F4	2	0.25	0.2	0.35	0.2	22	23	28	31					
24	F4	3	0.5	0.15	0.05	0.3	21	34	35	36		*	*	*	*
24	F4	4	0.3	0.2	0.15	0.35	17	19	38	40	*				
24	F4	5	0.1	0.2	0.25	0.45	14	15	16	46					

C.4 Overlapping Structure and Order of Choice Tasks

In our experiment, each subject first faced all 24 unique choice sets in the No Default treatment, and then faced each choice set twice more under the four default-setting rules, three blocks for each rule. These were arranged so that there was exactly one block of overlap between any two default-setting rules. After default-setting preference elicitation, three blocks were repeated again in the last 12 rounds where the default-setting rule was selected based on the subject's ranking. In total, we have 84 rounds.

We have six treatment orders based on six possible orders of Expert, Social, and Custom treatments. In addition, we mixed starting blocks and interweaved blocks so that subjects would rarely see two choice sets in the same order, to avoid subjects recognizing the block order. Treatment variations were assigned at the session level. Table C.1 lists the choice sequence that we used in our experiment.

Table C.1: Choice sequence

Session	No. of subjects	Treatment order	Choice set order
S1	9	No Default	A1, B2, C3, A4, B1, C2, A3, B4, C1, A2, B3, C4, D1, E2, F3, D4, E1, F2, D3, E4, F1, D2, E3, F4
		Random	A1, A2, B4, A3, B2, A4, C1, B3, C2, B1, C3, C4
		Expert	D1, D2, E4, D3, E2, D4, A1, E3, A2, E1, A3, A4
		Social	F1, F2, B4, F3, B2, F4, D1, B3, D2, B1, D3, D4
		Custom	E1, E2, C4, E3, C2, E4, F1, C3, F2, C1, F3, F4
		Choice of Default	B1, B2, D4, B3, D2, B4, E1, D3, E2, D1, E3, E4
S2	8	No Default	A4, A3, A2, A1, B4, B3, B2, B1, C4, C3, C2, C1, D4, D3, D2, D1, E4, E3, E2, E1, F4, F3, F2, F1
		Random	A4, A3, B1, A2, B3, A1, C4, B2, C3, B4, C2, C1
		Custom	D4, D3, E1, D2, E3, D1, A4, E2, A3, E4, A2, A1
		Expert	F4, F3, B1, F2, B3, F1, D4, B2, D3, B4, D2, D1
		Social	E4, E3, C1, E2, C3, E1, F4, C2, F3, C4, F2, F1
		Choice of Default	B4, B3, D1, B2, D3, B1, E4, D2, E3, D4, E2, E1
S5	12	No Default	B4, B3, B2, B1, C4, C3, C2, C1, D4, D3, D2, D1, E4, E3, E2, E1, F4, F3, F2, F1, A4, A3, A2, A1
		Random	B1, B2, C4, B3, C2, B4, D1, C3, D2, C1, D3, D4
		Social	E1, E2, F4, E3, F2, E4, B1, F3, B2, F1, B3, B4
		Expert	A1, A2, C4, A3, C2, A4, E1, C3, E2, C1, E3, E4
		Custom	F1, F2, D4, F3, D2, F4, A1, D3, A2, D1, A3, A4
		Choice of Default	C1, C2, E4, C3, E2, C4, F1, E3, F2, E1, F3, F4
S6	11	No Default	B1, C2, B3, C4, C1, B2, C3, B4, D1, E2, D3, E4, E1, D2, E3, D4, F1, A2, F3, A4, A1, F2, A3, F4
		Random	B4, B3, C1, B2, C3, B1, D4, C2, D3, C4, D2, D1
		Custom	E4, E3, F1, E2, F3, E1, B4, F2, B3, F4, B2, B1
		Social	A4, A3, C1, A2, C3, A1, E4, C2, E3, C4, E2, E1
		Expert	F4, F3, D1, F2, D3, F1, A4, D2, A3, D4, A2, A1
		Choice of Default	C4, C3, E1, C2, E3, C1, F4, E2, F3, E4, F2, F1
S8	8	No Default	C1, D2, E3, C4, D1, E2, C3, D4, E1, C2, D3, E4, F1, A2, B3, F4, A1, B2, F3, A4, B1, F2, A3, B4
		Random	C4, C2, D1, C3, D2, C1, E4, D3, E2, D4, E3, E1
		Custom	F4, F2, A1, F3, A2, F1, C4, A3, C2, A4, C3, C1
		Expert	B4, B2, D1, B3, D2, B1, F4, D3, F2, D4, F3, F1
		Social	A4, A2, E1, A3, E2, A1, B4, E3, B2, E4, B3, B1
		Choice of Default	D4, D2, F1, D3, F2, D1, A4, F3, A2, F4, A3, A1
S9	8	No Default	C1, D2, C3, D4, D1, C2, D3, C4, E1, F2, E3, F4, F1, E2, F3, E4, A1, B2, A3, B4, B1, A2, B3, A4
		Random	C1, C2, D4, C3, D2, C4, E1, D3, E2, D1, E3, E4
		Social	F1, F2, A4, F3, A2, F4, C1, A3, C2, A1, C3, C4
		Custom	B1, B2, D4, B3, D2, B4, F1, D3, F2, D1, F3, F4
		Expert	A1, A2, E4, A3, E2, A4, B1, E3, B2, E1, B3, B4
		Choice of Default	D1, D2, F4, D3, F2, D4, A1, F3, A2, F1, A3, A4

Session	No. of subjects	Treatment order	Choice set order
S10	11	No Default	D1, E2, D3, E4, E1, D2, E3, D4, F1, A2, F3, A4, A1, F2, A3, F4, B1, C2, B3, C4, C1, B2, C3, B4
		Random	D4, D3, E1, D2, E3, D1, F4, E2, F3, E4, F2, F1
		Expert	A4, A3, B1, A2, B3, A1, D4, B2, D3, B4, D2, D1
		Custom	B4, B3, F1, B2, F3, B1, C4, F2, C3, F4, C2, C1
		Social	C4, C3, E1, C2, E3, C1, A4, E2, A3, E4, A2, A1
		Choice of Default	E4, E3, A1, E2, A3, E1, B4, A2, B3, A4, B2, B1
S11	11	No Default	D1, E2, F3, D4, E1, F2, D3, E4, F1, D2, E3, F4, A1, B2, C3, A4, B1, C2, A3, B4, C1, A2, B3, C4
		Random	D1, D3, E4, D2, E3, D4, F1, E2, F3, E1, F2, F4
		Social	A1, A3, B4, A2, B3, A4, D1, B2, D3, B1, D2, D4
		Expert	C1, C3, E4, C2, E3, C4, A1, E2, A3, E1, A2, A4
		Custom	B1, B3, F4, B2, F3, B4, C1, F2, C3, F1, C2, C4
		Choice of Default	E1, E3, A4, E2, A3, E4, B1, A2, B3, A1, B2, B4
S13	6	No Default	E1, F2, E3, F4, F1, E2, F3, E4, A1, B2, A3, B4, B1, A2, B3, A4, C1, D2, C3, D4, D1, C2, D3, C4
		Random	E1, E2, F4, E3, F2, E4, A1, F3, A2, F1, A3, A4
		Expert	B1, B2, C4, B3, C2, B4, E1, C3, E2, C1, E3, E4
		Social	D1, D2, F4, D3, F2, D4, B1, F3, B2, F1, B3, B4
		Custom	C1, C2, A4, C3, A2, C4, D1, A3, D2, A1, D3, D4
		Choice of Default	F1, F2, B4, F3, B2, F4, C1, B3, C2, B1, C3, C4
S15	10	No Default	E4, E3, E2, E1, F4, F3, F2, F1, A4, A3, A2, A1, B4, B3, B2, B1, C4, C3, C2, C1, D4, D3, D2, D1
		Random	E1, E3, F4, E2, F3, E4, A1, F2, A3, F1, A2, A4
		Social	B1, B3, C4, B2, C3, B4, E1, C2, E3, C1, E2, E4
		Custom	D1, D3, F4, D2, F3, D4, B1, F2, B3, F1, B2, B4
		Expert	C1, C3, A4, C2, A3, C4, D1, A2, D3, A1, D2, D4
		Choice of Default	F1, F3, B4, F2, B3, F4, C1, B2, C3, B1, C2, C4
S16	11	No Default	F1, A2, F3, A4, A1, F2, A3, F4, B1, C2, B3, C4, C1, B2, C3, B4, D1, E2, D3, E4, E1, D2, E3, D4
		Random	F4, F2, A1, F3, A2, F1, B4, A3, B2, A4, B3, B1
		Expert	C4, C2, D1, C3, D2, C1, F4, D3, F2, D4, F3, F1
		Custom	E4, E2, A1, E3, A2, E1, C4, A3, C2, A4, C3, C1
		Social	D4, D2, B1, D3, B2, D1, E4, B3, E2, B4, E3, E1
		Choice of Default	A4, A2, C1, A3, C2, A1, D4, C3, D2, C4, D3, D1
S18	8	No Default	F4, F3, F2, F1, A4, A3, A2, A1, B4, B3, B2, B1, C4, C3, C2, C1, D4, D3, D2, D1, E4, E3, E2, E1
		Random	F4, F3, A1, F2, A3, F1, B4, A2, B3, A4, B2, B1
		Custom	C4, C3, D1, C2, D3, C1, F4, D2, F3, D4, F2, F1
		Social	E4, E3, A1, E2, A3, E1, C4, A2, C3, A4, C2, C1
		Expert	D4, D3, B1, D2, B3, D1, E4, B2, E3, B4, E2, E1
		Choice of Default	A4, A3, C1, A2, C3, A1, D4, C2, D3, C4, D2, D1

Statistical Packages

Our statistical analysis was conducted in R (R Core Team 2017). Obuchowski tests used the 'clust.bin.pair' command in the 'clust.bin.pair' package (Gopstein 2018). Rank-ordered logit regression, with robust standard errors, used the 'rologit' command in STATA (StataCorp. 2013). Figures made in 'ggplot2' (Wickham 2009).

C.5 Ranking of Default-Setting Rules and Risk Attitudes

In this section, we explore the connection between heterogeneity in risk attitudes and the ranking of default-setting rules. As described in Section 3.3, we coarsely scored each subject's risk aversion based on their No Default choices in the three Eckel and Grossman Eckel and Grossman (2002) style choice problems with equal likelihood of both outcomes for each lottery. In each of these choice sets, lotteries were scored from 1 (safest) to 5 (riskiest) and we added these scores to obtain a final score S between 3 (i.e. always choosing the safest option) and 15 (i.e. always choosing the riskiest option). Based on the score S, we assigned each subject to one of three groups: $S \leq 6$ (the most risk-averse group), $7 \leq S \leq 9$ (the moderately risk-averse group), and $S \geq 10$ (the least risk-averse group). The tables below count the fraction of subjects who ranked each rule in each position, for the three groups.

We construct a Borda count for each default-setting rule for each group. Overall, the aggregate ranking is Expert $\succ$ Custom $\succ$ Social $\succ$ No Default $\succ$ Random[2]. We do not find any apparent differences in the ranking of default-setting rules across groups with differrent risk profiles.

Table C.2: Ranking of default-setting rules by the most risk-averse group of subjects

Default type	#1	#2	#3	#4	#5
No Default	0.146	0.024	0.122	0.341	0.366
Random	0.049	0.122	0.146	0.268	0.415
Expert	0.488	0.293	0.122	0.024	0.073
Social	0.098	0.244	0.366	0.220	0.073
Custom	0.220	0.317	0.244	0.146	0.073

$n = 41$ subjects

[2]For the least risk-averse group, there is a switch of ranking between Social and No Default, thus the ranking is Expert $\succ$ Custom $\succ$ No Default $\succ$ Social $\succ$ Random.

Table C.3: Ranking of default-setting rules by the moderately risk-averse group of subjects

Default type	#1	#2	#3	#4	#5
No Default	0.171	0.195	0.171	0.195	0.268
Random	0.024	0.049	0.073	0.341	0.512
Expert	0.463	0.317	0.171	0.049	0.000
Social	0.073	0.220	0.341	0.293	0.073
Custom	0.268	0.220	0.244	0.122	0.146

$n = 41$ subjects

Table C.4: Ranking of default-setting rules by the least risk-averse group of subjects

Default type	#1	#2	#3	#4	#5
No Default	0.355	0.129	0.032	0.097	0.387
Random	0.065	0.065	0.194	0.452	0.226
Expert	0.290	0.290	0.226	0.065	0.129
Social	0.097	0.226	0.258	0.226	0.194
Custom	0.194	0.290	0.290	0.161	0.065

$n = 31$ subjects