Essays on Labor and Behavioral Economics

by

Felipe Augusto de Araujo

Bachelor of Arts, Federal University of Juiz de Fora, 2008

Master of Arts, Federal University of Minas Gerais, 2012

Master of Arts, University of Pittsburgh, 2016

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Felipe Augusto de Araujo

It was defended on

April 22nd 2020

and approved by

David B. Huffman, Department of Economics, University of Pittsburgh

Lise Vesterlund, Department of Economics, University of Pittsburgh

Alistair J. Wilson, Department of Economics, University of Pittsburgh

Rania Gihleb, Department of Economics, University of Pittsburgh

Alex Imas, Department of Social and Decision Sciences, Carnegie Mellon University

Essays on Labor and Behavioral Economics

Felipe Augusto de Araujo, PhD

University of Pittsburgh, 2020

This dissertation includes three essays on topics in labor and behavioral economics. The first chapter studies a long-lasting question in labor, namely if public-sector workers are paid more than their counterfactual wages in the private sector. The second chapter experimentally studies the problem of adverse selection occurring over time. Finally, the third chapter examines decision-making when agents become aware of hitherto unknown contingencies, also using experimental methods.

Table of Contents

List of Tables

List of Figures

Preface

To my parents, who have always supported my decisions;

To my wife, for loving me and putting up with my sometimes insanely long hours;

To my sons, for not being helpful at all;

To my friends, in Brazil and the US, for always being there for me;

To my big sister, because I once forgot to mention her in another graduation;

To my advisors, from whom I learned much, much more than from books and papers;

To the Labor Ministry of Brazil and the two public-sector exam organizers that shared essential data for this work;

This work could never have been completed without the support of my family and friends. I would also like to thank Isabela Tazima, Anna Clara Moreira and Ana Borges for excellent research assistance with the first chapter. And a special thank you to my committee chair, David B. Huffman, who was always available to gently steer me in the right direction and provide invaluable advice.

If I hadn't spent an estimated 400 hours during the PhD getting coffee with Mark Azic and others, this dissertation would have been finished earlier and would have been better. Or not.

1.0 Introduction

This dissertation is composed of three chapters: an empirical work on labor economics and two experimental papers on dynamic adverse selection and decisions under unawareness. Even though the chapters are not strictly connected, they all share the same motivation of understanding individual behavior in economic situations with complex information structures and challenging information processing and learning.

The first chapter—and my job market paper—studies a question with a long history in labor economics: are government employees paid more than their counterfactual private-sector wages? The main empirical challenge in this literature is the issue of selection bias: workers that sort into the public sector may be different in terms of unobservable characteristics than workers that sort into the private sector. To quote an early study by [Van Ophem, 1993], "the problem is not that the researcher does not want to take such factors [self selection] into account, but the problem is that the data do not contain such information". In this chapter, I create a unique dataset that does contain the necessary information. The data covers the Brazilian labor market, where public-sector jobs are allocated via exams. I construct a dataset with over 2 million individuals linking public-sector exam results to labor market outcomes. The data allow me to address two major sources of selection bias. Specifically, I compare outcomes for workers that have sought a public-sector job and had a very similar performances in the exams. Using a regression discontinuity design, I find a positive and large public-sector wage premium. The estimated coefficient is almost twice as large as simple OLS estimates, suggesting there is negative selection into the public sector. Finally, in order to better understand the selection mechanism, I conducted a survey with undergraduates at a Brazilian university. I collected information on career preferences and elicited full wage expectations for both sectors. I find that subjects who prefer to work in the public sector expect a positive and large wage premium, while those that prefer not to work for the government expect a negative wage gap. The survey results shed light on the nature of the negative selection.

The second chapter of this dissertation (joint with Alistair Wilson and Stephanie Wang), uses experimental methods to study an adverse selection environment that is common in many economically relevant situations. The chapter uses a new experimental design in which subjects make decisions on a series of common-value matching games where the value of an outside option deteriorates as time passes. A natural counterpart, for example, are situations where a potential employer looks at job applicants that have had different numbers of previous jobs and different spells of unemployment. All else equal, candidates with more job separations and longer spells are likely to be less productive. The experimental results indicate that while a minority understand the selection mechanism and act accordingly, most subjects fail to condition the quality of the outside option on the passage of time. For the latter types, choices are better described by a behavioral model in which agents notice the overall selection problem, but fail to comprehend the dynamic aspect of it. We test this model in an out-of-sample and out-of-context experiment and find support for the main predictions.

Finally, on the third chapter (joint with Evan Piermont), we design a web-based experiment to study the interaction between unawareness and risk-taking. In the experiment, subjects complete a pattern-finding task for which the space of valid patterns is easy to understand, but hard to enumerate. In other words, finding a new pattern is hard, but verifying its validity is easy. We use this new task to exogenously vary subjects' awareness by manipulating the type and extent of feedback. We find that subjects unawareness per se does not affect risk-taking. We find, however, that the context in which the risky decision is made is crucial: subjects exposed to unawareness are shown to be more risk averse when choices are made within the same context as the pattern-finding task, as opposed to risky choices in a neutral setting.

2.0 Selection Bias and the Returns to Public-Sector Employment

2.1 Introduction

Personnel policies in the public sector may be set through a political economy process, rather than by market forces Barro [1973], Alesina et al. [2000]; hence it is not obvious that wages in the public and private sectors should be the same for similar individuals. A potential gap between public and private sector wages could in turn have major implications. First, it may affect the allocation of talent, impacting the quality of public-sector workers and by extension the quality of public goods and services. Second, it could result in an inefficient use of resources if, for example, the public sector pays more than is necessary to match workers' outside options. Since the wage bill is usually the government's largest expense, this difference could impact its fiscal sustainability, especially in developing countries Gindling et al. [2019]. Third, the wage and hiring policies adopted by the public sector can affect equilibrium outcomes of the entire labor market Bradley et al. [2017], Albrecht et al. [2018]. Lastly, since the public sector employs, on average, 14 percent of the labor force[1], the impact of its wage policies on resource allocation, and public expenditures, is likely very large.

Because of these potentially profound implications for the functioning of economies, a large literature has sought to establish whether public and private sector wages differ.[2] A key challenge in testing for and measuring the size of the public-private wage gap, however, is the issue of selection bias. That is, workers that sort into the public sector might be different with respect to unobservable characteristics than workers that sort into the private sector. As [Van Ophem, 1993, p.206] stated, "the problem is not that the researcher does not want to take such factors [selection bias] into account, but the problem is that the data do not contain sufficient information".

Most previous studies have compared wages in both sectors using regression analysis controlling for observable covariates Smith [1976], Moulton [1990], Campos et al. [2017],

[1] World Bank's World Bureaucracy Indicators. Data is from 73 countries, both rich and poor, for the period 2011-2016.

[2] See surveys in Ehrenberg and Schwarz [1986], Bender [1998], Gregory and Borland [1999], Lausev [2014].

Finan et al. [2017], Gindling et al. [2019], which does not address the possibility that workers could select into sectors based on unobservables. For example, suppose that workers that end up in the public sector are not well-suited for the private sector, e.g., due to being more risk averse Bonin et al. [2007], Guiso and Paiella [2008]. This could mean that their counterfactual wages in the private sector are lower than those of workers who actually work in the private sector. Without correcting for this negative selection bias, comparing wages of public and private sector employees would underestimate the public-sector wage premium. Alternatively, there could be positive selection bias, for instance if more able workers are also more motivated by public service Besley and Ghatak [2005].

Another approach used in this literature are panel data models with worker fixed effects Krueger [1988], Rosholm and Smith [1996]. Here the main challenge stems from the few number of observations on sector transitions, which are necessary for identification. Additionally, without a credible source of exogenous separations, the problem of endogeneity of sector choice remains, as pointed out by Moulton [1990]. A third empirical strategy are switching regression models Maddala [1986], Van der Gaag and Vijverberg [1988], Van Ophem [1993]. This strategy attempts to correct for selection bias by first estimating a sector-choice equation using a probit model based on observable characteristics of workers and jobs, and subsequently using the estimated probabilities to correct the original estimates. These models, however, are restricted to using observable covariates to explain sector choice, and also rely on numerous parametric assumptions for identification.

This paper provides a new estimate of public-private wage gaps while controlling for selection bias. The evidence comes from the Brazilian labor market where, like in other developing countries (e.g., India and Indonesia), public sector jobs are allocated via exams. The key innovation is the construction of a novel dataset combining test-taking records *and* labor market outcomes. The idea is to compare individuals who are similar in that (i) they applied for a public-sector job (something that is typically unobserved), and (ii) got similar test results except for falling just above of just below a threshold for getting the job. Thus, for these individuals, it is arguably random (in a regression discontinuity sense) whether they get the public-sector job or not. Additionally, the data on test performance are matched with data on the same individual's wages and sector of employment, which make it possible to compare public and private sector wages for individuals who were effectively randomly assigned to either sector.

Using a fuzzy regression discontinuity (RD) design, I find a positive and large public-sector wage premium of approximately 48 percent. I further estimate the model restricted to high-school and college-educated workers, and find the relative wage gap to be higher for those with lower education levels. The main contribution of the paper, however, is not the size of the estimated gap, but rather to show that the coefficient is almost twice as large as the coefficients from OLS wage regressions.[3] This difference points to a negative selection bias term and highlights the value of the RD approach. To the best of my knowledge, this paper provides the first evidence on the sign and magnitude of the selection bias using data on public-sector exams.

Finally, I conduct a survey experiment with undergraduate students from a large Brazilian university in order to better understand the selection mechanism. I gather data on career prospects and preferences, and, most importantly, elicit subjective wage expectations for both sectors. If the estimated public-sector wage premium was in fact due to negative selection, one would expect students that declare a preference for working in the public sector to have higher expectations for the wage premium compared to students that do not prefer the public sector. That is because characteristics that are unobserved to the researcher are known by the individual. That is precisely what I find: The average expected wage premium among those that prefer the public sector is positive and large, corresponding to approximately 29 percent of mean expected wages. In contrast, students not interested in a government job expect a public-sector wage *penalty.*

This paper relates primarily to the literature on public-private wage gaps. Many previous studies have examined the earnings gap in individual countries, such as Romania Voinea and Mihaescu [2012], Canada Mueller [1998], France Bargain and Melly [2008], Russia Gimpelson et al. [2015], Brazil Foguel et al. [2000], Germany Melly [2005], United States Schanzenbach [2015], among others. There are also papers that study the question using data for multiple countries Mizala et al. [2011], Gindling et al. [2019], Finan et al. [2017]. Most of these

[3] In terms of the size of the wage premium, Gindling et al. [2019] uses survey data for 91 countries and find a wage premium between 40-60 percent for 10 of those countries, and Bewerunge et al. [2012] find a premium of 39 percent for federal workers in the United States.

studies compare wages for private and public sector workers while controlling for observable characteristics, an empirical strategy that does not address the problem of selection bias.

Gonzaga and Firpo [2010] use a different approach from the previous literature and is the paper closest to mine. They estimate a difference-in-differences model with worker fixed effects exploring the privatization of state-owned firms (SOF) 1990's Brazil. Using a dataset of Brazil's entire formal labor market — which is one of the datasets used in this paper —, they compare workers that were laid off following the privatization of SOFs with workers that were laid off from other, private-sector firms. Those who lost their public-sector jobs and went on to work at the private sector experienced decreases in pay that were 11 percent higher than workers fired from private-sector firms who found work at another private firm. This gap is the authors' estimate of the public sector wage premium. This paper, however, only studies workers at SOFs and exclude all workers on general public administration, which accounts for the bulk of employment in the public sector. Moreover, the empirical strategy does not address all possible sources of selection bias. In particular, we do not know which of the workers laid off from private firms ever got close to obtaining a job in the public sector, nor if they ever applied to one, which is something I observe in my dataset.

My results also speak to a more recent literature that tries to understand selection into the public sector. Dal Bó et al. [2013] conduct a field experiment in Mexico as part of a government effort to hire 300 (temporary) workers for a socially oriented position. Both wages and job offers were randomized, and candidates completed intelligence tests and a questionnaire about motivation for public service. The authors find that higher wage offers attract a larger and higher-quality candidate pool, as measured by both test results and by current or past earnings. Moreover, they find no difference in the levels of public-sector motivation between wage conditions. Note, however, that Dal Bó et al. [2013] do not explore the issue of wage gaps, and their results do not rule out the possibility that counterfactual private-sector wages for the (better) applicants are lower than the public-sector wage.

Finally, the paper also adds to the literature on subjective wage expectations and its usefulness in both understanding a particular phenomenon and predicting behavior Manski [2004], Arcidiacono et al. [2012], Reuben et al. [2017]. In particular, I use subjective expected wage distributions to shed light on the mechanism behind the negative selection bias term uncovered in the empirical estimations.

The remainder of the paper is organized as follows. Section 2.2 describes the institutional environment and the datasets. Section 2.3 presents OLS wage regressions that will serve as a benchmark with which to compare the RD results. Section 2.4 discusses the issue of selection bias and the assumptions of the RD design. Sections 2.5 and 2.6 present the results from the RD estimations and the instrumental variables approach, respectively. Section 2.7 breaks the results down by education level, while Section 2.8 discusses the results from the survey experiment. Section 2.9 concludes.

2.2 Institutional Background and Data

2.2.1 Public-sector Exams

The requirement that the government hire via pubic-sector exams is enshrined in Brazil's Constitution.[4] The only exceptions are political appointees, such as ministers and lower-level aides, and elected officials, which combined represent a small fraction of overall employment in the public sector. Whenever the government (federal, state or municipal), a state institution (e.g., public universities) or a state-controlled firm decide to hire workers it must organize a public-sector exam.[5] Moreover, a separate exam is required for each job-employer combination: If, say, the Central Bank and the Labor Ministry each wants to hire an economist, there has to be two separate exams even though both institutions are part of the federal government.

The government has to make an upcoming exam known to the public, which it does by publishing a notice, known as *edital*, in its official daily newsletter. This notice contains details about the job (hours, wage, benefits, etc.), the minimum requirements for admission (education level, prior experience, etc.) and information about the exam itself (date and time, content, classification criteria, etc.). The exams are mostly comprised of written tests; for

[4]The Constitution states that "entry into a public-sector job or position requires prior approval in a public-sector exam of written tests and/or evaluation of degrees(...), except for jobs defined by law as being of free appointment and removal".

[5]For simplicity, I will refer to the three types of employers as the *government.*

some jobs, such as police officers and judges, they may also include physical and psychological components. Participation in the exams is open to the entire working-age population. The only requirements are being up to date with military (only males) and voting (everyone) obligations, completing an online form, and paying a small fee.

The most common arrangement is for the government to outsource the entire selection process to a third-party organization. These *exam organizers* are responsible for creating and grading the exams, and for the logistics of registrations and test-taking. When the selection process is completed, the government is *obligated* to offer the job(s) according to the final ranking provided by the exam organizer: first offer goes to the first place, second offer to second place, and so on. Moreover, as long as there are enough candidates, it has to hire at least as many people as the number of openings it advertised.[6]

There are dozens of such exam organizers in Brazil. In this paper, I use data obtained from two of these companies, both of which non-profit foundations that agreed to share the data for research purposes only. They are concentrated in the southeastern part of the country and are among the 10 or so largest such companies in Brazil. The data cover all public-sector exams handled by these companies from 2007 to 2017. Each entry in the dataset contains a job identifier, the candidate's final grade and ranking, and a few demographics.[7] Job characteristics were manually collected from over 400 public notices (*editais*) and merged with the data on exam results. Lastly, each individual is identified by the national tax number (*cpf*), which allows me to perfectly match test-taking results with labor market outcomes using the RAIS (*Relação Anual de Informações Sociais*; see Section 2.2.2) dataset.

To the best of my knowledge, this is the first time that matched data on exam results and labor market outcomes have been compiled and used for research purposes. As such, there were many steps involved in preparing the dataset.[8] First, I dropped exams for university entrance, non-government jobs, for temporary positions, and for internal promotions.

[6]This obligation has been legally established by a series of judicial decisions, including by the Supreme Court of Brazil. Occasionally, however, the government will hire fewer people than advertised for budgetary or other reasons. Those cases invariably end up in the court system and the rights of the employee usually prevail.

[7]Most candidates do not get assigned a final ranking as each exam will only rank a pre-specified number of the highest-scoring candidates — usually 3 times the number of openings. I come back to this point when discussing the regression discontinuity design on Section 2.4.2.

[8]A detailed discussion of the procedures can be found in the Data Appendix.

Second, I excluded jobs with serious data problems. These include: inconsistent or missing ranking information, jobs allocated to different cities without proper identification, jobs with openings for different departments within the same employer, etc. Third, I restrict attention to individuals aged 18-55 who held a job at anytime during 2017.

Table 2.1 provides a summary of the exams dataset. The final sample covers approximately 3,000 different jobs from 125 employers, and contains exam results for over 2 million individuals. Note, however, that I do not observe the universe of public-sector exams. As discussed further in Section 2.4, the empirical strategy requires only that the subset of exams that I do observe be sufficient in the sense that (just) approved candidates are more likely to end up working in the public sector than (just) not-approved candidates.

Table 2.1: Summary Statistics

Panel A: Individual Test-Takers		
Candidates	N	2,090,022
Hires	N	98,305
Female		0.55
High-school degree		0.46
College degree		0.44
Age	Mean	35.8
Exams Taken	Mean	2.8
Panel B: Public-Sector Exams		
Exams	N	3,092
Openings per exam	Mean	23.6
	Median	4
	Min	1
	Max	2,500
Candidates per opening	Mean	101.5
	Median	45.1
	Min	0.2
	Max	3,077
Education requirement	Less than high-school	0.13
	High-school	0.33
	College	0.52
Gov't level	Municipal	0.43
	State	0.54
	Federal	0.03

Note: Sample used for main analyses. Includes only individuals aged 18-55 and that had a formal-sector job in 2017. Include exams from both organizers.

2.2.2 RAIS

The RAIS is an employer-employee match dataset covering the entire formal sector of the Brazilian economy. It does not include information on the self-employed, those in the informal sector, interns, domestic workers and some other minor categories. The data are assembled in the first quarter of every year by the Brazilian Labor Ministry. Each entry in the dataset is a summary of both employer and employee characteristics for each formal job held during the previous year. The information is reported by the employer, who is subject to fines in case of non compliance. From the workers' perspective, the information on RAIS is used to define eligibility for a number of worker-related social programs. Information on the worker includes average wages, hours worked, education, and gender; data on the employer include legal status, number of employees, geographic location, and many others. Crucially, both workers and firms are identified by their national tax numbers, *cpf* and *cnpj*, respectively.

For this paper, the two most important variables are average hourly wages and the legal status of the employer (either pubic or private). Earnings information are from average wages during the year. The hourly wage is calculated based on the average number of hours worked in a week. The variable used to identify the legal status of the employer (*natureza juridica*) assumes many different values. Examples are "municipality", "public-traded corporation", "limited liability corporation", "mixed enterprise", etc. I classify as public sector direct governments (municipalities, states, federal government), government-controlled institutions (e.g., public health agencies, public universities, etc.) and government-controlled companies (e.g., state oil company, state electricity generator, etc.). All other legal entities are classified as private-sector employers[9]. For the cases when a worker has more than one entry in a year, I first keep the jobs that were active as of December 31st of the corresponding year. If multiple jobs are still left, I keep the job with the highest hourly wage and, if still tied, I select the job with the earliest start date. If multiple jobs are still left, I choose one at random.

The main limitation of the RAIS dataset is that is is restricted to the formal sector of

[9]See Data Appendix for more details.

the economy. There are two reasons why this limitation is not particularly important for this paper. First, more than 92 percent of the test-taking individuals held a formal sector job at some point during the period studied. This indicates that the relevant labor market for those taking public-sector exams is the formal sector, not the informal one.[10] Second, the primary goal is to estimate the sign and magnitude of the selection bias, and to do so I use as benchmark wage regressions using the same RAIS dataset.

2.3 Wage Regressions

Before proceeding with the regression discontinuity estimates, it is worth considering the evidence from simple wage regressions. Table 2.2 reports OLS regressions of log hourly wages on years of education, experience, experience^2, gender, region/state fixed effects and an indicator for public-sector job. I use a random sample of approximately 1.2 million workers who held a job in 2017. The sample was chosen to match the distribution of test-takers' labor market outcomes on key variables, namely, years of education, state or region, and age. Results are shown for the entire sample and also separately for high-school and college graduates, the two categories that correspond to approximately 85% of the exam test-takers. Table 2.2 presents evidence of a substantial wage gap in favor of the public sector. Considering the full sample first, public-sector workers with the same years of education, experience, gender and from the same region or state earn, on average, 25% more compared to private-sector workers. Moreover, the observed wage gap is larger for high-school graduates (42%) than for college graduates (13%).

[10] Albrecht et al. [2018] finds a similar pattern in a study of the Colombian labor market.

Table 2.2: Wage Regressions for Random Sample of Workers

	Full Sample	High-School Graduates	College Graduates
Coefficient on public-sector	0.249***	0.418***	0.130***
	(0.002)	(0.003)	(0.002)
% Public-Sector	0.17	0.08	0.27
Mean Hourly Wage	22.73	11.71	35.11
Median Hourly Wage	13.16	9.03	24.33
R^2	0.420	0.436	0.150
N	1,257,019	542,810	537,840

Note: OLS regression on random sample of workers who held a formal job in 2017 and were between 18 and 55 years old. Dependent variable is log hourly wages and is restricted to be at least 75% of the 2017 federal minimum and at most equal to the public-sector legal maximum (245 BRL), for public-sector workers, or two times the public-sector legal maximum for private-sector workers. Controls include years of education, experience, experience^2, gender, and state or region fixed-effects.

2.4 Identification Strategy

2.4.1 The Problem of Self-Selection

To understand the role of selection in estimates of the public-private wage gap, it is useful to think in terms of potential outcomes Rubin [1974, 1977], Angrist and Pischke [2008]. Let Y_i denote hourly wages for worker i, and let Y_{0i} and Y_{1i} represent potential wages for the same individual in the private and public sectors, respectively. Define the indicator variable $G_i \in \{0, 1\}$ which is 1 if the worker has a public-sector job and 0 if she has a job in the private sector. For any given worker, we only get to observe Y_{1i} if $G_i = 1$, or Y_{0i} if $G_i = 0$. Hence, we can write observed wages in terms of potential wages as follows:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})G_i \tag{2.1}$$

Suppose we take a random sample from the working population and compare the wages of private and public-sector workers. We would obtain the following:

$$
\begin{aligned}
\mathbb{E}[Y_i|G_i = 1] - \mathbb{E}[Y_i|G_i = 0] =& \mathbb{E}[Y_{1i} - Y_{0i}|G_i = 1] \\
&+ \mathbb{E}[Y_{0i}|G_i = 1] - \mathbb{E}[Y_{0i}|G_i = 0]
\end{aligned} \tag{2.2}
$$

The first right-hand-side term in (2) is the average treatment effect on the treated. It is the difference, for public-sector workers, between their current wages and the wages they would have received had they worked in the private sector. The second line of (2) is the *selection bias* component. It is the difference between potential private-sector wages for those currently working for the government and those working in the private-sector.

Note that the sample used includes workers that have taken *at least* one public-sector exam. It is possible, however, that part of the selection is coming from the decision to take an exam. In practice, this channel is unlikely to be of first importance in the case of Brazil. Taking a public-sector exam is not particularly expensive or time demanding. Exams typically happen on weekends and candidates below the poverty line can apply for a fee waiver. Given the low overall cost of taking a public-sector exam and the large potential benefits, one would expect a large number of people to take exams even if their chances are small and the public-sector is not their preferred option. In the survey discussed on Section 2.8, for instance, most students that say the public-sector is not their preferred option also say there is a high chance they will take at least one such exam.

The sign and magnitude of the selection bias is the main object of interest of this paper. It is not clear, a priori, if the sign on the selection bias should be negative or positive. I consider each possibility in turn.

2.4.1.1 Positive Selection Bias Suppose that a worker's productivity is the same for the private and public sectors. Assume also that government wages are set above the average wages adjusting for education and experience. In that case, if the hiring mechanism used by the Brazilian government is successful in sorting the most able candidates, the result would be a positive selection bias: counterfactual private-sector wages for those currently occupying public-sector jobs would be higher than potential outcomes for private-sector workers. In this case, estimates of $E[Y_{1i} - Y_{0i}|G_i = 1]$, the true public-sector wage premium, would be *smaller* than the simple wage comparison $E[Y_i|G_i = 1]$ - $E[Y_i|G_i = 0]$.

2.4.1.2 Negative Selection Bias A negative bias would result if, for instance, the hiring mechanism adopted by the government favored individuals with lower potential private-sector earnings. Since the public-sector has high job security and is less responsive to productivity, it is conceivable that it disproportionately attracts candidates with lower overall productivity. Moreover, this individuals would have lower opportunity costs and would thus be more willing to invest the time necessary to be approved in a public-sector exam. If this were the case, estimates of the true public-sector wage premium would be *larger* than the naive wage comparison.

2.4.2 Regression Discontinuity

The regression discontinuity (RD) design makes use of institutional idiosyncrasies to exploit variations in outcomes close to a fixed threshold. The main assumption necessary for identification is that the values of unobserved covariates vary smoothly around the threshold Hahn et al. [2001]. In the context of exam test-taking, the idea is that two people on either side of the cutoff are very similar in terms of underlying ability, or motivation, or other unobserved components that are correlated with the outcome of interest. If this assumption is satisfied and we further observe a discontinuity in the probability of treatment assignment (i.e. working in the public sector), then observed differences in the outcome of interest can be interpreted as treatment effects.

Usually, applications of the RD design use information from few selection criteria (for example, exams), but many individuals. For example, Abdulkadiroğlu et al. [2014] study the impact of classroom composition on learning outcomes for students in grades 7-12. The authors explore discontinuities in admissions tests for six highly-selective schools in the United States. Since there are many applicants and few schools, it is necessary to determine which students can be deemed as being *close* to the threshold on either side. This is a common issue in RD applications and has given rise to different methods of estimation and of optimal bandwidth selection Imbens and Kalyanaraman [2012], Calonico et al. [2014]. In contrast, the unique dataset used in this paper contains the test results for *many* individuals across *many* different exams. For each exam, I essentially compare the labor market outcomes of the last individual to have been approved to the outcomes of the first individual *not* to be approved. Since I have the results for many different exams, I obtain a large sample while focusing on those very close to the threshold.

Another benefit of the data used here is that each exam allocates, on average, a small number of individuals to treatment. This is not the case the case for most RD applications. For example, Ost et al. [2018] study the returns to college for academic marginal students by exploring discontinuities on universities' dismissal rules based on GPAs. Their sample includes 20,000 at-risk students, of which more than 10,000 end up being dismissed. In order for the RD assumptions to be valid, the authors have to restrict attention to only a fraction of those students that are closest to the dismissal threshold. Since the RD design focuses on individuals close to the assignment cutoff, its treatment estimates are referred to as local average treatment effects. In my dataset, on the other hand, 26 percent of the exams advertise only 1 opening and fully 63 percent of them advertise 5 or less. For that reason, the estimated treatment effect is *less* local than usual, as for most exams the sample used for the RD estimates includes all of the treated individuals.

I use the national tax number to combine information on public-sector exam results and labor market outcomes for each individual test-taker. The empirical strategy results in a discrete running variable, as explained in more detail later. Note also that if I had access to *all* exams taken by individuals in the sample, I would be able to use a sharp RD design. That is because a public-sector exam is the only available path to obtaining a public-sector job. However, since I observe the exam results of only two exam organizers, the identification comes from the discontinuity in the *probability* of obtaining a public-sector job. The crucial assumption for identification is that potential outcomes vary continuously with the variable used to determine treatment status - the *running variable* - while the observed outcomes have a jump, or discontinuity, at the relevant values Hahn et al. [2001], Imbens and Lemieux [2008].

2.4.3 The Running Variable

Since I have individual results from over 3,000 different exams, I construct the (discrete) running variable as to compare the labor market outcomes of a candidate that was the *last* to be approved with the outcomes of a candidate that was the *first* not to be approved in the exam. One way to do this is to use the number of openings advertised in the official notice to construct such a variable. Note, however, that being approved in the exam does not mean the candidate will *take* the job. Moreover, during the legal time window the government is allowed hire *more* workers than it advertised. Finally, many exams list 0 as the number of targeted hires, and most of those end up hiring multiple workers. For these reasons, it is not possible to use the number of advertised openings to define the running variable.

To construct a valid running variable, I examined the official notice for each one of the exams and collected, among other information, the tax identification number of the exam-specific employer. Using both the individual and employer tax number, I was able to verify, for each classified test-taker, whether she ended up working for that particular employer. More specifically, for each classified candidate, I check if she started working for the government employer at any time during a three-year window starting 60 days after the exam had taken place. I consider a three-year window because most of the exams have a duration of two years, which can be renewed for another two years at the discretion of the employer. Only during that period is the government legally allowed to hire among the classified workers. The duration time starts to count after the exam has been validated, which happens at least 2-3 months after the exam has taken place[11].

The running variable is defined as the distance, in ranking positions, to the last person to have been hired by the exam-specific employer. Candidates with a positive value for the running variable have been *approved* in the exam and, as such, had the option to take the job. Those with negative values for the running variable did not get approved and did not get a job offer from that employer. For candidates that were classified at public-sector exam e, the running variable R_i^e is constructed as follows:

$$R_i^e = Rank_{LH}^e - Rank_i^e \tag{2.3}$$

[11] See Appendix B for more details on constructing the running variable.

where $Rank^e_{LH}$ is the ranking of the last candidate to be *hired* by the exam-specific employer, and $Rank^e_i$ is the ranking of candidate i on exam e. For example, consider an actual exam to hire four analysts with a degree in Economics, Business or Accounting for a state public health institution in 2013. The last candidate to be hired by that institution, satisfying the timing requirements, obtained a final ranking of 5. Every candidate with a ranking better than 5 will then be assigned a positive running variable. In this particular example, the candidate ranked 3^{rd} is the first one still in the dataset[12], and thus receive a running variable equal to 1, followed by the candidates with ranking equal to 2 and 1, respectively. Similarly, candidates that obtained rankings lower than 5 were assigned negative values for the running variable.

Following de Chaisemartin and Behaghel [2019], I do not include the last candidate hired in the analysis. The intuition behind it is simple: candidates that are approved in the exam can decide not to take the job, and non-approved candidates could have decided not to take the job had they been approved instead. Hence, an approved candidate that is known to have accepted the job is fundamentally different from the comparison groups. In the context of random assignment to waitlists for oversubscribed training programs, de Chaisemartin and Behaghel [2019] shows that including the last person to take up the offer result in inconsistent estimates.

2.4.4 RD Sample and Discontinuities

Since there are over 3,000 different jobs, with different wages and educational requirements, it is important to construct the RD sample to ensure comparability among individuals on equally-distant sides of the admissions cutoff. Consider, for example, an exam that hires only two workers. For that exam, I observe individuals with values of 1 and 2 for the running variable, but not for higher values. Since there are not many openings, there will not be many approved candidates. On the negative side, however, I observe candidates with very low values for the running variable. That is simply because there are many more non-approved than approved candidates.

[12] The candidate ranked 4^{th} met one of the exclusion criteria: not aged between 18-55 in 2017, lower than minimum wage, not holding a formal sector job in 2017, etc.

Recall that the empirical strategy compares the outcomes of candidates that are equally distant from the approval cutoff, and since candidates applying to the same job are most similar in terms of observed and, arguably, unobserved characteristics, I construct the RD sample to guarantee that any exam that has an observation for running variable x also has an observation for -x. For that reason, I restrict the RD sample to only include exams that have at least one observation for each value x and -x for the running variable, for $x \in [-15, 15]$.

For the RD results, I estimate the following equation on a number of different outcomes:

$$Y_i = \alpha + \delta R_i + \beta C_i + \gamma R_i C_i + \epsilon_i \tag{2.4}$$

where Y_i is the outcome of interest (indicator for job in the public sector, hourly wage, and others), R_i is the running variable defined above, and C_i is an indicator variable which is equal to 1 if R_i is positive. An observation is a person-exam, and errors are clustered at the ranking level Kolesár and Rothe [2018]. Since some exams have multiple candidates with the same ranking value, the regression weighs each observation by the inverse of the number of units of the same exam at each different value for the running variable. This is equivalent to using variables that are the average of each exam-running variable combination.

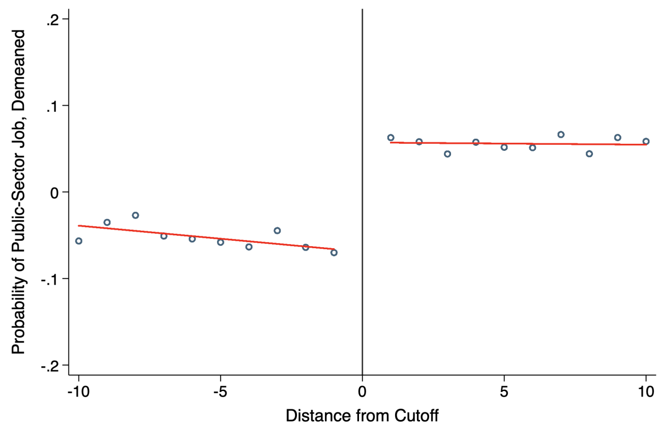
2.5 Regression Discontinuity Results

2.5.1 Probability of Public-Sector Employment

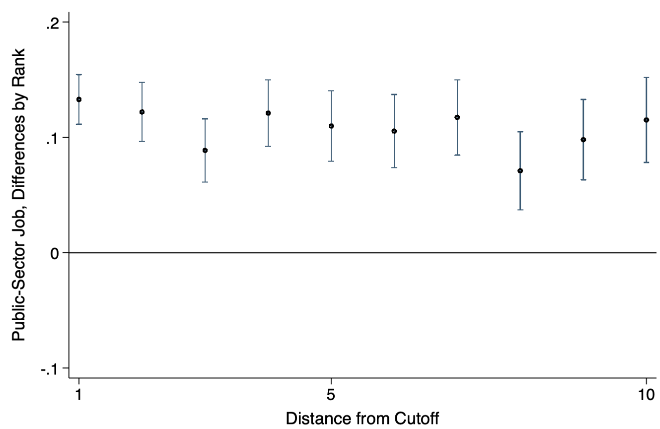
I first examine whether being just above the exam-specific approval cutoff has any impact on the probability of holding a public-sector job in 2017, conditional on having a formal-sector job. Figure 2.1a shows a clear discontinuity at the approval cutoff, suggesting that the fuzzy RD approach is valid. The graph plots the (demeaned) fraction of candidates working in the public sector as a function of the running variable for the interval -10 to 10. The fitted lines come from a regression of the indicator for public-sector job on the running variable, an indicator for the running variable being above zero, and their interaction (see equation 4). Being just above the approval cutoff increases the likelihood of holding a public-sector job by approximately 12.5 percentage points. Figure 2.1b plots the differences in probabilities by absolute value of the running variable. These are direct comparisons between individuals that applied to the same set of jobs and that are equally distant from, but on different sides of, the approval cutoff. The confidence intervals come from regressions estimated separately for each pair $(x, -x)$, for $x \in [1, 10]$. Every one of the pairwise differences is positive and statistically significant.

Finally, note that candidates that fall below the cutoff also have high rates of public-sector employment: approximately 65% of non-approved candidates in my RD sample end up working in the public sector. This indicates that individuals in the sample are also taking exams from other organizers. This fact, while weakening the power of the instrument, does not bias the results as long as the assumptions of the RD design are valid.

Finally, I look at outcomes for the latest available year (2017) for a number of reasons. Most important of those is that, since I do not observe the universe of public-sector exams, looking at outcomes after, for example, two years after an individual's exam would artificially inflate the discontinuity. For example, an individual that got really close to getting a particular job, but didn't, is likely to obtain another one, outside of my sample, in the following years. If I were to restrict the labor market outcomes to this hypothetical 2-year window, I would wrongly assign that individual as not having obtained a public-sector job. A second important reason relates to the exclusion restriction in the instrumental variables estimates. The IV regression requires that the instrument affect the outcome variable (wages) solely via the probability of obtaining a job in the public sector. If failing to obtain a passing grade affects earnings potentials in other ways (e.g., negative market signaling, impact on morale, etc.), the exclusion restriction could be compromised. Focusing on outcomes for the latest available year means that, for most candidates, exam results and the measurement of labor market outcomes are many years apart. It is thus less likely that those threats to the exclusion restriction are valid.



(a) Discontinuity in Public-Sector Status



(b) Differences in Public-Job Status, Ranks x and -x

Figure 2.1: Discontinuity in Probability of Public-Sector Job

Note: Figure (A): Discontinuity on restricted sample: 0.126^{***} (0.009), N=63,720. Linear regression with indicator for public-sector job as dependent variable. Robust standard errors clustered at the running variable. Figure (B): Differences in share of workers in public-sector by absolute values of running variable. Confidence intervals (95%) come from regressions with dummy indicating positive value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

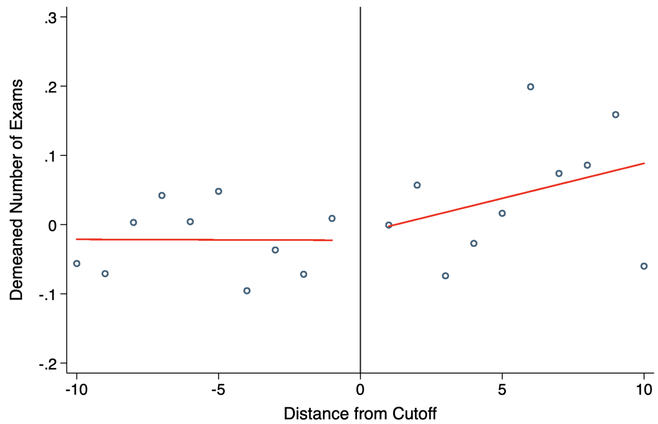
2.5.2 Specification Checks

As discussed in section 2.4.2, the crucial assumption for identification in a RD design is that unobserved confounds vary smoothly across the threshold used to assign treatment status. By definition, it is not possible to directly test an assumption on unobservables. It is possible, however, and highly desirable, to examine the distribution of key variables across the threshold. Usually the most important specification test for a regression discontinuity analysis is the density of the running variable around the cutoff McCrary [2008]. The density test is used to determine if participants are able to manipulate the variable at the threshold boundary, which would result in many more individuals being just above the cutoff than just below (or vice versa). Since the running variable here is defined as the rank position in a particular exam, there is, by construction, the same number of observations for values 1 and -1 (and 2 and -2, and 3 and -3, etc.) of the running variable, which makes sorting impossible. There is, however, one variable that candidates could plausibly manipulate: the number of exams taken. Specifically, if approved candidates take more exams than comparable non-approved candidates, the assumption of similar unobserved covariates could be deemed less convincing.

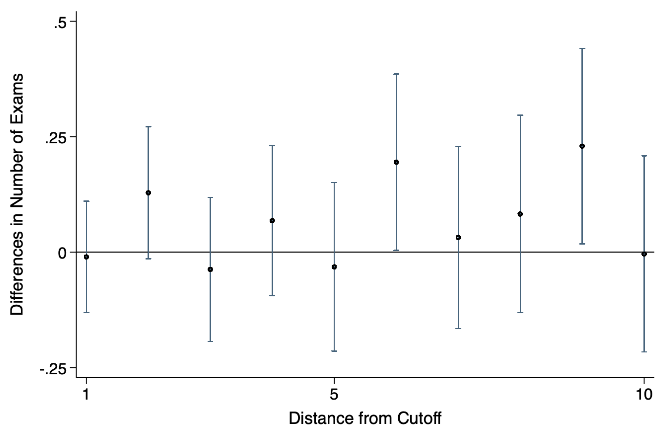
Figure 2.2a plots the (demeaned) average number of exams taken in the period 2007 to 2017 as a function of the running variable. The fitted line comes from a regression with the demeaned number of exams as dependent variable. As before, Figure 2.2b plots the estimated difference by absolute value of the running variable. I find no evidence of manipulation of the number of exams attempted: the estimated coefficient for the the bandwidth [-10,10] is very low (0.009) and not statistically significant.

Table 2.3 reports discontinuity estimates on various observable variables using three different ranking bandwidths. Focusing on the specification checks, there is no evidence of a discrete jump on either age or the gender of the candidate. I also check for discontinuities in the share of candidates with each education level, since high-school (college) graduates can compete for jobs that require *less* than a high-school (college) degree. That is, if test-takers on either side of the cutoff are truly comparable, then one would expect educational attainment to vary smoothly across the threshold. That is precisely what I observe. Coefficient estimates for an indicator of high-school or college degree are small (less than 1 percentage point) and not statistically significant.

Finally, I test for discontinuities on public-sector wages. The goal is to gauge how representative the sample of government jobs used here is. Recall that most of the test-takers



(a) Discontinuity on Number of Exams Taken



(b) Differences in Number of Exams, Rank X and -X

Figure 2.2: Discontinuity on Number of Exams Attempted, 2007 to 2017

Note: Figure (A): Discontinuity on restricted sample: 0.010 (0.047), N=63,719. Linear regression with number of exams attempted as dependent variable. Robust standard errors clustered at the running variable. Figure (B): Differences in average number of exams by absolute value of running variable. Confidence intervals (95%) come from regressions with dummy indicating positive value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

who are not approved in the observed exams end up working for the government anyway, which means they are taking exams from other organizers. And if different organizers are associated with different types of government jobs, we would observe a discontinuity in terms of wages for those in the public sector. Table 2.3 shows that, on average, candidates that

Table 2.3: Discontinuity on Selected Variables and Bandwidths

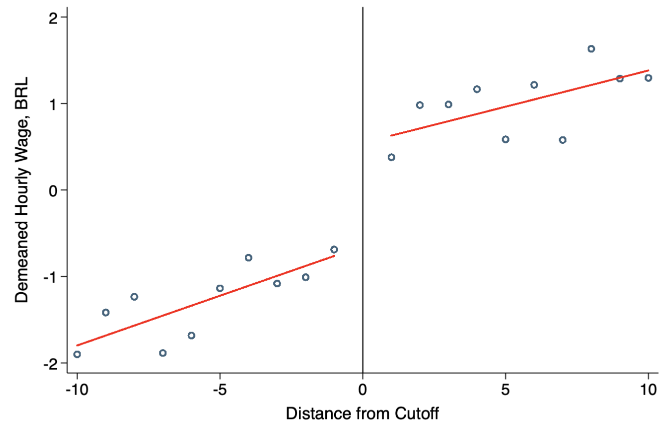
	Ranking Bandwidth		
Variable	$[-5, 5]$	$[-10, 10]$	$[-15, 15]$
Public-Sector Job	0.133***(0.007)	0.126*** (0.009)	0.119*** (0.009)
N	36,664	63,720	82,792
Hourly Wage	1.216** (0.413)	1.196*** (0.249)	1.305*** (0.234)
N	36,516	63,441	82,420
Number of Exams	0.037 (0.065)	0.009 (0.047)	-0.006 (0.038)
N	36,663	63,719	82,791
Age	0.301 (0.203)	0.163 (0.132)	0.064 (0.130)
N	36,662	63,718	82,789
Female	-0.007 (0.004)	0.002 (0.002)	0.000 (0.004)
N	36,664	63,720	82,792
High-school Degree	0.010 (0.012)	0.009 (0.010)	0.009 (0.009)
N	36,664	63,720	82,792
College Degree	-0.008 (0.012)	-0.008 (0.009)	-0.012 (0.008)
N	36,664	63,720	82,792
Hourly Wage, Public-Sector	-0.058 (0.360)	-0.428 (0.348)	-0.289 (0.405)
N	24,370	41,253	53,674

Note: Indicator for public-sector job for 2017 and includes government agencies and government controlled companies. Hourly wage in 2017 is measured in Brazilian Reais (BRL). Number of exams include any exam taken from the two organizers in the period 2007 to 2017. Age, in years, is measured in 2017.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

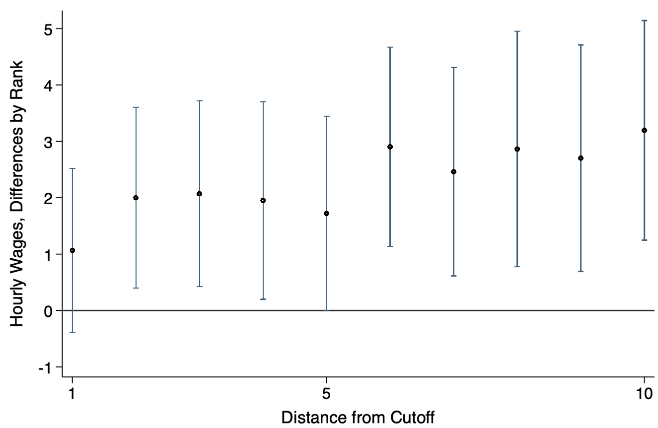
failed to be approved in an exam in my sample, but nonetheless ended up working in the public sector, earn slightly *higher* wages than candidates that were approved and took the public-sector job. The difference, however, is not statistically significant.

2.5.3 Discontinuity in Hourly Wages

The main outcome of interest is hourly wages for those working in the formal sector in 2017. Figure 2.3a plots the average demeaned hourly wages as function of the running variable. Estimates of equation (4) with demeaned hourly wages as dependent variable for bandwidth [-10,10] reveal a sharp discontinuity at the approval cutoff. Candidates just above the cutoff earn, on average, 1.19 Brazilian *Reais* (BRL; $p < 0.01$) more per hour than workers just below the cutoff. The fitted lines comes from regression (4) with demeaned hourly wages as dependent variable.



(a) Discontinuity in Demeaned Hourly Wages



(b) Differences in Hourly Wages, Ranks X and -X

Figure 2.3: Discontinuity in Hourly Wage

Note: Figure (A): Discontinuity on restricted sample: 1.192^{***} (0.249), N=63,441. Linear regression with demeaned hourly wage as dependent variable. Robust standard errors clustered at the running variable. Figure (B): Differences in hourly wages by values of running variable with opposite signs. Confidence intervals (95%) from regressions with hourly wage as dependent variable and dummy indicating positive value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

As discussed before, most of the jobs have observations at negative values of the running variable, since there are many more candidates than openings. However, exams that have few openings usually do not have any observations for values 3 and above of the running variable. To ensure comparability, the RD sample keeps, for each absolute value of the running variable, only those jobs with at least one observations on either side of the cutoff. Figure 2.3b plots the estimated difference in hourly wages by the absolute value of the running variable, where the confidence intervals come from the similar regressions estimated separately for each pair $(x, -x)$ of the running variable, for $x \in [1, 10]$. For each distance from the cutoff, average wages for approved candidates are higher than average wages for non-approved candidates, and 9 out of 10 pairwise comparisons are statistically significant.

2.6 IV Estimates of the Public-Sector Wage Premium

In this section, I use the discontinuity as an instrument and estimate the returns to a public-sector job. The IV estimate is simply the ratio of the estimated discontinuity in hourly wages to the estimated discontinuity in the probability of public-sector employment Angrist and Pischke [2008]. Table 2.4 presents the IV estimates for three different bandwidths, with and without exam-fixed effects. Focusing on the estimates with the bandwidth [-10,10] and exam fixed-effects, the IV estimates result in a large and positive public-sector wage premium. I estimate that an individual earns, on average, 48% more in a public-sector job versus a private-sector job. All of regression specifications include controls for gender, years of education and experience, number of tests taken, and number of tests attempted.

First, note that these estimates are larger, in both absolute and relative terms, compared to the estimates from the Mincerian regressions. Using a random sample of 1.2 million workers, the estimated wage gap from a wage regression was 25% (see Section 2.3). This is evidence, therefore, that the selection bias discussed in Section 2.4 is substantially *negative*: potential private-sector earnings for workers who sort into the public sector are lower than private-sector wages for comparable workers who sort into the private sector. Thus, failing to control for selection bias leads to an underestimation of the public-sector wage premium.

In order for the IV estimates to have a causal interpretation, the exclusion restriction need to be valid. In the present context, it means that being just above the approval cutoff for an exam impacts earnings solely via the increased probability of obtaining a public-sector

Table 2.4: IV Estimates of Returns to Public-Sector Employment

	Ranking Bandwidth					
	[-5,5]	[-5,5]	[-10,10]	[-10,10]	[-15,15]	[-15,15]
Coefficient on Public-Sector	0.489***	0.475***	0.523***	0.485***	0.538***	0.504***
	(0.062)	(0.084)	(0.058)	(0.062)	(0.054)	(0.059)
Exam Fixed-Effects	No	Yes	No	Yes	No	Yes
F-statistic	139.38	155.76	156.70	157.07	171.04	171.90
Mean Hourly Wage	26.75		25.35		24.80	
Median Hourly Wage	20.53		20.08		19.89	
N	36,513		63,436		82,414	

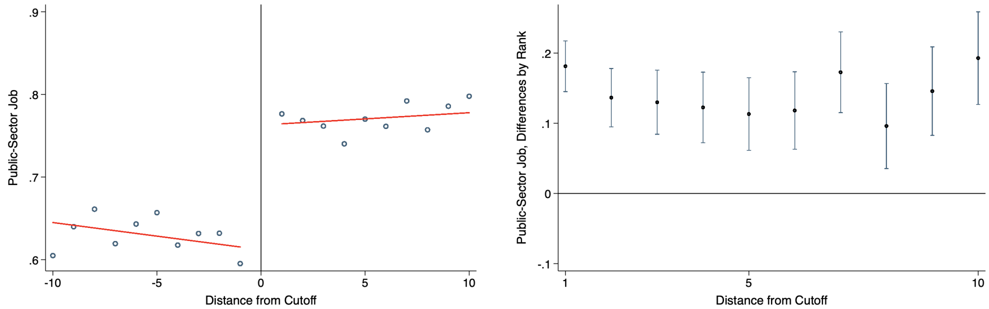
Note: Instrumental variables estimates for the public-sector wage premium. Depended variable is log of hourly wages and excluded instruments are the discontinuities in the probability of holding a public-sector job in 2017. For all specifications an observation is a person-exam, and all regressions include controls for years of education, experience, experience^2, gender, state or region fixed effects and number of exams taken.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

job. One possible threat to the exclusion restriction is if negative exam results affect wages directly via another channel, such as lower morale or negative market signaling. Since labor market outcomes are measured in 2017 and the exam results are for the years 2007 to 2015, there is a span of at least two years between test results are labor market measures. This time gap makes it less likely that effects on morale or signaling could affect wages directly. Finally, as to the signaling value of test results, note, first, that it does not matter for public-sector employment; as to the private-sector, I have found no evidence that employers actually use exam results as screening devices.

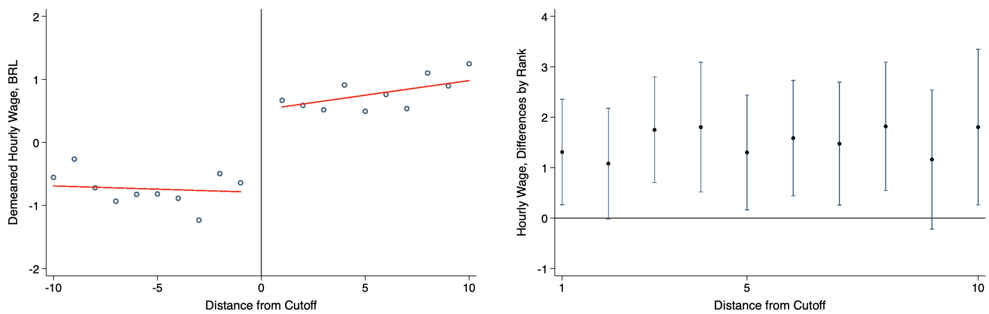
2.7 Heterogeneity by Education Level

The literature on public-private wage usually finds returns that are larger for lower-educated individuals, which tend to decrease, or even reverse sign, for higher-education groups. In order to look at the different education groups separately, I construct specific RD samples for high-school and college graduates, following the same steps as before. That is because individuals with different educational levels often take an exam for the same job.

For example, I observe many college-educated candidates taking exams for public-sector jobs that require only a high-school degree.



(a) Public-Sector Job: Discontinuity and Pairwise Differences



(b) Hourly Wages: Discontinuity and Pairwise Differences

Figure 2.4: Discontinuity in Public-Sector Job and Hourly Wages, High-School

Note: Figure (A): Discontinuity on restricted sample: 0.151*** (0.021), N=27,102. Linear regression with indicator for public-sector job as dependent variable. Robust standard errors clustered at the running variable. The graph on the right plots the differences in the indicator for public-sector job, by absolute value of the running variable. Figure (B): Discontinuity on restricted sample: 1.309*** (0.149), N=26,984. Linear regression with demeaned hourly wage as dependent variable. Robust standard errors clustered at the running variable. The graph on the right plots the differences in hourly wages, by absolute value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

Figure 2.4 presents the RD results for workers with a high-school degree as of 2017. Figure 2.4a and Figure 2.4b plot the discontinuities at the cutoff and the pairwise differences for the probability of public-sector employment and for hourly wages, respectively. High-school graduates just above the approval cutoff are 15.1 percentage points more likely to be working in the public-sector than those just below it. I also estimate a large and positive discontinuity on demeaned hourly wages of about 1.31 BRL ($p < 0.01$). Finally, the pairwise comparisons for the each absolute value of the running variable reveal the differences are always positive and almost always significant for each point in the interval [-10,10].

Figure 2.5 presents the results for individuals with a college degree as of 2017. Figure 2.5a and Figure 2.5b plot the discontinuity and the pairwise differences in the probability of public-sector employment and hourly wages, respectively. College graduates just above the approval cutoff are 10.3 percentage points more likely to be working in the public-sector than those just below it, and earn, on average, 1.11 BRL ($p < 0.05$) more. As with high-school graduates, the pairwise difference are all the in same direction, and most of them are statistically significant.

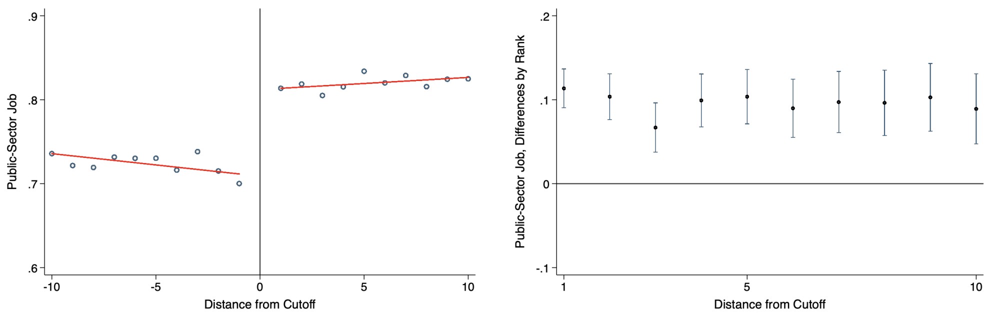
Finally, Table 5 presents the IV estimates of the returns to public-sector employment for each education group separately. For both groups, the IV estimates, which take the selection bias into account, are higher than the OLS estimates. This indicates that the selection bias term is negative for both groups. The difference, relative to the gap estimated in wage regressions, is much higher for college graduates (45 percent versus 13 percent) than for high-school graduates (50 percent versus 42 percent).

2.8 Survey Experiment

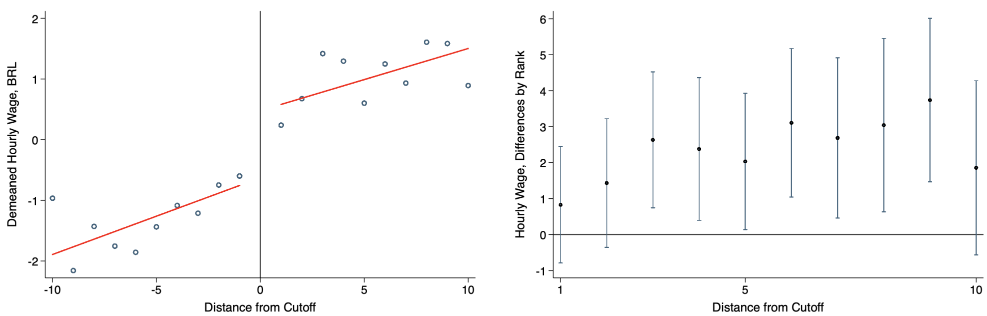
To better understand the selection mechanism, I conducted a survey with undergraduates from a Brazilian university. The survey was administered online and was sent to students via the university administration. Students who completed the survey were entered to win one of 75 gift-cards for a local bookstore worth 100 BRL (49 USD, 2018 PPP dollars). Among many other questions, I asked students about their preferred sector to work for and about their expected future wages.[13] A total of 845 students provided complete answers.

For each student, I elicited subjective wage expectations for both an entry-level position and for a job after 10 years of experience, and for both the private and public sectors. The average age at graduation is 24, hence in the latter case students were asked to estimate their

[13] See Appendix C for more details about the survey, including the translated questionnaire.



(a) Public-Sector Job: Discontinuity and Differences by Running Variable



(b) Hourly Wages: Discontinuity and Differences by Running Variable

Figure 2.5: Discontinuity in Public-Sector Job and Hourly Wages, College Graduates

Note: Figure (A): Discontinuity on restricted sample: 0.103*** (0.010), N=38,060. Linear regression with indicator for public-sector job as dependent variable. Robust standard errors clustered at the running variable. The graph on the right plots the differences in the indicator for public-sector job, by absolute value of the running variable. Figure (B): Discontinuity on restricted sample: 1.107**(0.416), N=37,938. Linear regression with demeaned hourly wage as dependent variable. Robust standard errors clustered at the running variable. The graph on the right plots the differences in hourly wages, by absolute value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

Table 2.5: IV Estimates of Returns to a Public-Sector Job, by Education Level

	Ranking Bandwidth					
	[-5,5]	[-5,5]	[-10,10]	[-10,10]	[-15,15]	[-15,15]
Panel A: High-School Graduates						
Coefficient on Public-Sector	0.429***	0.405***	0.484***	0.495***	0.564***	0.565***
	(0.046)	(0.038)	(0.068)	(0.074)	(0.085)	(0.086)
Exam Fixed-Effects	No	Yes	No	Yes	No	Yes
F-statistic	231.56	239.77	59.41	57.98	101.17	97.54
Mean Hourly Wage	16.49		16.20		16.08	
Median Hourly Wage	14.71		14.84		15.02	
N	15,571		26,983		35,846	
Panel B: College Graduates						
Coefficient on Public-Sector	0.275***	0.304***	0.459***	0.483***	0.434***	0.461***
	(0.083)	(0.095)	(0.153)	(0.159)	(0.123)	(0.127)
Exam Fixed-Effects	No	Yes	No	Yes	No	Yes
F-statistic	80.51	75.36	72.51	71.80	98.89	102.31
Mean Hourly Wage	32.08		30.93		30.52	
Median Hourly Wage	25.32		24.65		24.44	
N	23,407		37,937		47,314	

Note: Instrumental variables estimates for the public-sector wage premium. Dependent variable is the log of hourly wages and excluded instruments are the discontinuities in the probability of holding a public-sector job in 2017. For all specifications an observation is a person-exam, and all regressions include controls for experience, experience^2, gender, state or region fixed effects and number of exams taken.
***p¡0.01 **p¡0.05 *p¡0.1

earnings at around age 34, which is similar to the average age of workers in the labor market data (33.94). For each question, subjects were first asked about the minimum and maximum wages they could conceivably get. I then used these bounds to created four intermediate points and elicited subjects' subjective probability of their wages being at least as high as each intermediate value. Finally, I used those four CDF points to fit a log-normal distribution for each one of the respondents.

Results from the survey could be informative for two main reasons. First, the selection effect could operate via the mere choice to take an exam if, for example, only those with a low outside option choose to take an exam. The test-taking data, however, only includes information for individuals that have taken at least one public-sector exam. As argued before, the characteristics of the exams make it unlikely that this is the main selection channel. Indeed, results from the survey support this conjecture. Even though half of the participants do not list the public sector as their preferred outcome, only 8.4 percent said their chances of taking a public-sector exam were "low" or "very low".

Second, data on wage expectations can shed light on the underlying mechanisms. In particular, assume that students know more about their own abilities and motivations than does the researcher. Since public-sector wages are less responsive to effort, one would expect the students with lower (unobserved) ability to sort into the public-sector *and* to expect lower private-sector wages. That is, students that declare a preference for the public sector would display a higher expected public-sector wage premium. Figure 2.6 plots the expected public-private wage gap for those that prefer the public sector versus those that are either indifferent of that prefer the private sector. The graph clearly shows the distribution of the public-sector preferred group shifted to the right. For that group, the expected wage gap is 16.41 BRL (p ¡ 0.000, se = 1.61), while those that do not prefer the public sector expect a small *negative* wage gap equal to -2.92 BRL (p = 0.06, se = 1.54).

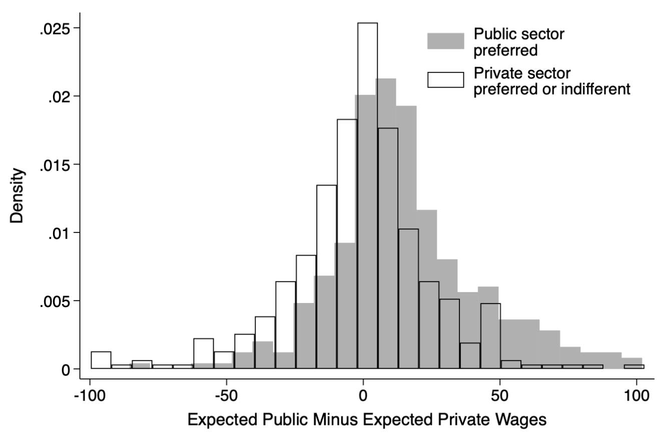
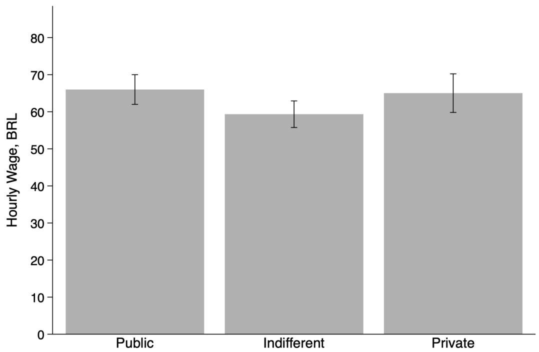


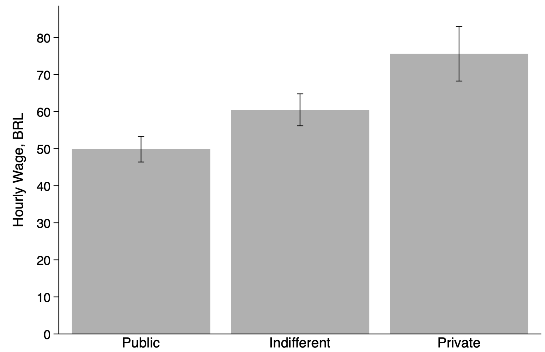
Figure 2.6: Expected Public-Private Wage Gap by Sector of Preference

Note: Histogram of individual expected (mean) wage gaps after 10 years of experience. Preference for sector comes from answer to question: "In case you decide to look for a job after graduating, would you prefer a job in the public or the private sector? (1) Public sector; (2) Private sector; (3) Indifferent"

Note also that, consistent with negative selection on unobserved ability, differences in the expected wage gap come almost entirely from differences in expected wages in the private sector. Figure 2.7 plots the expected private and public-sector wages by choice of sector. There is no significant difference in expectations regarding public-sector wages. Expected private-sector wages, on the other hand, are markedly different between groups: students that prefer the public-sector have significantly lower expectations for private-sector wages, followed by the group of students that are indifferent between the sectors. Those that declare a preference for private-sector jobs, on the other hand, have the highest wage expectation.



(a) Expected Public-Sector Wages



(b) Expected Private-Sector Wages

Figure 2.7: Expected Public and Private-Sector Wages by Sector Preference

Note: Bars represent 95% confidence intervals. Preference for sector comes from answer to question: "In case you decide to look for a job after graduating, would you prefer a job in the public or the private sector? (1) Public sector; (2) Private sector; (3) Indifferent". Figure (A): Mean expected public-sector wages after 10 years of experience. Figure (B): Mean expected private-sector wages after 10 years experience.

2.9 Conclusion

This paper constructs a new dataset to study an old question in labor economics: are government workers paid more than their counterfactual private-sector wages? Using a fuzzy regression discontinuity design combining public-sector exam results and labor market outcomes for over 2 million individuals, I find a large and positive public-sector wage premium. In the main specification, I find a wage gap in favor of the public sector of approximately 48 percent, which is twice as large as the wage premium estimated via OLS. This difference suggests that there is negative selection into the public-sector: counterfactual private-sector wages for workers who sort into the public-sector are lower than wages of similar workers who sort into the private-sector. Results from a survey experiment eliciting expected wage distributions from a sample of Brazilian undergraduates support the selection hypothesis. The expected wage gap for the subjects that declare a preference for the public sector is positive and large, while the expected wage gap for subjects that do not prefer the public sector is small and negative.

An alternative explanation is that workers have sector-specific productivity and sort into their highest-productivity sector. That seems unlikely, however, as most of the positions offered government have similar counterparts in the private sector — police officers, firefighters and judges being possible exceptions. Moreover, for this explanation to be consistent with the RD and survey results, students would have to have a clear understanding of their levels of productivity on both sectors even before having any experience in the labor market.

Given the nature of the data, it is not possible to separate the estimates of the wage premium that are due to the public-sector job per se and the fraction that is a direct consequence of the hiring mechanism used in Brazil. In either case, the result of a negative selection effect is important. First, the end result is the same in terms of its potential economy-wide effects and on the fiscal constraints of the state. Second, other populous developing countries (e.g., India and Indonesia) use a similar system to select public-sector employees, and are thus likely to experience similar effects. Finally, the large negative selection bias reported here has important implications for interpreting previous results based on OLS regressions with controls for observable variables.

3.0 The Times They Are a-Changing: Learning in a Dynamic Adverse Selection Experiment

3.1 Introduction

In many situations of economic interest the passage of time carries with it important strategic implications. In labor markets, good workers are hired and retained at higher rates, and so prolonged unemployment can serve as a negative signal to future employers. For durable goods such as houses, long periods on the market can provide prospective buyers with a stronger bargaining position. In health-insurance markets, long prior spells without insurance may signal adverse selection for customers newly seeking a policy. These forces can also be present in simple consumer settings, where the quality of produce on offer at a farmer's market will be lower later in the afternoon than first thing in the morning, with earlier shoppers having picked through the best offerings. Such forces might even loom in the market for academic papers, where the length of time a paper has been around in working paper form can act as a negative signal, creating the need for a decreasingly ambitious submission strategy.

While situations with dynamically accruing adverse selection are commonplace, evidence for how decision makers respond to these forces is predominantly derived from behavior in static situations, such as sealed-bid common-value auctions or simple lemons problems. While the experimental literature certainly demonstrates initial failures of the Bayes-Nash equilibrium predictions, there is evidence in many settings that experience pushes subjects *toward* the equilibrium. For example, though many fall for the winner's curse early on, the learning process pushes them to bid less as they gain experience. While behavioral concepts such as Cursed Equilibrium may predict the early behavior, the standard Bayes-Nash equilibrium can still be a good long-run predictor, an "as if" outcome, without the need for subjects to understand the environment introspectively. However, similar "as ifs" in richer settings where subjects need to be responsive to non-obvious conditioning variables require greater sophistication. In order for behavior to converge towards the equilibrium agents need to have a flexible enough model of the environment to admit a conditional response to the "right" variables. A useful analogy here is to running a regression, where an omitted variable can lead to bias in all of the other estimates. Whenever subjects' models of the world are not flexible enough, long-run learned behavior can differ substantially from the Bayes-Nash equilibrium, potentially reversing policy-relevant comparative statics.

Our experimental environment provides a test-tube for examining these ideas. Specifically, in our dynamic matching environment, participants need to allow for the passing of time as a conditioning variable. In our main experimental treatment, subjects are formed into groups with each member initially assigned to an object with an independently drawn common value. Each subject discovers their assigned object's value at an exogenously determined point in time, and is then given a chance to exchange it for an unknown rematching option. In our environment, exchange results in given-up objects becoming the rematching option for subsequent movers, leading to adverse selection. So long as participants keep objects with high values and give those with low-values, adverse selection will increase over time.

Given this accruing selection, *equilibrium* behavior will be highly sensitive to the passing of time, with greater willingness to keep a low-value object at later points in the game. Moreover, introspectively reaching an equilibrium outcome requires only a relatively simple *qualitative* model of other players: that they keep good items and exchange bad ones according to a cutoff. As the best-response cutoffs in our environment are relatively insensitive to the specific values of others' cutoffs, a decreasing response to time comes about for most plausible beliefs about others, in particular whether they are believed to be stationary or not. While reaching the solution introspectively requires only minimal sophistication, another possibility is that equilibrium behavior is reached only in the long run through experience. A simple requirement for convergence to equilibrium in our setting (as all information sets are reached with positive probability) is that subjects allow for a time-dependent best response. Experienced bad (good) outcomes that occur later (earlier) in the game must be considered distinctly when forming a response. In the regression analogy, given their observable outcomes from past play as the left-hand side variable, participants need to allow for time as a right-hand-side variable.

Our experimental results find that aggregate behavior *is* qualitatively in line with the equilibrium comparative static: participants are less likely to give up low-value objects as time passes. However, there are substantial deviations from the equilibrium point predictions. Looking at individual-level data there is a great deal of heterogeneity, with a large proportion of subjects failing to respond to the passage of time at all, exhibiting a stationary response even after extensive experience. Similar effects are found across five robustness treatments that vary the feedback subjects are given on the environment.

Taken together, our baseline and robustness results point to approximately two-thirds of the subjects maintaining a too-simple model of the world that omits time as a causal factor. While these subjects show no movement towards learning the qualitative conditioning variable as the session proceeds, they do show a significant learning effect, adjusting their behavior to the presence of adverse selection. Going back to the regression analogy, omitting time as a right-hand-side variable creates bias in the regression intercept, absorbing the unconditional expected effect of time. While this change in the stationary level from a fully cursed response lowers the costs of stationary subjects' mistakes, it also serves to inhibit their ability to learn the actual relationship. For example, a fully-cursed subject would observe large ex-post losses only when moving late in the game, giving up objects of below-average value, but rematching very often to the worst-case values, potentially helping them recognize time as a conditioning variable. In contrast, our adaptive but stationary subjects instead make counterfactual but unobserved errors early on (failing to give up below-average objects) and then make comparatively smaller errors later on in the game.

In our discussion section we explore the extent to which a boundedly rational equilibrium model can explain the outcomes. In particular, we show that a steady-state learning model where subjects use past play to form expectations under a misspecified assumption of stationarity can rationalize behavior. As this explanatory model is formulated post hoc, we conduct two further treatments as out-of-sample tests, explicitly testing the long-run feedback mechanism in the model. In a simplified, longer-run version of our experiment, we vary the way subjects receive role assignments. Similar to the original sessions, our first treatment assigns roles (the time of choice) randomly in each new supergame, so subjects gain experience across the conditioning variable. In our second treatment, subjects' exogenously assigned roles are fixed at the session level, so a player makes her choice at the same time period in each new supergame. Despite identical Bayes-Nash and Cursed Equilibrium predictions, subjects that learn against a stationary model will end up at very different points in the two treatments. By manipulating the assignment at the session level, we force boundedly rational agents in the *fixed-role* treatment to experience outcomes at a single point in time, and so a simple model of the world will still lead to an expectation conditioned on time. Per the behavioral model, we find stark differences in the long-run response, providing a strong out-of-sample validation, both qualitatively and quantitatively. Our paper provides evidence that boundedly rational learning models (see for example, Esponda, 2008, Esponda and Pouzo, 2016, and Spiegler 2016 for a corresponding graph-theoretic language) can be successful at organizing long-run behavior, in our case with subjects maintaining an overly simple model of a static world. Variations on our experimental setting can be fertile ground for examining related questions, and in helping to document what regularities might exist in the subjects' model-based learning.

While our experimental results definitely sound a note of caution for models that require substantial strategic sophistication to converge toward equilibrium, we should note that there is also a glass-half-full interpretation: A large minority do condition on time. This subgroup's late-session behavior is well-explained by the Bayes-Nash equilibrium, but their time-conditioning emerges very early on in the session. As more direct evidence of their understanding, written statements in a peer-advice treatment indicate a clear grasp of the theoretical mechanics. When we consider the selection effects likely to be present for professionals in finance, human resources, and actuarial disciplines, our results not only favor the equilibrium predictions, they also suggest they might be reached introspectively.

The paper is structured as follows. Section two reviews the related literature. Section three contains the experimental design and procedures, and section four presents the model and hypotheses. The main results are presented in sections five, and in section six we discuss the heterogeneity in response, outline a behavioral model and provide an out-of-sample test. Finally, section seven concludes.

3.2 Literature Review

Our study contributes to the growing theoretical [Eyster and Rabin, 2005, Jehiel, 2005, Jehiel and Samet, 2007, Jehiel and Koessler, 2008, Esponda, 2008] and experimental literatures [Esponda and Vespa, 2014, 2015] on failures to account for how others' private information will affect them in strategic settings. Experimental and empirical studies have primarily focused on three settings: auctions (see Kagel and Levin, 2002 for a survey), voting, and informed sellers. One well-documented case is the *winner's curse*, the systematic overbidding found in common-value auctions. A leading theoretical explanation for this effect is that bidders fail to infer decision-relevant information on the value for the item they are bidding on, conditional on a relevant hypothetical: their bid winning the auction. Modeling this, Eyster and Rabin [2005] allow for subjects to best respond to others' *expected* action, failing to incorporate (or imperfectly incorporating, if partially cursed) how others' actions are correlated with their private information.

A number of experimental studies have focused on determining the extent to which the winner's curse can be explained by this conjecture. For example, Charness and Levin [2009] have participants engage as buyers in an individual version of the informed-seller problem. Ivanov et al. [2010] have players bid in a common-value second-price auction where the value of the object is the highest signal in the group (the maximal game), thereby controlling for beliefs about their opponents' private information. Both studies continue to find deviations from the standard equilibrium prediction, suggesting that incorrect beliefs about other players' information are one source, but not the only one, of the failure to best respond. For cursed behavior in voting, Esponda and Vespa [2014] find that most participants in a simple voting decision problem with minimal computational demands are unable to think hypothetically. That is, they do not condition their votes on the event that their vote is pivotal (and the subsequent information on the common state). Moreover, a smaller fraction of subjects is also unable to infer the other (computerized) voters' information from their actual votes. Similarly, Esponda and Vespa [2015] found that most participants were not able to correctly account for sample-selection driven by other players' private information.[1]

[1]See Enke [2017], who also examines belief updating in selected samples, and Jin et al. [2015] who

Our experimental setup expands this literature by offering a novel setting that can be easily modified to explore various bounded rational models of learning to detect regularities in people's misspecified perceptions of the strategic setting.

Thus far, the experimental literature has focused on the importance of sequential rather than simultaneous play in reaching closer to equilibrium behavior in these strategic settings. For example, a significant share of participants who received explicit feedback about the computerized players' choices in the sequential treatment of Esponda and Vespa [2014] were able to correctly extract information from those observed choices. Similarly, players are more likely to adjust their thresholds to account for the selection problem if they were actually pivotal in the previous round [Esponda and Vespa, 2015]. A number of experiments on sealed-bid vs. clock auctions have found closer to equilibrium bidding behavior when bidders are able to observe the decisions of other bidders (Levin et al., 1996; Kagel, 1995). Carrillo and Palfrey [2009] find that second movers in the sequential version of their two-sided adverse selection setup behave more in line with equilibrium predictions than the first movers or players in the simultaneous version. Ngangoué and Weizsäcker [2017] is another recent example where traders neglect the information contained in the hypothetical value of the price in a simultaneous market, but react to realized prices in line with standard theory in a sequential market. However, while the literature has identified sequentiality as the key to subjects understanding the equilibrium thinking, our paper suggests that sequentiality alone might not be enough, and that it is crucial that agents have a flexible-enough model of the world.

Our study also speaks to the substantial theoretical literature interested in dynamic adverse selection environments (Hendel et al., 2005; Daley and Green, 2012; Gershkov and Perry, 2012; Chang, 2014; Guerrieri and Shimer, 2014; and Fuchs and Skrzypacz, 2015). One focus has been on asset markets where sellers have private information about the quality of the asset [Chang, 2014, Guerrieri and Shimer, 2014]. Similarly, the current and past owners of an object in our setup could know the value of the object, while those who have never

examine the response to empty messages in a disclosure environment. In both studies, sample selection creates a tension between the naive expectation and the correct one. In our setting we instead examine the within-subject response to an observed conditioning variable, namely *time* and focus more on how subjects learn about the selection forces.

held the object do not. Although our players only make a binary choice on whether to keep the object or trade it for another in the early rounds of the experiment, they state a cutoff value for trading the object in later rounds, much like the price setting done by sellers and buyers in the asset markets. Our experimental results suggest that these models should take seriously behavioral agents with misspecified models of the dynamic adverse selection environment.

3.3 Design

We conducted 30 experimental sessions with a total of 516 undergraduate subjects. The experiments were all computer-based and took place at the Pittsburgh Experimental Economics Laboratory (PEEL). Sessions lasted approximately 90 minutes and payment averaged $25.64, including a $6 participation fee. We have eight different treatments: two main treatments, five robustness treatments, and two additional treatments for the out-of-sample test of a behavioral model. To approximate the sequential structure of our studies the next two sections focus on describing the two main treatments: a *Selection* (S) treatment, which induces dynamic adverse selection; and a *No Selection* (NS) control treatment that removes adverse selection and has a stationary best response.

Our *S* and *NS* sessions both consist of 21 repetitions of the main supergame, broken up into: part (i) (supergames 1–5), which introduces subjects to the environment; part (ii) (supergames 6–20) and part (iii) (supergame 21), which add strategy methods; and part (iv), which elicits information on risk preferences and strategic thinking. Before each part, instructions were read aloud to the subjects, alongside handouts and an overhead presentation.[2] The environment in both treatments has a similar sequential structure, with one key difference: in *S* supergames three randomly chosen subjects are matched together into a group to play a game; in *NS* supergames an individual subject makes choices in an isolated decision problem. We next describe the *S*-treatment environment in more detail

[2]Detailed instructions, presentation slides, and screenshots of the experimental interface are available on the Appendix to Chapter 3.

before coming back to describe the *NS*-treatment.

3.3.1 Selection Treatment

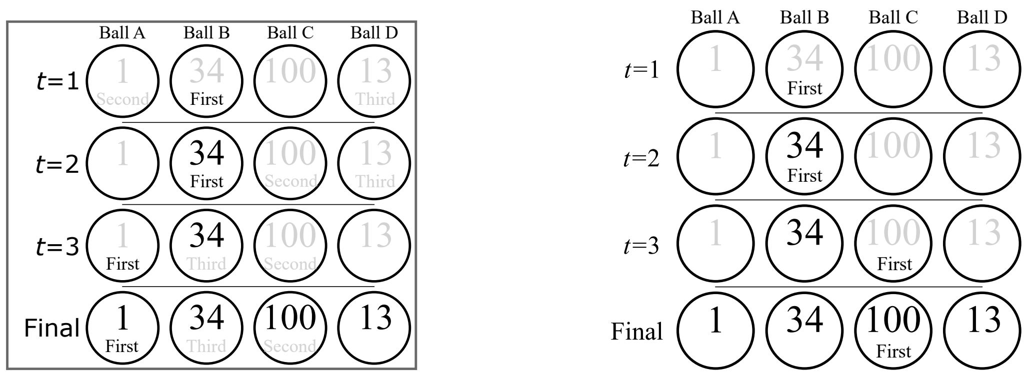
The primary uncertainty in each of our supergames is generated by drawing four numbered balls, labeled as *Balls A–D*. Each ball is assigned a random value θ through an independent draw over the integers 1–100 (with proportionate monetary values from \$0.10 to \$10.00) according to a distribution F, which has an expected value of 50.5.[3] A group of three players are randomly assigned a mover position, which we refer to as *first*, *second* and *third* mover. Each group member takes one of the four balls in turn, randomly and without replacement. As the three players each hold a initial ball, a single ball remains unheld. This unheld ball is the initial rematching population in our game.

An example matching is illustrated in Figure 3.1, where the first line shows an example initial matching. In the illustrated example the first mover is matched to *Ball B*, the second mover to *Ball A* and the third mover to *Ball D*, or $\langle 1B, 2A, 3D\rangle$ for short. In this example the leftover unheld ball is *Ball C*.

Though players know which of the four balls they have been assigned at the start of the supergame, they do not start out knowing the assigned ball's value, nor the balls (or values) held by other group members. In each round, the three players flip fair coins. If it lands heads they learn their held ball's value, and if the coin lands tails they do not learn the value and must wait to flip again. It is possible for more than one player to move in the same period, for instance, if both first and second movers flip a head in their first try. Lastly, if a player has not seen their held ball's value in rounds one or two (flipping tails in both) then the value is revealed to them in round three with certainty.[4] Each subject makes only one decision per supergame. In the round when they see their ball's value, the player makes one, and only one, payoff relevant decision:

[3]The distribution used in our experiments is a discrete uniform with additional point masses at the two extreme points. Precisely, the probability mass function puts a $\frac{51}{200}$ mass on the two values 1 and 100 and a $\frac{1}{200}$ weight on each of the integers 2–99. This distribution was chosen to make the selection problem more salient, and to generate sharper predictions for the Bayes-Nash equilibrium.

[4]We later report on treatments that eliminate the added uncertainty of the coin flips, with similar results. We introduced the coin flips in the initial experiments in order to collect within-supergame data for each participant and to make the passage of *time* more salient.



(a) *Selection (3 players)* (b) *No Selection (1 player)*

Figure 3.1: Example Supergames

Either Keep the currently held known-value ball as the final supergame outcome.

Or Take the unknown rematching ball; the currently held ball becomes the rematching ball.

To make clear the process and intuition of the game, consider the example illustrated in Figure 3.1. The figure here takes the point of view of the first mover, where figure elements in black represent information that is known to the first mover at each point in time, while elements in gray represent unknowns. In the example, though the first mover knows she is holding *Ball B* in the first round ($t = 1$), its value remains unknown to her as she fails the coin flip. The first mover does **not** know which balls the other two players are initially holding, nor their coin flip outcomes, nor their decisions. She only knows that they are present and that their decisions are potentially affecting the rematching ball.

In the illustrated example, the initial matching is $\langle 1B, 2A, 3D\rangle$ and *Ball C* is initially unheld. However, unbeknown to the first mover, the second mover flips a head and sees his held ball's value is one and decides to switch, while the third mover flips tails and does not learn her value. The interim matching is therefore $\langle 1B, 2C, 3D\rangle$ where the rematching ball is now the released *Ball A.* In the second round, the first mover flips a head, and sees that her held ball's value is 34. She decides to release this ball and rematches to the currently unheld *Ball A*, and the matching becomes $\langle 1A, 2C, 3D\rangle$. After her round-two decision (and again, unknown to the first mover) the second mover does not act as he has already made a decision, while the third mover flips a head and decides to give up her 13-ball, rematching to the *Ball B* that was just given up by the first mover, shifting the match to $\langle 1A, 2C, 3B\rangle$. By round three, all three participants have made a decision, and so no other choices are made, and thus the final matching is $\langle 1A, 2C, 3B\rangle$. At the end of the supergame all four balls' values are made common knowledge—though which balls other players are assigned to is not—and the first mover learns that the ball she rematched to has a value of one.

Supergames one to five exactly mirror the procedure above: Subjects make a binary decision to *keep* or *switch* only in the round where their ball value is revealed. The second part of each session then adds a partial strategy method. Specifically, in supergames 6–20 participants are asked to provide a cutoff in each round, indicating the lowest value for which they would keep their held ball contingent on seeing its value that round. If they receive information, the decision to keep or switch is resolved according to the stated cutoff; if they do not, they must wait until the next round, when they provide another cutoff. As in supergames 1–5, subjects only implemented decision is in the round where the ball's value is revealed, and no other decisions—or cutoffs—are recorded afterwards. Finally, in part (iii) we use a complete strategy method in which subjects are not informed about whether or not information was received in each round, and we collect their minimum-acceptable cutoff values in all three rounds of supergame 21 with certainty.[5,6] Strategic feedback on the other participants is purposefully limited in our baseline S game.[7] At the end of each of our S-supergames, each group member sees the values of the four drawn balls, as well as the particular ball he/she is holding at the end and (if relevant) the identity and value of the ball they were initially matched to. Participants *do not see* strategic feedback. That is, they observe neither the identity of the balls held by the other two group members at the end of the supergame, nor the balls others were initially holding, nor their choices.

Subjects' final payments for the session are the sum of: a \$6 show-up fee; \$0.10 times

[5] In expectation one-quarter of subject data in supergames 6–20 will have data from all three round cutoffs, one quarter with cutoffs from rounds one and two only, and one half of the data only has an elicited first-round cutoff.

[6] In part (iv) at the end of each session we collect survey information, and incentivize the following elicitations: (a) risk preferences (using the dynamically optimized sequential experimentation method from Wang et al. 2010); (b) a three-question Cognitive Reflection Test [Frederick, 2005]; and (c) a version of the Monty Hall problem. One participant per session was selected for payment in the part (iv) elicitations.

[7] We examine the effects of alternative feedback in Section 3.5.1.

the value of their final held ball ($0.10 to $10.00) from two randomly selected supergames from 1–20; and $0.10 times the value of their final held ball in supergame 21. Excluding the part (iv) payments the experiment therefore has a minimum possible payment of $6.30 and a maximum of $36.00.

3.3.2 No Selection Treatment

Our *No Selection* (NS) treatment is designed to have the same structure as the *S*-treatment game, except that we turn off the dynamic adverse selection. This is achieved by making a single change to the environment: each group has just one member. As such, each supergame is a decision problem with a single participant in the role of first mover. As there are four balls, and only one of them is held by the agent, there are three unheld balls. In whichever round the first-mover sees their held ball's value, if they decide to switch their ball, they receive a random selection from the three unheld balls. Our *NS* sessions therefore replicate the incentives and timing from the *S* sessions, but without the other group members. We illustrate a parallel example supergame for the NS environment in Figure 3.1(B).

3.4 Model and Hypotheses

The games described above are dynamic assignment problems over a finite set of common-value objects. The objects (the long-side) are initially assigned randomly to the short-side of the market (the game's participants). Private information on the held object's value arrives randomly over time, according to an exogenous process (in the experiment, the coin flips).

With a single decision maker, the rematching pool is never affected by other participants' decisions. As such, the risk-neutral prediction in our *NS* treatment is that subjects are stationary and use a minimal acceptable cutoff of 51 for retaining a ball. The cutoff rule gives up balls valued 50 or below (beneath the expected value of 50.5) and keeps balls valued 51 or higher.

Though risk-aversion or risk-lovingness might lead to alternative cutoffs, the passing of time conveys no information on the expected rematching value, and decision makers are predicted to be stationary across supergame rounds.

Hypothesis 1 (NS-Treatment)**.** *Subjects use stationary decision-making cutoffs in the* NS *treatment*

In contrast to the control, when there are multiple players the arrival of private information leads to adverse selection on the rematching pool. Whenever other players give up objects with (privately) observed low values, and keep objects with high values, the rematching pool will become selected. As private information arrives slowly, adverse selection accrues over time. In early periods, it is less likely that others have received information, so the rematching pool is less likely to be selected. In later periods, it is more likely that others have received information, leading to an increasing likelihood that the rematching pool is adversely selected.

Because the environment is sequential and involves each player making a single decision, the equilibrium predictions can be solved inductively, where best-response calculations are entirely backward looking. This is in contrast to many other situations that examine "cursed" behavior over hypothetical future events. For example, in common-value auctions, optimal decision-making requires the bidder to act as if concentrating solely on the hypothetical event that their bid wins the auction, inferring information contained in this event on the object's value. Similarly, in common-value voting, the voter has to focus on the hypothetical event that her vote is pivotal. In our environment, the optimal response is conditioned on time, where the hypothetical thinking relates to how other participants have acted in *previous* periods.

From the point of view a player making their decision at time t there are two distinct random variables: the initially assigned objects to each player θ_i^0, each distributed identically to an initial draw $\theta \sim F$; and the rematching object, θ_t^R, with a distribution that varies over time. Once the held object's value becomes known, the optimal risk-neutral response is to give up held objects if their value is lower than the expected value of rematching, and to keep higher values.[8] The rematching random variable θ_t^R and the policy cutoff $\mu_t^\star$ can be calculated inductively from the first-mover seeing her object's value in the first round ($t = 1$).[9] For the base case the rematching value θ_1^R in period one is just an *iid* draw from the initial-value distribution F, as no other participant has had a chance to exchange their object yet. The policy for a risk-neutral first-round, first-mover is summarized by a cutoff equal to the expected-value of a single draw from F. The player moving at $t = 1$ therefore keeps values of 51 and higher, and discards values of 50 and lower, similar to the *NS* rule.

For the inductive step we define the event that the player moving at time t sees their value as $\mathcal{I}_t$, and the joint event that they both see their value *and* choose to switch as $\mathcal{S}_t$ (with complement $\mathcal{S}_t^c$). Given the rematching random variable θ_t^R in period t, and the previous player's policy cutoff $\mu_t^\star$, the rematching value θ_{t+1}^R in period $t + 1$ is defined by:[10]

$$\theta_{t+1}^R \left| \mathcal{I}_{t+1} \right. = \Pr\left\{\mathcal{S}_t; \mu_t^\star \left| \mathcal{I}_{t+1} \right.\right\} \cdot \left(\theta \left| \theta < \mu_t^\star \right.\right) + \Pr\left\{\mathcal{S}_t^C; \mu_t^\star \left| \mathcal{I}_{t+1} \right.\right\} \cdot \left(\theta_t^R \left| \mathcal{I}_{t+1}, \mathcal{S}_t^C \right.\right). \qquad (3.1)$$

The optimal policy cutoff for the player making a decision at $t + 1$ is then given by the expectation of rematching $\mu_{t+1}^\star = \mathbb{E}\left(\theta_{t+1}^R \left| \mathcal{I}_{t+1} \right.\right)$. Given the induction in (1) it is clear the solution to the model is unique where the best response is entirely backward looking on what previous players might have done.

The risk-neutral Bayes-Nash Equilibrium predictions for the S-treatment vary from a predicted cutoff of 51 for the first mover in the first round, to a cutoff of 23 for the third mover in the third round.[11] This represents a substantial response to adverse selection by the end of the supergame, reducing the expected value of rematching by almost half. To put this in context, if the other two agents were fully informed on the other three balls' values and perfectly sorted so the remaining unheld ball was the worst of the three, its expected

[8] In our experiments the action set is discrete as the ball values are in $\Theta = \{1, \ldots, 100\}$, and so the cutoff can be summarized instead by $\min\{x \in \Theta : x \geq \mu_t^\star\}$, the minimal acceptable ball value.

[9] For the theory, instead of indexing time by the round number, we do it by round-mover. So the first mover in round 1 is $t = 1$; the second mover in round 1 is $t = 2$; the third mover in round 1 is $t = 3$; the first mover in round 2 is $t = 4$; *etc.*

[10] For example, given the base case the next step in the induction has the second-mover see her value and infer that $\Pr\left\{\mathcal{S}_1; \mu_1^\star \left| \mathcal{I}_2 \right.\right\} = \Pr\left\{\mathcal{S}_1\right\} = \Pr\left\{\mathcal{I}_1\right\} \cdot F\left(\mu_1^\star\right) = \frac{1}{4}$, given a one-half probability the first mover observes their value, and a one-half probability that their observed ball's value is lower than the first-round cutoff. The effective CDF for the rematching pool in period two is therefore $\frac{1}{4} \cdot F\left(x \left| \theta < 50.5 \right.\right) + \frac{3}{4} \cdot F(x)$, with expected value $\mu_2^\star$.

[11] The risk-neutral PBE cutoffs for rounds 1, 2, and 3 are, respectively: $(51, 35, 28)$ for the first-mover; $(42, 31, 25)$ for the second-mover; and $(35, 28, 23)$ for the third-mover.

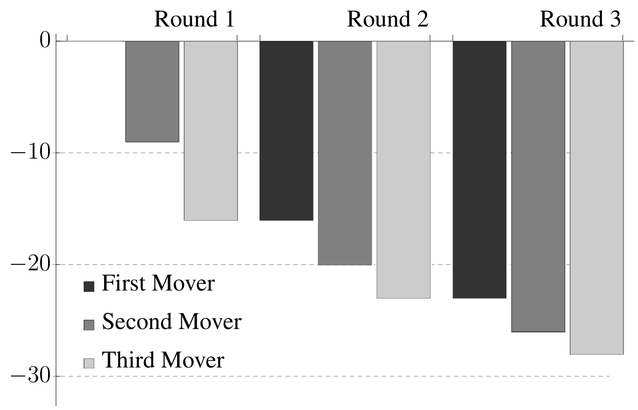


Figure 3.2: Predicted Adverse-Selection Accruing Over Supergame

value would be $\mu_{(3)} = 16.4$. That is, by the end of the last round over 75 percent of the adverse selection possible under full information and perfect sorting has occurred. Figure 3.2 expresses the risk-neutral Bayes-Nash Equilibrium predictions as the intensity of the adverse selection, graphing the difference $(\mu_t^* - \bar{\mu})$, where $\bar{\mu}$ is the expected value of the original distribution F. Within each round the cutoffs are decreasing as the different roles take turns to move, but across rounds there is a flatness from the third mover in round one to the first mover in round two, and from the third mover in round two to the first mover in round three.[12] Importantly, within each role the PBE predictions indicate strictly decreasing cutoffs, reflecting the increased adverse selection as the game unfolds.

While the equilibrium cutoffs are unique under risk neutrality, the decreasing pattern holds in equilibrium for both risk-loving and risk-averse preferences. Moreover, decreasing cutoffs will be predicted even without sophisticated equilibrium beliefs on others' behavior. For example, a simple belief that other participants use a stationary (non-boundary) cutoff rule that gives up low-valued objects and keeps high ones yields best-response cutoffs with quantitatively similar predictions to the equilibrium (see Figure A1.1 in the Appendix).

Using a cutoff μ' different from the equilibrium one has two largely offsetting effects

[12]The reason for the non-decreasing parts is the conditioning in equation (1). For example, the first mover who sees their value in round two (the fourth mover, so the event $\mathcal{I}_4$ in the induction) knows that they did not switch in round one. So in the language of the induction $\Pr\{\mathcal{S}_1 | \mathcal{I}_4\} = 0$, as $\mathcal{S}_1 \subset \mathcal{I}_1$ and $\Pr\{\mathcal{I}_1 \cap \mathcal{I}_4\} = 0$ in our information structure—the same reasoning applies to the first mover in round three. Note, if players were different for each decision, the cutoffs would be strictly decreasing, as conditioning on $\mathcal{I}_{t+1}$ would be uninformative to prior periods.

in the inductive calculation given in (1). On the one hand increasing the cutoff increases the likelihood of selection, $\Pr\{\mathcal{S}_t;\mu'\}$, as there are more values for which the object is exchanged. On the other, it decreases how bad the selection is when it does occur, making the rematching distribution given a switch more favorable to succeeding agents. In our experimental parameterization, the best-response cutoffs are quantitatively similar to the equilibrium cutoffs for a large set of beliefs on others' behavior. This robustness to the cutoffs used by others is a useful feature of our dynamic environment: even with substantial deviations from equilibrium by others, the empirical best response for individuals in our sessions is essentially identical to the PBE. Instead of requiring highly accurate beliefs on the cutoffs used by others, the strategic sophistication required to deploy a decreasing cutoff is to understand that information arrives slowly over time for all players, and that others will give up low-valued objects and keep high-valued ones. Given the similarity in the different agents' tasks within the supergame, understanding this can be achieved by projecting one's own experiences and behavior onto others.

While the levels in Figure 3.2 are calculated using the risk neutral PBE, the following hypotheses are independent of the subjects' risk aversion and are robust to subjects' beliefs on others' strategic behavior. As the majority of our results elicit subjects' precise cutoff rules, we specify our hypotheses over such cutoffs.[13]

Hypothesis 2 (S-Treatment)**.** *Subjects use strictly decreasing cutoffs over the supergame in the* S-*Treatment*

In addition to the qualitative direction of cutoffs within treatments, we can also make comparisons across treatments, as the first-mover in the first round of our *S*-treatment faces an identical problem to the *NS*-treatment.

Hypothesis 3 (First decision equivalence)**.** *The distribution of first-round first-mover cutoffs in the* S-*treatment is identical to that in* NS

Below we outline the aggregate experimental results from these two treatments, where our focus will be on the late-session play after subjects have acquired extensive experience

[13]We examine the behavior in supergames one to five in the Appendix, where we show that subjects act as if they are using a monotone rule that keeps high value balls and gives up on low values.

with the environment. After outlining the main results, and checking their robustness, we turn to subjects learning behavior over the session in section 3.5.2, where we propose and test a behavioral model of steady-state learning that rationalizes the long-run deviations from equilibrium.

3.5 Aggregate Results

We now describe the main experimental results, comparing the behavior in the decision environments with and without adverse selection, examining the three hypotheses above. The aggregate results for the *S*- and *NS*-treatments are illustrated in Figure 3.3, where the figure presents all data from subjects in the first-mover role where a cutoff is elicited. The focus on first movers provides the cleanest comparison across treatments because (i) the PBE prediction is identical for the first-mover in the first round, and (ii) the changes in the optimal cutoffs across rounds are largest for first-movers.[14] The figure indicates the first-mover subjects' responses relative to the expected cutoff without selection ($\mu_1^\star = 51$). While the equilibrium theory (the circles) and empirical best response (the triangles) predict no adverse selection in *NS*, the prediction in the *S* treatment is for selection to accrue across the three rounds (the gray circles), with much of the selection accruing by the round two decision.

Three patterns emerge from Figure 3.3: (i) aggregate subject behavior does respond to the passage of time in *S* supergames, but the adjustment to the adverse selection falls short of the equilibrium predictions; (ii) behavior is qualitatively different across treatment and control; and (iii) while aggregate behavior in the *NS* control is statistically indistinguishable from the risk-neutral theory, behavior in the *S* treatment is significantly different.

Table 3.1 provides random-effect regression results to complement the figure. The table reports estimated (absolute) cutoff for first movers across rounds one to three, where we

[14]Results and conclusions are statistically and numerically similar with a focus on all rounds and mover roles. In the Appendix we provide evidence from the first five supergames where we did not explicitly elicit cutoffs; subjects' behavior in these rounds with the binary keep/switch action is consistent with the use of a cutoff.

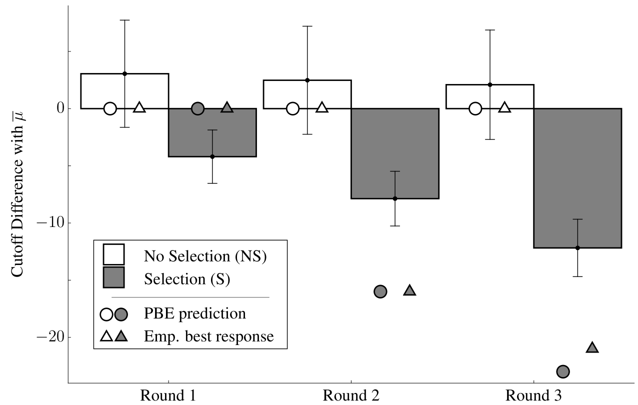


Figure 3.3: First-Mover Cutoffs (Supergame 6–21)

Note: Bars depict 95 percent confidence intervals from a random-effects estimation across all cutoffs in supergames 6–21. Empirical best responses calculated using cutoff distributions.

separately estimate first-mover behavior in supergames 11 to 20 and in the full-strategy method supergame 21.[15] Aggregate estimates are produced by regressing the chosen first-mover cutoff μ^j_{ist} (subject i, supergame s, round t, and session type $j \in \{\text{NoSel}, \text{Sel}\}$) on a set of mutually exclusive treatment-round dummies. The estimated aggregate cutoff $\hat{\mu}^j_t$ for session type j and supergame round t, allows us to make statistical inference over the equilibrium hypotheses.

[15]We focus here on results in the latter half of the session. Results for supergames 6 to 20 are in Appendix Table A3.1; results for subjects in the second- and third-mover roles are in Table A3.2. Qualitative results are similar to those reported in Table 3.1.

Table 3.1: Average First-Mover Cutoff (Relative to 51) by Round in *No Selection* and *Selection*

Treatment	**Cutoff for**	**Theory**	**Supergame 11 to 20**				**Supergame 21**			
			N	**Estimate**	***p*-values**		N	**Estimate**	***p*-values**	
		$\mu^{\star}$		$\hat{\mu}$	$\hat{\mu}=\hat{\mu}_1^{NS}$	$\hat{\mu}=\mu_t^{\star j}$		$\hat{\mu}$	$\mu=\hat{\mu}_1^{NS}$	$\mu=\mu_t^{\star j}$
No Selection (NS)	Rd. 1, $\hat{\mu}_1^{NS}$	[0]	110	+3.8	–	0.119	33	+1.8	–	0.479
				(2.4)				(2.5)		
	Rd. 2, $\hat{\mu}_2^{NS}$	[0]	49	+3.1	0.213	0.216	33	+1.9	0.907	0.449
				(2.5)				(2.5)		
	Rd. 3, $\hat{\mu}_3^{NS}$	[0]	27	+2.9	0.268	0.244	33	+2.3	0.620	0.360
				(2.5)				(2.5)		
	Joint Tests:			0.335‡		0.238§		0.875‡		0.817§
Selection (S)	Rd. 1, $\hat{\mu}_1^{S}$	[0]	460	-4.4	**0.003**	**0.000**	66	-7.5	**0.018**	**0.013**
				(1.23)				(3.04)		
	Rd. 2, $\hat{\mu}_2^{S}$	[-16]	212	-8.0	**0.000**	**0.000**	66	-11.1	**0.001**	0.107
				(1.28)				(3.0)		
	R. 3, $\hat{\mu}_3^{S}$	[-23]	113	-11.9	**0.000**	**0.000**	66	-14.7	**0.000**	**0.006**
				(1.4)				(3.0)		
	Joint Test:			**0.000**‡		**0.000**§		**0.000**‡		**0.000**§

Note: Figures derived from a single random-effects least-squares regression for the relative cutoff (choice-51) against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 170/137/33 Total/*Selection*/*No Selection* first-mover subjects across supergames 11–20, and 55/22/33 in supergame 21. *Selection* treatment exclude subjects in the second- and third-mover roles (these figures given in the Appendix). †−Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t) or the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Hypothesis 1 is a basic check for the control environment: given the stationary *NS* environment, are the aggregate cutoffs in this treatment stationary across the supergame? Inspecting the *NS* coefficients in Table 3.1, we verify that the first-round cutoffs are 54.8 (so +3.8 relative to the risk-neutral cutoff of 51). This decreases slightly over the course of each supergame, to +2–3 in rounds two and three. Examining each coefficient in turn we test whether the average cutoffs used in each treatment-round are equal to the coefficients used in round one, reporting the p-values in the $H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ column. Individually, neither the second- nor third-round's *NS* coefficients are significantly different from the first round. Examining Hypothesis 1 directly with a Wald test for the same cutoff for all three rounds of the control ($H_0 : \hat{\mu}_1^{NS} = \hat{\mu}_2^{NS} = \hat{\mu}_3^{NS}$) we fail to reject for both supergames 11–20 ($p = 0.355$) and for supergame 21 ($p = 0.875$).

Beyond just stationarity in the *NS* cutoffs, we also fail to reject the stronger hypothesis that aggregate behavior in *NS* is both stationary and equal to the risk-neutral prediction. Examining each coefficient separately, we fail to reject the risk-neutral predictions for all *NS* round coefficients (the $H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$ column), where we also fail to reject the risk-neutral PBE prediction jointly with a Wald test ($p = 0.238$).

Result 1 (Control Stationarity)**.** *We cannot reject that average behavior in the* NS *control sessions is stationary, nor can we reject the risk-neutral PBE prediction.*

Given that aggregate behavior in our control is well-behaved, we turn to an examination of aggregate behavior in the environment with adverse selection. The bottom half of Table 3.1 provides the average first-mover S-treatment cutoffs across the three rounds, $\hat{\mu}_1^S$, $\hat{\mu}_2^S$ and $\hat{\mu}_3^S$, again breaking the estimates up into those obtained in supergames 11–20 and supergame 21. Our coarsest prediction for the *Selection* treatment is that the cutoffs decrease, indicating that subjects respond to the adverse selection accruing over time (Hypothesis 2). The hypothesis is tested by examining a Wald test for stationary cutoffs, $H_0 : \hat{\mu}_1^S = \hat{\mu}_2^S = \hat{\mu}_3^S$. Unlike the control where we fail to reject stationarity, we strongly reject it in the S-treatment ($p < 0.001$) in favor of the PBE prediction of strictly decreasing cutoffs.

Though qualitative behavior is in line with the theory, the aggregate levels in S are far from the PBE predictions. As illustrated in Figure 3.3, subjects' behavior does not fully internalize the predicted degree of adverse selection. For supergames 11 to 20, the relative level is just over half the predicted magnitude. For supergame 21 on its own, the relative magnitude is closer to the prediction, representing 64 percent of the predicted adverse selection effect of -23. However, the attenuated response relative to theory becomes more pronounced when we consider that subjects start out with lower cutoffs in the very first round. While the behavioral shift across the three rounds is significantly less than zero ($\hat{\mu}_3^S - \hat{\mu}_1^S = -7.65$ in supergame 21, $p < 0.001$), the magnitude of the observed change in cutoffs is a third of the theoretical prediction.

The relative drops in willingness to rematch across the supergame are therefore less pronounced than the equilibrium predictions, but the different behavior in the first-round of the S-treatment also jumps out as an anomaly. Despite an equivalent decision for first movers in the very first round, the provided cutoffs in the S supergames ($\hat{\mu}_1^S$) are significantly lower than both the NS cutoffs ($p = 0.002$) and the risk-neutral prediction ($p = 0.004$). Moreover, this effect becomes more pronounced if we focus just on play at the end of the session in Supergame 21.

Result 2 (Treatment Dynamics)**.** *We reject that aggregate behavior in the* S *treatment is stationary, as the cutoffs have a significant and strictly decreasing trend. However, the dynamic reaction is significantly different from the theoretical prediction.*

Result 3 (First Round Non-Equivalence)**.** *Average first-round cutoffs in the* S *treatment are significantly lower than both the* NS *results and the risk-neutral prediction.*

In Section 3.5.2 we show that these two aggregate patterns from the S-treatment (a negative but shallow slope, with a lower intercept) are a product of individual heterogeneity. Two behavioral types emerge: (a) Sophisticated subjects who change their cutoffs across supergames, starting close to the risk-neutral value in the first round; and (b) Coarse-reasoning subjects that use a constant response across the supergame, but where the level of this response responds downward due to (unconditional) experienced selection. When mixed, the aggregate behavior is shifted downward with an attenuated response across time. Before analyzing the results from individual heterogeneity, we briefly outline results from four robustness treatments, varying the strategic feedback that subjects receive.

3.5.1 Summary of Robustness Treatments

Our robustness treatments manipulate the information subjects receive in the dynamic adverse selection environment. Details of these treatments and the results are relegated to the Online Appendix for interested readers, as the main findings replicate the above, and seem to be driven by the same learning behavior discussed in the next section. Our intention here is to provide the reader with a concise summary of the treatments and the qualitative results.

Robustness Treatment 1 (S-Across)**.** *Additional strategic feedback across the* S-*treatment supergames.*

Here we replicate the S-treatment, but the subjects are now completely informed on all players' actions at the end of each supergame. Looking back to Figure 3.1(A) in the design, whereas the S-treatment only informed subjects on their own choices (the elements in black), in the *S-Across*-treatment subjects are informed of all elements in the figure once the supergame has ended. The treatment results mirror those in the S-treatment.

Robustness Treatment 2 (S-Within)**.** *Additional strategic feedback within the* S-*treatment supergame.*

Where *S-Across* provided feedback at the end of each supergame, this treatment modifies the information structure within the supergame so that subjects are informed about switches along the path of play. Rather than time, the relevant conditioning variable is the observation of a switch by the other participants. In the Figure 3.1(A) example, the first mover would know that the second-mover had switched when they made their choice in round 2. We come back to this treatment to talk about individual-level results in the next section, but we find qualitatively similar effects to the S-treatment at the aggregate level. Subjects respond to the appropriate signal (here an observed move, not the passage of time), but the size of the response is attenuated.

Robustness Treatment 3 (S-Peer)**.** *Peer advice on the final choice (supergame 21) in the* S-*treatment environment.*

These sessions are identical to the *S*-treatment, except for supergame 21.[16] In the final supergame (which is paid with certainty), subjects are first matched into chat groups of three. After chatting, each member is matched with members from other chat groups for the final *S* supergame. Crucially, one of the three chat-group members is selected at random and that participant's supergame 21 outcomes determines the payoff for the entire chat group. As such, each team member has an incentive to explain the environment to others.

Even though many groups do have chat members who explain the underlying tensions in the game to the other participants,[17] the end behavior in supergame 21 is not significantly different from that observed in the *S*-treatment environment.

The results from our robustness treatments mirror the findings in our *Selection* treatment. Given the regularity across our robustness treatments, we focus on breaking down behavior at the individual level in the Selection treatment, to show that the aggregate results obscure important heterogeneity. Moreover, we show that the subject-level data from the robustness treatments do show substantial stability in the rates of subject types.

3.5.2 Subject Heterogeneity and Learning Behavior

In Section 3.5 we provide evidence that average cutoffs are significantly decreasing across the *S* supergames and are stationary in the *NS* control. To an extent this represent a victory for the theory as a qualitative prediction on aggregate behavior. However, in this section, we show that the averages mask an important heterogeneity in behavior. While a substantial number of subjects do use strictly decreasing cutoffs in the *S*-treatment, a majority use stationary cutoffs that are unresponsive to the theoretical conditioning variable. In this section, we examine individual-level results to better understand the within-subject response.

To describe individual behavior we define a simple type-scheme based on each partic-

[16] Results from this treatment were included in Table 3.1 for the columns examining Supergames 11–20 as the treatment is identical up to this point, but not for the results examining Supergame 21.

[17] Every team chat from all *S-Peer* sessions are included in Appendix E for interested readers. Example explanations: *"As the rounds go on, the chances that the ball the computer is holding has a really small value increases[...] because in previous rounds, if someone had a small value they probably switched and gave it to the computer";"So here are my thoughts: The chance of you getting a low # that someone else switched out is based on which mover you are and what round it is. Typically I go with ˜ 50 if I am mover 1 or 2 on the first round[...] Then drop down for each subsequent round. Because you get stuck with what you switch too and as time goes on that is much more likely to be a low # ".*

ipant's choices in the final supergame.[18] Specifically, we dichotomize subjects as either *Decreasing* or *Non-Decreasing*, where for the *Non-Decreasing* types we further break down the total fraction that are *Stationary*. A *Decreasing* subject is one whose final supergame cutoffs satisfy $\mu_1^i > \mu_2^i \geq \mu_3^i$, where a *Stationary* subject satisfies $\mu_1^i = \mu_2^i = \mu_3^i$. In addition to the knife-edge definitions, we create a parallel family of definitions over a variable $\epsilon > 0$, where an ϵ-*Decreasing* type satisfies $\mu_1^i \geq \mu_2^i + \epsilon$ and $\mu_2^i \geq \mu_3^i$, and an ϵ-*Stationary* type satisfies $|\mu_1^i - \mu_2^i|, |\mu_1^i - \mu_3^i| < \epsilon$.[19]

Table 3.2: Type Proportions

	Subjects	**Decreasing**		**Non-Decreasing**			
				Total		**Stationary**	
		Exact	$\epsilon = 2.5$	Exact	$\epsilon = 2.5$	Exact	$\epsilon = 2.5$
No Selection	33	3.0%	3.0%	97.0%	97.0%	57.6%	75.8%
Selection	66	42.4%	37.9%	57.6%	62.1%	36.4%	47.0%
Robustness:							
S-Across	60	28.3%	23.3%	71.7%	76.7%	45.0%	60.0%
S-Peer	72	34.7%	30.6%	65.3%	69.4%	47.2%	55.6%
Selection+Robustness	198	35.4%	30.8%	64.6%	69.2%	43.0%	54.0%

Table 3.2 provides the type composition in our *NS*-treatment and *S*-treatments (including robustness treatments). Focusing on the type definitions with $\epsilon = 2.5$, we find that all but one subject in our *NS*-treatment uses non-decreasing cutoffs, with a large (slim) majority being ϵ (exactly) stationary. In contrast, pooling across our adverse selection treatments — where we expect decreasing behavior —, we find that only a third of subjects use a decreasing cutoff profile. Comparing the fraction of decreasing types between *S* and *NS* suggests that approximately one in every three subjects is responsive to the theoretical prediction. A majority of participants, however, are invariant to the key qualitative prediction, and are instead better classified as using a stationary cutoff across the supergame.

Though our type dichotomy is based on behavior in the last supergame, the types are highly predictive of the response within earlier supergames. We next examine behavior in

[18] The final supergame represents the point where subjects have maximal experience with the task and where we ramp up the monetary incentive by an order of magnitude as this supergame is paid for sure.

[19] Figure A4.1 in the Appendix provides the type fractions as we vary ϵ from 0 to 10 to indicate the robustness across definitions.

supergames 6 to 20, where we elicit cutoffs using a partial strategy method, breaking it down by blocks of five supergames. Note that for each block of five supergames we expect to observe: (i) at least one measurement of the first-mover, first-round cutoff $\mu^i_{1,1}$ from 87 percent of subjects; and (ii) at least one measurement of the within-subject dynamic response between rounds 1 and 2 $\Delta\mu^i := \mu^i_{j,1} - \mu^i_{j,2}$ from 97 percent of subjects (for all mover roles j). For each data subsample we run random-effects regressions for each measure against dummies interacting the subject's type (based on supergame 21 behavior) and dummies for each block of five supergames. The regression results are provided in Table 3.[20]

Table 3.3: Behavior by Type Across the Session

Type	**Supergames** 6–10	**Supergames** 11–15	**Supergames** 16–20	**ΔSession**
	NS-Treatment (All)			
First-round cutoff, μ_1	+1.8 (2.4)	+3.9 (2.4)	+3.7 (2.4)	+2.0*** (0.7)
Cutoff change, $\Delta\mu$	+0.2 (0.8)	−1.3* (0.8)	−0.4 (0.8)	−0.5 (0.8)
	S-Treatment (Non-Decreasing)			
First-round cutoff, μ_1	−1.7 (1.6)	−2.6* (1.6)	−4.1*** (1.6)	−2.4*** (0.8)
Cutoff change, $\Delta\mu$	−0.3 (0.5)	−0.5 (0.5)	−0.9* (0.5)	−0.6 (0.5)
	S-Treatment (Decreasing)			
First-round cutoff, μ_1	−2.0 (2.3)	−3.8 (2.3)	−2.3 (2.3)	−0.3 (1.1)
Cutoff change, $\Delta\mu$	−6.6*** (0.7)	−7.9*** (0.7)	−8.6*** (0.7)	−2.0*** (0.7)

Note: Coefficients (and standard errors in parentheses) derived from two random-effects regressions from 1,125/1,233 observations of the first-mover first-round cutoff/cutoff-change within supergame over 159 participants. Significance at the following confidence levels: *–90 percent; **–95 percent; ***–99 percent.

[20] For the adverse-selection treatments we pool subjects from both S and *S-Across*, as these supergames are theoretically identical. We do not include data from *S-Explicit* or *S-Within* as the extensive-form games here have distinct predictions, nor do we include data from *S-Peer* as the type classifications are made *after* the chat rounds.

Table 3.3 therefore provides the treatment-type averages for the first-round first-mover cutoff, μ_1 (measured relative to 51, the risk-neutral prediction), and the participant's dynamic cutoff change within the supergame, $\Delta\mu$. Examining the first-round cutoffs (the decision point where there is no adverse selection in any treatment) we find that: (i) Subjects in the *NS* treatment use cutoffs that are consistent with slightly risk-*loving* preferences, where cutoffs increase with experience across the session.[21] (ii) In contrast to *NS*, non-decreasing subjects in the *S*-treatment use initial first-mover cutoffs that start out consistent with a risk-neutral response, but which decrease significantly over the course of the session. (iii) First-round–first-mover cutoffs for decreasing subjects are not significantly different from the risk-neutral PBE prediction in any supergame block for the *S*-treatment and do not change much over the session.

The second set of results in the table examine how subjects change their cutoffs within the supergame in reaction to time, the theoretically relevant conditioning variable. We find that: (iv) Subjects classified as non-decreasing in supergame 21 do not have a significant cutoff response to time in prior supergames.[22] This is true in environments both with and without adverse selection. (v) Finally, subjects classified as decreasing in supergame 21 show significant within-supergame reductions in the cutoff in earlier supergames, even in the 6–10 block. We also do find evidence that the magnitude of their within-supergame decrease over time increases across the session. By the last supergame block the cutoff difference is -8.6, which is not too far from the PBE prediction of 11.3.[23]

From Table 3.3 we conclude that the type classifications based on supergame 21 are useful for understanding subjects' behavior in prior supergames. While this result may not be surprising for the non-decreasing subjects, it does speak to the stability of their classification. For the decreasing types, the results indicate that these subjects understand

[21] The Δ*Session* column provides the difference between the first and third block of cutoff-eliciting supergames.

[22] Non-decreasing subjects in supergames 16-20 in the *S*-treatment do have a significant decrease in their cutoff. However, the change is quantitatively small (-0.9), and insignificant when we look at the joint hypothesis across all three blocks ($p = 0.304$).

[23] Since subjects are equally likely to be assigned first, second or third-mover roles, the expected PBE prediction takes account of the frequency of each mover role, so: $1/3(-16-9-7) = -11.33$. While we can reject the equilibrium point prediction ($p < 0.001$), the average decreasing subjects' behavior after extensive experience shows they internalize 76 percent of the predicted level.

the qualitative time component of the game early on.[24] Moreover, the chat logs from the *S-Peer* treatment suggest that approximately a quarter of the subjects use the chat to *explain* the game's dynamic adverse selection mechanic to others. The conclusion is that the minority of decreasing subjects introspectively understand the environment, adjusting their levels with experience. By the end of the sessions, these subjects average behavior is approximately 4/5 of the way along the PBE prediction.

Second, the results suggest that the non-decreasing subjects become more pessimistic over their stationary cutoff as the S-treatment session proceeds. This trend is absent in the NS-treatment; where if anything the NS-subjects move in the opposite direction, becoming more optimistic about the outside option as the session proceeds.[25] What then is going on?

Before diving into our behavioral model, we shift briefly to standard explanations for how an equilibrium "as if" might work without an introspective understanding of the game. Through repeated exposure to the environment, a subject should be able to build a direct understanding of how their resulting supergame outcome V responds to the observed signals — the exogenous time t at which they received information, $\mathcal{I}_t$, and the value of the initially held ball, θ^0 — and their chosen cutoff rule $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_t, \ldots, \mu_T)$. The expected supergame value is given by:

$$\mathbb{E}V(\boldsymbol{\mu}) = \sum_t \Pr\{\mathcal{I}_t\} \cdot \Big[\Pr\{\theta^0 < \mu_t\} \cdot \mathbb{E}\left(\theta^R|\mathcal{I}_t\right) + \Pr\{\theta^0 \geq \mu_t)\} \cdot \mathbb{E}\left(\theta^0|\theta^0 \geq \mu_t\right)\Big]. \quad (3.2)$$

Through access to enough data on past play an agent would be able to arrive at the equilibrium policy vector $\boldsymbol{\mu}^\star$ simply by forming an accurate estimate of the rematching value at each point in time, $\mathbb{E}\left(\theta^R\,|t\right)$. If this conditional expectation is learned, risk-neutral agents optimize their outcome by setting their cutoff at each point in time to $\mu_t^\star := \mathbb{E}\left(\theta^R\,|t\right)$. In the long-run, any subject that switches their assigned ball with positive probability will be able to generate an accurate estimate of $\mathbb{E}\left(\theta^R\,|t\right)$ through sample averages. Given that exchanging the lowest-possible-value ball (1) is a strictly dominant response and that the

[24]Looking just at decreasing types, 75–80 percent have a cutoff-difference in excess of -2.5 in each of the prior supergame blocks. Of the 29 subjects with data in all three blocks, 21 are consistently negative in all three.

[25]Given the between-subject identification, we should clarify that about 30 percent of the subjects classified as non-decreasing in *NS* would be expected to be decreasing types were they counterfactually placed in our *S* environment. However, we cannot separately identify these sophisticated subjects in the *NS* environment.

receipt of information $\mathcal{I}_t$ is exogenous with full support, any participant that allows for the expected rematching value to be time dependent will in the steady-state act as if they introspectively understand the best response.

While conditional expectations can be quickly estimated by looking at the average of recent experiences, *understanding* that conditioning is required is harder to learn. One of the main findings from our experiments is that a majority of our participants are entirely insensitive to time, even after 21 supergame repetitions. Moreover, for those participants that do condition on time, it does not seem that they reach this understanding through trial and error. Instead, they seem to do so early on in our sessions through an introspective understanding from the game's description.[26] While sophisticated types learn the level of the effect, we do not see many 'eureka' moments where a stationary subject becomes decreasing midway through the session.

In principle the majority's failure to condition could be captured by behavioral models such as Cursed Equilibrium [Eyster and Rabin, 2005], where boundedly rational agents take entirely unconditional expectations. However, this type of model would lead to no cutoff differences for cursed players between the *S*- and *NS*-treatments. Instead, we find significant differences between the treatments for the non-decreasing types, where non-decreasing *S*-treatment participants use lower cutoffs (relative to *NS*), making their failure to condition less costly overall.

Long-run behavior for the stationary participants is therefore poorly modeled by Cursed Equilibrium, which does not allow for the cursed types to adapt their unconditional expectations. Instead we examine an alternative model with steady-state learning for the stationary types. While we reject the simplest implementation of this behavioral model (learning the unconditional value of rematching), we show that a model based on the unconditional supergame value (i.e., including outcomes both kept and rematched) performs well at explaining the results. However, recognizing that this model is post-hoc, we go on to test

[26] The best predictor for being classified as a decreasing type in the selection treatments is score on the cognitive reflection task (CRT). Marginal effects from a probit estimate suggest each correct answer on the 3-question CRT increases the likelihood of being classified as a decreasing type by 23 percent ($p = 0.000$). In contrast, while there are predictive effects on being a sophisticated decreasing type from other covariates—gender (higher probability for males), risk-aversion (higher probability with greater risk aversion) and behavior in a Monty Hall problem (higher probability for those with a sophisticated response)—these covariates are only marginally significant with p-values in the 0.069–0.079 range.

the model out-of-sample with two additional experimental treatments.

3.5.3 Maintained (but Wrong) Models of the World

Our behavioral model assumes that a fraction λ of the population is boundedly rational, and ignore time as a conditioning variable; the remaining fraction $(1-\lambda)$ correctly condition on time and best respond.

Our initial approach for modeling the stationary boundedly rational type was to have them be uncertain over the rematching outcome θ^R. While a best-responding agent will set their cutoff at time t to $\mu_t = \mathbb{E}\left(\theta^R \,|t\right)$, agents that ignore time but do respond to the experienced outcomes will in the steady state set their cutoffs to

$$\overline{\mu} = \mathbb{E}\theta^R \sum_t \Pr\left\{\mathcal{I}_t\right\} \cdot \mathbb{E}\left(\theta^R \,|t\right). \tag{3.3}$$

Given the game's structure, for any λ and boundedly rational player cutoff $\overline{\mu}$ we can solve for the equilibrium strategies of the best-responding player (a time-varying cutoff $\mu_t^\star(\lambda, \overline{\mu})$) through the rematching random variable $\theta^R(t)$. The behavioral model is then closed by finding the fixed-point solution $\overline{\mu}^\star$ to equation (3.3): the stationary cutoff that matches the unconditional value of rematching.

Under risk-neutrality the model predicts the sophisticated type's cutoffs in the selection treatment are almost-identical to the PBE prediction for all values of $\lambda \in [0, 1)$, and predicts the boundedly rational player uses a constant relative cutoff of $\overline{\mu}^\star = -15$.[27] In contrast, because the rematching model is well-specified for our NS-treatment, the prediction here is that both types use a stationary relative cutoff of zero.

A behavioral model focused on learning the rematching value reproduces a qualitative feature of our data: boundedly rational players do react to the presence of adverse selection (differences between the NS and S treatments), adjusting their levels in response. However, quantitatively the model's prediction for the boundedly rational types is too large at -15,

[27] Changes in λ do not lead to large changes in the model predictions—a function of the previously alluded to property that best response in our specific game is not sensitive to the *level* of other players' cutoffs. By way of example, when $\lambda = 1/10$ the unconditional value of rematching with a cutoff that keeps objects of 36 or better (a relative cutoff of -15) is 35.53; when $\lambda = 9/10$ the unconditional rematching expectation is 35.64. For the sophisticated types, across this same range of λ the cutoffs used are identical except for two time periods where the cutoff differs by 1.

as the average relative cutoff used by non-decreasing types in the final supergame was -5.3. Matching the qualitative feature but missing the level could plausibly be due to other assumptions in the model.

Risk neutrality is one strong assumption in the model, and one that is rejected by a separate risk elicitation at the end of our sessions. We ask subjects to make choices over a sequence of objective lottery pairs and find that most participants exhibit risk aversion. Using the median participants' measured risk preference, the prediction for the boundedly rational type in the S-treatment is a relative cutoff of -34.[28] We repeat the same exercise for the NS treatment and estimate the relative cutoff the subject should use. The median subject's behavior in our objective risk elicitation suggests a relative cutoff in the NS treatment of approximately -21. Rationalizing the quantitative selection effects observed in our experiments would instead require the large majority of subjects to be risk loving. The failure of this first model led us to conduct a number of exploratory regressions examining what aspects (if any) from the previous supergame experience and measured risk preferences predicts non-decreasing participants final cutoffs.

Pooling together non-decreasing subjects from the S- and NS-treatments,[29] we regressed the supergame 21 cutoff on a number of observables. In the appendix we lay out all of the specifications examined, however, here we focus on a single summary regression (that maximized the adjusted R-squared).

$$\hat{\mu} = -\underset{(1.5)}{4.5} \cdot S - \underset{(2.5)}{0.1} \cdot NS + \underset{(1.6)}{3.3} \cdot \hat{\sigma}\left(\overline{V}_i\right) + \underset{(1.6)}{1.7} \cdot \hat{\sigma}\left(\overline{\theta}_i^R\right) + \underset{(1.3)}{3.9} \cdot \hat{\sigma}\left(\text{Risk}_i\right). \tag{3.4}$$

The first two terms in the regression allow for separate intercepts in the S- and NS- data. The last three terms represent subject-level variation (each measured in standard-deviations from the sample average) over: (i) $\overline{V}_i$, the average experienced final outcomes in the first 10 supergames; (ii) $\overline{\theta}_i^R$, the average experienced rematching outcome in the first 10 supergames

[28]Our elicitation is designed to estimate a subject-level CRRA coefficient ρ for the utility function $u(x) = \frac{x^{1-\rho}}{1-\rho}$. Maintaining CRRA we calculate the subject's relative certainty equivalent as $\text{Risk}_i = u^{-1}\left(\mathbb{E}u(\theta^R)\right) - 51$.

[29]Our non-decreasing sample is made up of 32 subjects from NS, 41 subjects from S, and 46 subjects from S-Across which had an identical supergame. We do not include S-Peer as the final supergame choices are contaminated by the chat.

(the final outcome conditional on rematching); and (iii) Risk_i, the subject's risk preference from the elicitation task, measured as a certainty equivalent.[30]

The significant variables in the regression are: (i) the measured risk preference, with an increase in the final cutoff of 3.9 per standard deviation increase in the calculated certainty equivalent for a fresh draw ($p = 0.003$); and (ii) the average final outcome in supergames 1–10, with an increase in the final cutoffs of 3.3 per standard deviation increase ($p = 0.031$). After controlling for risk aversion and the experienced outcomes (a function of treatment) we do not find significant differences between the S- and NS-treatments ($p = 0.120$). Neither do we find significant effects from the subject's rematching experiences ($p = 0.291$).

These regression results led us to consider another explanatory behavioral model. Instead of rematching outcomes, we shifted the feedback variable for the boundedly rational subjects to be the final supergame values. Subjects' cutoffs are set to the unconditional average of their previous supergame-level outcomes, and not just the rematching outcomes. The new equation for the stationary type's cutoffs is:

$$\overline{\mu} = \mathbb{E}V \sum_t \Pr\left\{\mathcal{I}_t\right\} \cdot \left[F\left(\overline{\mu}\right) \cdot \mathbb{E}\left(\theta^R \left| t\right.\right) + \left(1 - F\left(\overline{\mu}\right)\right) \cdot \mathbb{E}\left(\theta^0 \left|\theta^0 \geq \overline{\mu}\right.\right)\right]. \tag{3.5}$$

The learning model defined by (3.5) has the boundedly rational type's expectation formed over final supergame values which mixes positively selected outcomes from retained high values to the adversely selected outcomes from rematching. For our S-treatment the risk-neutral prediction for the boundedly rational type using a final-value referent is a relative cutoff of $\overline{\mu}^{\star} = +10$. Using this rule, a participant would compare her current object to her previous supergame average outcome of just over 60, and so would keep all objects with values higher than 61, and give up all lower values. Moreover, unlike the rematching-based referent, this learning model is also misspecified for the NS treatment, where a boundedly rational type would use a cutoff of +18.

This second behavioral model is also quantitatively far from the observed behavior under risk neutrality. While the model predicts a relative cutoff of +10 for the boundedly rational

[30] We also examined experience in later supergames, but did not find significant effects. Convergence seems to happen quite quickly.

type in S, the average non-decreasing participant's supergame-21 cutoff here was -5.3.[31] However, unlike our initial model, the measured risk-aversion is now able to rationalize observed behavior across both treatments.[32]

The behavioral model where the boundedly rational agents use their final supergame outcomes as a reference to decide whether or not to give up a current value remains easy to solve when accounting for risk preferences. We illustrate the behavioral model's prediction as we shift the underlying risk preferences in Figure 3.4. In terms of preferences, we use a CRRA utility function within the model, varying the coefficient ρ. Rather than the coefficient ρ, the horizontal axis in 3.4 indicates the certainty equivalent to a single draw from our value distribution, expressed as a relative cutoff, $\mathrm{CE}(\rho)$. Risk neutrality is therefore located at 0, where greater risk aversion is indicated by increasingly negative relative cutoffs. On the vertical axis, we indicate the learning model's prediction (based on final outcomes) for the boundedly rational type under the relevant risk preference.[33]

The solid line in Figure 3.4 indicates the cutoff in our S-treatment, while the dashed line indicates the model cutoff in *NS*. Both predictions are increasing on the certainty equivalent (decreasing on risk aversion). Superimposed on the boundedly rational model predictions, we indicate elements from the data. The interquartile range and median for the non-decreasing participants' risk preferences ($[\mathrm{CE}(\hat{\rho}_{0.75}), \mathrm{CE}(\hat{\rho}_{0.25})]$, and $\mathrm{CE}(\hat{\rho}_{0.50})$, respectively) are shown as a gray band/line running vertically. Running horizontally, we illustrate the average final-supergame cutoffs for non-decreasing subjects, labeled $\hat{\mu}_{\mathrm{S}}$ and $\hat{\mu}_{\mathrm{NS}}$, for the S- and *NS*-treatments, respectively.

Using the median risk coefficient of $\hat{\rho}_{0.5} = 0.73$, which has a certainty equivalent of -22, the behavioral model squares the conflict between the elicited risk aversion and some of the seemingly risk-loving choices in the *NS*-treatment. At the median level of elicited risk aversion, the boundedly rational model predicts a steady-state cutoff of $\bar{\mu}_V = 53$ in the *NS*-treatment and $\bar{\mu}_V = 43$ in the S-treatment, comparing well to the observed supergame-21

[31] Similarly, the average relative supergame-21 cutoff in *NS* was +2.0, far from the predicted cutoff of +18 even if we were to allow for a half of the participants being sophisticated types with a zero cutoff.

[32] Recall that equation (3.4) indicates that measured risk preferences do have a strong predictive effect.

[33] For this figures we fix $\lambda = \frac{2}{3}$. However, there is again little substantive change as we vary λ. In Figure XX in the Online appendix we present the analog for Figure 4 over the S-treatment predictions at $\lambda = \frac{1}{10}$ and $\lambda = \frac{9}{10}$.

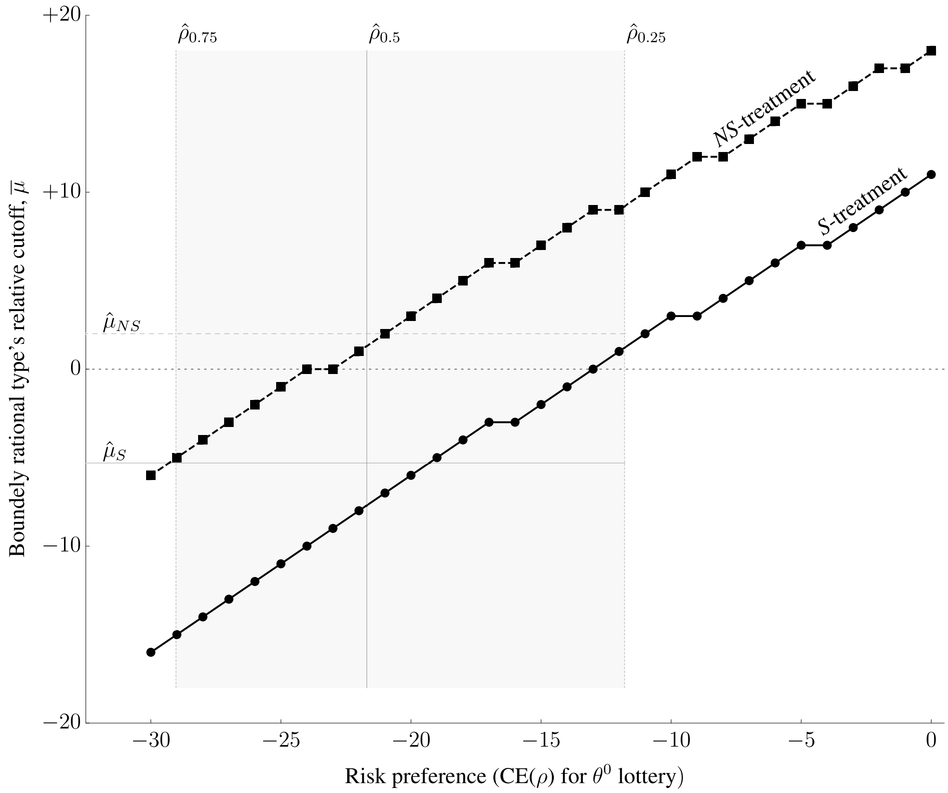


Figure 3.4: Boundedly Rational Agent Cutoffs by Risk-Aversion

Note: Shaded region shows interquartile range for elicited values for ρ among non-decreasing subjects.

cutoff for non-decreasing subjects of 53 and 45.7, respectively.[34]

The stationary behavioral model that has subjects comparing their drawn objects to a referent expectation based on final *outcomes* thus matches the aggregate results in both treatment and control, once risk aversion is considered. However, we should note that we did not consider this model prior to conducting our main experiments, and our arrival at it came about as an *ex-post* description of behavior. The model requires two distinct features to accurately predict levels: that subjects form expectations over final outcomes and not rematching outcomes (this was not obvious to us at the start of the study); and that risk aversion modifies the cutoff from the risk-neutral level (this *was* anticipated, hence the risk elicitation).

[34]For the interquartile range of risk parameters $[0.44, 0.93]$ the steady-state prediction for the boundedly rational types predicts values of $\bar{\mu}_V$ between 60 and 46 for the *NS*-treatment (subject data is 60.0 to 47.8) and between 52 and 36 for the *S*-treatment (subject data is 51 to 40).

One way of addressing the post-hoc nature of the model is to look at subject-level variation. As the model was formed from an examination of aggregate behavior, individual-level variation can help us identify the two mechanisms that underlie it. In the Appendix (Table A.5.1) we provide subject-level regressions that do just that, where we show that the non-decreasing subjects' final supergame 21 cutoffs vary significantly and in the direction of the behavioral model with: (i) their idiosyncratic experienced final outcomes in prior supergames; and (ii) their elicited risk aversion. While individual heterogeneity corroborates the two mechanisms in the model, a stronger test requires us to assess the model out of sample. Stronger still is to move out of context to a simpler strategically distinct environment.

3.5.4 Out-of-Sample Test of Behavioral Model

To test our boundedly rational model we conduct two further robustness treatments, labeled *S-Random* and *S-Fixed*. These treatments were planned and conducted *after* formulating the model. The treatments' three aims were to: (i) simplify the environment to three movers acting at fixed points;[35] (ii) double the number of supergame repetitions (made feasible through the simplification); and (iii) vary the experience that individual subjects obtain, in a way that alters the boundedly rational model but not the Perfect Bayesian (nor Cursed) equilibrium prediction. In achieving these aims the new treatments provide a strong examination not only of the steady-state learning model's predictive power, but also a validation of the key learning mechanism within the model.

The *S-Random* and *S-Fixed* environments both have the three movers act sequentially, each knowing that the previous mover observed their held-object's value for sure and made a keep/switch decision. The risk-neutral PBE prediction for the environment has the first mover use a cutoff of 51 (no adverse selection), the second mover 32 (-19), and the third mover 22 (-29). The reduction in the expected rematching outcome across the supergame is therefore quantitatively similar to our S environment,[36] but with the decrease happening more quickly.

[35] We thank Dan Levin for suggesting this modification.

[36] Adverse selection of 0 for a first mover in the first round; -20 for a second-mover in round two, and -30 for a third mover in the last round.

In the *S-Random* treatment 72 subjects played 40 supergames, where the mover roles—first, second or third—, are randomly assigned at the beginning of each new supergame. In the *S-Fixed* treatment a further 72 subjects also played 40 supergames, but the three roles were randomly assigned at the start and held fixed throughout the session (so each subject experiences the same mover role across the entire session).[37] In supergame 41 of both treatments subjects are given an equal chance of being assigned to each role, and are asked to provide cutoffs for each roles using a strategy method (similar to supergame 21 in the initial treatments).[38] Dichotomizing subjects into *Decreasing/Non-decreasing* by their supergame 41 behavior we find that 46 percent are decreasing in the random-role treatment, and 33.3 percent with fixed roles.

The out-of-sample prediction for our model relates to the non-decreasing types. Specifically, the boundedly rational model predicts an adverse selection effect of -8 (under the median risk parameter $\rho = 0.73$) for the non-decreasing participants in *S-Random*. This prediction balances out the (equally likely) chances of being a first, second or third mover. In contrast, for *S-Fixed* the subjects only experience outcomes for a single role. As such the same model predicts different steady-state behavior for the boundedly rational type: +2 for the first mover, -9 for the second mover and -14 for the third mover.

In Table 3.4 we examine the supergame 40 cutoffs of subjects in the two treatments through a joint regression, where we include dummies for all role-type and treatment interactions (and so exclude the constant).[39] We report regression results that only include the $\epsilon-$non-decreasing subjects — 87 out of 144 subjects. For each estimated coefficient we then test whether it is significantly different from the PBE prediction, and from the relevant behavioral prediction $\bar{\mu}_V$. Moreover, in the last four columns we present joint tests across the vector of three coefficients against: (i) the PBE prediction; (ii) the prediction for the boundedly rational type in the fixed-role treatment; (iii) the prediction for the boundedly rational type in the random-role treatment; and (iv) a null of cutoff stationarity across the

[37] See also Fudenberg and Vespa [2018] for an investigation of the effect of experiencing different roles in a signaling game.

[38] Note the slight asymmetry in supergame 41: in *S-Random* we ask for strategies over roles the subject experienced; in *S-Fixed* this question asks them to make decisions outside of their previous experience within the experiment.

[39] We focus here on supergame 40 as the last-round behavior as supergame 41 was used to classify participants.

Table 3.4: Supergame 40 Relative Cutoffs: Out-of-Sample Test

Treatment	**Mover**						**Joint Tests (p-values)**			
	First, $\hat{\mu}_1$		Second, $\hat{\mu}_2$		Third, $\hat{\mu}_3$		$\boldsymbol{\mu}^\star$	$\bar{\boldsymbol{\mu}}_V^{\text{Fix.}}$	$\bar{\boldsymbol{\mu}}_V^{\text{Rdm.}}$	$\boldsymbol{\mu_1 = \mu_2 = \mu_3}$
S-Fixed	-1.9		-9.5	***	-19.4	**	**.00**	.44	**.03**	**.01**
S-Random	-9.5	*	-12.2		-7.2	***	**.00**	**.03**	.84	.70

Note: Significantly different from the relevant scalar PBE prediction in $\boldsymbol{\mu}^\star = (0, -19, -29)$ at the following confidence levels: * - 90 percent; ** - 95 percent; *** - 99 percent. Significantly different from relevant scalar treatment-specific behavioral prediction in $\bar{\boldsymbol{\mu}}_V^{\text{Fix.}} = (2, -9, -14)$ or $\bar{\boldsymbol{\mu}}_V^{\text{Rdm.}} = (-8, -8, -8)$ at the following confidence levels: $^\bullet$ - 90 percent; $^{\bullet\bullet}$ - 95 percent; $^{\bullet\bullet\bullet}$ - 99 percent.

different roles.

The results validate the boundedly rational model, both quantitatively and qualitatively. For the quantitative result, and focusing on the joint tests, we reject the PBE prediction with 99 percent confidence in all tests. Similarly, we can reject *the other* treatment's behavioral prediction with at least 95 percent confidence in all cases. In contrast, for the relevant behavioral prediction $\bar{\mu}_V$, we cannot reject in either the fixed or random treatments. Moreover, we fail to reject the behavioral model in all six tests against the role-specific cutoffs.

Qualitatively, our two treatments are predicted to have distinct outcomes under the behavioral model. Though all of the non-decreasing subjects reveal a misunderstanding of how mover-roles affect expected values in their last supergame, we do observe distinct behavior across roles in the two treatments. In the fixed role treatments, the non-decreasing subjects (randomly assigned to each role) have significantly different cutoffs from one another, and we can reject stationarity across roles with 99 percent confidence. In contrast, in the treatment with random roles, subjects taking on each role in supergame 40 (again, randomly assigned) do not have distinct cutoffs from one another, and we fail to reject stationarity. As such, this represents a strong test of the model's *mechanism*. We summarize the result as follows:

Result 4 (Behavioral Model)**.** *Long-run behavior for participants in our dynamic adverse* *selection environment strongly respond to experience. However, a majority of subjects never condition on time. Their long-run choices can be modeled with a boundedly rational steady-state model, with distinct predictions from both the PBE and from boundedly rational models of initial response.*

3.5.5 Dynamic Adverse Selection in Decision Problems

Our *S*-treatments (and behavioral model) are derived for behavior in a strategic setting, where part of the problem is forming a coherent model of the other players. In contrast, our *NS* control is a decision a problem, without any ambiguity over others' play. In this section, we outline two decision-problem treatments that have dynamic adverse selection hardwired into the environment. Across these two treatments we show that (i) most participants still fail to account for selection, indicating that the failure is in being able to understand how selection affects the expected outcomes; and (ii) learning effects are attenuated, with agents' misspecified models of the environment being less affected by experience.

Our decision problem variant of the *S*-treatment makes the dynamic adverse selection in the game *explicit*. Per our other treatments, the agents make one and only one decision at an exogenously determined time — rounds one, two, or three. However, unlike in our *NS*-treatment, where choosing to rematch leads to an equal chance of matching with each of the three objects in the rematching pool, in our explicit selection treatments the rematching pool changes by round. Participants are informed that if they switch their held object in round one they will be randomly and fairly matched to one of the three rematching objects. If they switch in round two, the best rematching object is excluded, and they will be fairly rematched to the other two. Finally, if they choose to switch in round three both the best *and* second-best rematching objects are excluded, forcing the rematching object to be the worst of the three.

The expected value of rematching in rounds one to three can be calculated via three order statistics: the min, median and max of three draws, $\theta_{\text{min}} = 16.4$, $\theta_{\text{med}} = 50.5$ and $\theta_{\text{max}} = 94.6$, respectively. The expected rematching in round one is the average across all three $1/3(\theta_{\text{min}} + \theta_{\text{med}} + \theta_{\text{max}}) = 50.5$, in round two it is $1/2(\theta_{\text{min}} + \theta_{\text{med}}) = 33.5$, while in round three rematching has value $\theta_{\min} = 16.4$.

Our two treatments hold constant this structure, but differ in the exogenous likelihoods of seeing the initially held object's value in each period. Our *S-Dec* environment mirrors the standard games with a chance of observing the held ball in rounds 1–3 of $(1/2, 1/4, 1/4)$. Our *S-Dec-High* environment increases the likelihood of observing the held ball in the later periods, with an observation chance in rounds 1–3 of $(1/4, 1/4, 1/2)$.[40]

The treatment hypotheses are therefore that: (i) A large majority of participants will continue to make mistakes through the use of stationary strategies across the supergame periods; and (ii) for the stationary agents, the (constant) cutoffs will be lower on *S-Dec-High* as this treatment results in more more extensive experience with the very bad outcomes.

Table 3.5: Behavior by Non-Decreasing Types Across the Session

Outcome	**Supergames**			**ΔSession**	
	6–10	11–15	16–20		
	S-Dec (67% non-decreasing)				
First-round cutoff, μ_1	+3.8 (3.1)	+2.8 (3.1)	+3.7 (3.1)	−0.1 (0.7)	
Cutoff change, $\Delta\mu$	−0.3 (1.3)	−0.2 (1.2)	−1.3 (1.2)	−1.0 (1.1)	
	S-Dec-High (55% non-decreasing)				
First-round cutoff, μ_1	+2.0 (3.4)	+2.4 (3.4)	+0.2 (3.4)	−1.8 (0.8)	⋆⋆
Cutoff change, $\Delta\mu$	−0.1 (1.2)	−0.5 (1.3)	+0.4 (1.3)	+0.5 (0.9)	

Note: Coefficients (and standard errors in parentheses) derived from two random-effects regressions from 1,080/688 observations of the first-round cutoff/cutoff-change within supergame over 72 participants. Significance at the following confidence levels: ⋆–90 percent; ⋆⋆–95 percent; ⋆⋆⋆–99 percent.

Similar to our S-treatment, aggregate behavior is qualitative in line with the PBE predictions, but the quantitative predictions are off.[41] As before, this is driven by heterogeneity:

[40] In the instructions we move from using a coin-flip metaphor to using an urn draw without replacement.

[41] Table XX in the appendix presents full results from the *S-Dec* and *S-Dec-High* treatments across all

a majority of subjects are classified as ϵ-non-decreasing — 67% for S-Dec and 55% for S-Dec-High. Here we focus on the behavior of the non-decreasing types. Table 3.5 reports regressions results with first-round cutoffs and changes from first to second round cutoffs as outcomes, broken down by blocks of five supergames.

We observe two main results. First, it does not seem to be the case that thinking about others' beliefs is the major hurdle for achieving equilibrium. This is in line with previous work by Charness and Levin [2009] and Esponda and Vespa [2014]. The second interesting result is that there appears to be less learning going on relative to the Selection treatments. In the S-Dec treatment, we find no learning effects across the periods, and we find only a small effect on our S-Dec-High treatment where we increase the probability of late-round decisions and thus increase the likelihood of experiencing bad outcomes.

3.6 Conclusion

In this study, we use a novel experimental design that implements a common-value matching environment. The environment sets up a dynamic adverse selection problem, similar in its strategic tensions to those present in labor markets, housing markets, and mating markets, among others. The main prediction is that subjects adapt to the dynamic adverse selection by conditioning their responses on time, where the particulars of our environment make this qualitative prediction robust to risk preferences and the beliefs over the quantitative response of others. Moreover, our environment is sequential with directly experienced conditioning variables, where the strategic thinking required in is entirely backwards-looking. As such, the previous literature should make us optimistic for equilibrium in this setting. However, while a substantial minority do respond to the adverse selection in a discerning way, the majority fail to adjust their valuations over time, maintaining a stationary response.

A number of additional treatments show the results are robust to changes in the environment. Three robustness treatments increase the provision of feedback, while a fourth treatment allows for peer feedback. Taken together with a complementary result in Martínez-

rounds.

Marquina et al. 2017 in a simultaneous setting without uncertainty, our results suggest that the gains from sequentiality in previous studies were primarily driven by removing uncertainty. Further, while our sophisticated minority are able to clearly explain the adverse-selection mechanism to their peers, very few of the stationary subjects actually understand the advice, and instead stick with their stationary strategies.

While the modal stationary response place a cloud over the equilibrium predictions for our dynamic environment, we do see a silver-lining: Our sophisticated minority seem to understand the equilibrium introspectively. Their valuations are decreasing with time from the first supergames that we can observe this response; and the fact that they can explain the game's selection mechanic to others speak to their deeper understanding. When we think of professionals operating in dynamic markets—for example in finance, insurance, and labor markets—selection forces would seem to make the behavior of our sophisticated *minority* more representative.

Outside professional settings where expert decision-makers are likely to introspectively understand the strategic forces, our results point to the need for alternative theoretical models.[42] In our discussion section we consider a boundedly rational model that can help explain long-run outcomes. In this steady-state learning model, subjects learn the overall expectation of the supergame, which they use to make their (stationary) decisions. Two new treatments provide an out-of-sample test for this behavioral model's long-run prediction, where the results provide a strong validation.

In two simplified version of our original setup, we vary whether subjects experience all conditioning outcomes, or just one. Consistent with the predictions of the behavioral learning model, we find that subjects who only experience outcomes at a fixed point in time have a starkly different responses to those who experience outcomes at a random point in time, matching both the directions and levels predicted by the behavioral model. Future work using variations of our experimental setting can continue to probe for patterns and

[42]The results also point to the power of informational nudges. For example, Hanna et al. [2014] show that seaweed farmers fail to optimize production even when given all the relevant data, but pointing out the relevant conditioning variable corrects the problem. Behavioral models are useful here, as they can help us understand whether the correcting nudge produces societal benefits. For example, in a CV audit study Bellemare et al. [2018] document a failure to condition on unemployment duration when selecting candidates. Correcting this non-response might have a social cost if the planners' aims are the reduction of persistent unemployment.

regularities in subjects' maintained models, thereby sharpening the power of behavioral theoretic predictions as a complement to classical ones.

4.0 Unawareness and Risk Taking: The Role of Context

4.1 Introduction

Economic agents must often make decisions in environments they do not fully understand, either because the environment is too complicated or because the complexity is not readily apparent, leading agents to learn less than they should. As novel information is discovered, agents' perceptions of the world change, potentially leading to changes in behavior beyond what can be attributed directly to the information acquisition. For instance, the discovery of HIV made salient the possibility of a permanent and life-threatening sexually transmitted disease; the 2008 financial crisis shed a bright light on the systemic correlation in the returns to mortgage backed securities, hitherto unrecognized by economic agents ; the 20$^{\text{th}}$ century's struggle with nuclear arms and climate change illuminated humanity's ability to alter entire ecosystems.

Following such revelations, behavior is bound to change; although how, and, importantly, why, is much less clear. On the one hand, these discoveries might lead to learning useful information, in which case changes in behavior can be attributed to standard conditioning on this new information. For example, condom use increased substantially as the result of AIDS education Moran et al. [1990]. On the other hand, the very act of learning about previously inconceivable events might change agents' perceptions of their own knowledge. The discovery of a novel (and at the time, deadly) sexually transmitted disease might raise alarms about the existence of yet other undiscovered consequences of unprotected sex, resulting in an increase in safer sex practices even when participants are known to be HIV-negative. Likewise, even after the dust has settled on mortgage backed securities, investors might act more cautiously in anticipation of the next unforeseen event. In other words, exposure to unawareness can lead agents to change their preferences in ways not directly attributable to information acquisition.[1]

[1] There are many different definitions of unawareness and our upcoming discussions would meet some, but not others. We provide a formal treatment in Section 4.5, but as a working definition we take unawareness to be the inability to properly understand the space of available actions and the consequences of those actions.

Being made aware of one's own unawareness, and the resulting change in beliefs and behaviors, is a topic mostly absent from economic investigations, likely due to the difficulty in modeling and measuring unawareness. In this paper, we propose a new method to study unawareness in a controlled laboratory setting. Our experimental design creates unawareness in a natural way: subjects are given a task whose complexity is not obvious. This allows us to distinguish between exposure to complexity (simply doing the task) and exposure to unawareness (recognizing that one's understanding was incomplete). Specifically, we ask subjects to find combinations of 3 cards, from an array of 12 cards, that conform to a logical pattern. In our control treatment, subjects are told beforehand the total number of valid combinations, leaving no room for unawareness. We induce unawareness in two other treatments by (1) withdrawing any reference to the number of valid 3-card combinations or (2) providing the information only after the task is completed. The experimental design also allows us to consider subjects' introspection: their awareness of their own unawareness, which is measured by performance self-assessments.

We are interested in studying the impact of experiencing unawareness on risk preferences, which we measure using a version of the investment game Gneezy and Potters [1997]. Additionally, for each unawareness treatment we describe the risky asset either in a neutral way—using virtual coin flips—or in a framed context that invokes the complex task. We find that exposure to unawareness alone does *not* affect risk taking: there are no differences in the amount invested in the risky option across our unawareness treatments. However, investment in the risky asset is significantly lower when the risky decision is framed in connection to the complex task, and *only* for the treatments that induce unawareness. To follow on our earlier examples, we can expect post-2008 investors to become more risk averse in their investment decisions but not in choices in unrelated domains, such as speedy driving, drug use, or unprotected sex. Likewise, the discovery of HIV might well have changed sexual practices without inducing safer portfolio choices.

Our paper is related to a growing theoretical (Modica and Rustichini, 1994, Dekel et al., 1998, Karni and Vierø, 2017, Piermont, 2019a; among others) and experimental Mengel et al. [2016], Ma and Schipper [2017] literature studying the properties and consequences of decision-making under unawareness. In particular, Mengel et al. [2016] find that exposure to unawareness causes the decision-maker to be more *sensitive* to objective risk as measured by the variance of a simple lottery. Ma and Schipper [2017], on the other hand, find no treatment effect on risk taking using a different manipulation to induce unawareness. Our results support the conclusion of a null effect on risk preferences, while highlighting the *context-dependent* nature of the impact of unawareness on decision-making.

There are two important distinctions between ours and the earlier experiments by Mengel et al. [2016] and Ma and Schipper [2017]. First, our design allows for having a measure of subjects' awareness of their own unawareness. As discussed in detail in Section 4.4.2.3, this measure mediates our treatment effects. Another important distinctive feature of our design is that the unawareness-inducing task is separate from the risk elicitation task, while both earlier studies use the same task both to induce unawareness and to measure risk preferences. We believe that such distance brings the design closer to real-world counterparts and , importantly, allows us to modulate the context of in which we assess risk taking. It likely also makes our experiment less susceptible to experimenter demand effects.[2]

The paper is organized as follows. Section 4.2 describes our novel experimental task. Sections 4.3 and 4.4 present the experimental design and results for our first-stage and main experiments, respectively. Section 4.5 provides a formal treatment of unawareness, and Section 4.6 concludes.

4.2 The Set Finding Task

To induce unawareness in subjects, we used a pattern matching task based on the card game SET™. A card in our game is described by a triple $\langle number, color, symbol\rangle$, where $n \in \{1, 2, 3, 4\}$, $c \in \{purple, red, orange, teal\}$ and $s \in \{triangle, square, circle, star\}$. The cards were presented visually, as in Figure 4.1, where the card $\langle n, c, s\rangle$ had n c-colored s-shaped symbols (i.e., the third card in Figure 4.1 is $\langle 3, orange, circle\rangle$ and has 3 orange circles). With three attributes each ranging over four values, there are a total of 64 unique cards. We refer to these 64 cards as the *deck*.

[2]See de Quidt et al. [2019] for a comprehensive survey on experimenter demand effects.

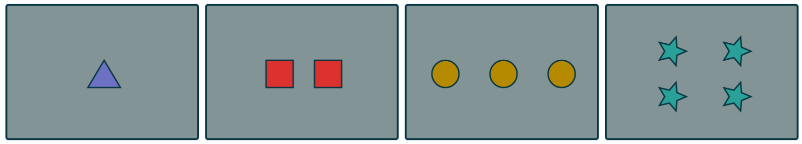


Figure 4.1: Example Of Cards Used In The Experimental Task

For the main task of the experiments, subjects were shown a selection of 12 cards drawn from the deck, which we refer to as their *hand*. From their hand, subjects had to select groups of three cards to form *sets*.[3] A *set* is any group of three cards such that the value for each attribute, $\langle number, color, symbol\rangle$, is either the same for *all* three cards or *pairwise* distinct for all three cards. Formally, a group of three cards, $\big(\langle n_1, c_1, s_1\rangle, \langle n_2, c_2, s_2\rangle, \langle n_3, c_3, s_3\rangle\big)$, is a *set* if

$$a_1 = a_2 = a_3 \qquad \text{or} \qquad a_1 \neq a_2, \quad a_1 \neq a_3, \text{ and } \; a_2 \neq a_3,$$

for each attribute $a \in \{n, c, s\}$.

Figure 4.2 provides an example. The first row constitute a set, whereas the second does not, as $s_1 = triangle$, but $s_2 = s_3 = circle$.

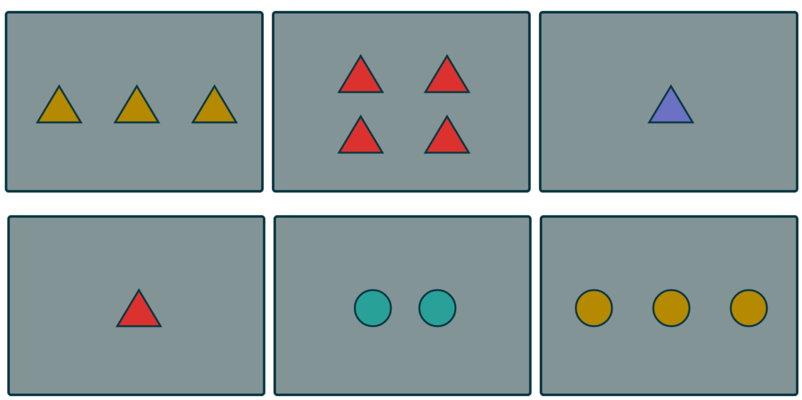


Figure 4.2: A Set And A Non-Set

The set-finding task has a number of desirable properties. First, the space of valid strategies is easy to delineate. That is, the rules of the game and the actions and consequences available to the subjects are straightforward to describe; the task is not open ended or

[3]The experimental design is explained in detail in Sections 4.3 and 4.4.

ambiguous. Second, despite it being easy to verify if three cards form a set, finding sets is hard, and the space of possible strategies (i.e., efficient ways to search for sets) is enormous and very hard to properly understand. Observing a hand of 12 cards does not easily reveal how many possible sets might be found from it, and even after interacting with the task, subjects make large mistakes when estimating their performance. Third, and to the best of our knowledge, the task has never been used in experiments before. Hence most subject will not have prior experience interacting with it. Fourth, the task appears to be gender and demographic neutral. And finally, it is a lot of fun.

Our experimental design is divided in two stages. In the first stage, we conducted simple experiments to measure participants' understanding and performance in the task. As explained in more detail later, we tested different combinations of task difficulty and levels of monetary incentives, and elicited participants' beliefs about their own performance. We used these first-stage results to aid in the design our main experiments, in which we study the impact of unawareness on risk preferences. Table 4.1 provides descriptive statistics for both stages. The subsequent Sections 4.3 and 4.4 describe each stage's design and results in detail.

Table 4.1: Descriptive Statistics

		Experiments	
		First-stage	Main
Number of subjects		246	971
Average payment (usd)		2.15	2.04
Completion time (min)		14.0	10.6
Female		.43	.49
Age	18-30	.38	.30
	31-45	.46	.46
	46-65	.16	.20
	65+	.00	.04
Education	High-school	.11	.11
	Some college	.38	.35
	College	.42	.39
	Masters+	.09	.15

Note: Table contains demographic information and descriptive statistics for participants in the first-stage and main experiments. Includes only participants that correctly answered the comprehension quiz and completed the experiment.

4.3 First-Stage Experiment

4.3.1 Experimental Design

We ran four sessions of the first-stage experiments on March and April of 2019. The experiments were done online using Amazon Mechanical Turk. We advertised it as a "short decision-making experiment" with a guaranteed payment of \$0.50 upon completion and the chance of an additional payment.[4]

The experiment begins by describing the composition of the cards and the resulting 64-card deck. We then describe the properties of a *set*—using a number of examples for clarity—and have subjects complete a comprehension quiz which asks them to classify five different 3-card combinations as either a *set* or *not a set*. Only subjects that answered 4 or more questions correctly were allowed to continue; the others were automatically eliminated and paid the \$0.50 participation fee.[5]

The experiment consisted of two rounds. For each round subjects were shown a 12-card hand and paid a piece rate for each set found during a 120-second time window. Subjects faced a 5-second penalty if they either selected three cards that did not form a set, or if they selected a set that had already been found. This rule was intended to minimize the incentives for random clicking. After each round, participants were asked to estimate the share—from 0 to 100%—of the total number of sets they believed they had found. We incentivized this elicitation by paying an extra 50 cents if the estimate was within 5% of the true value, on either side.[6]

We implemented a 2×3 design with the goal of assessing subjects' performance and understanding of the task, with both between and within-subject variation. First, each subject was randomly assigned to either the *low* (\$0.10) or *high* (\$0.35) piece rate treatment.

[4]See Appendix **??** for complete screenshots and instructions. Interested readers can complete the experiment by accessing `https://faep.herokuapp.com/` and typing as a username any combination of 5 or more alphanumeric characters.

[5]We used four different quiz versions that we believed, ex-ante, to be equally difficult. Ex-post analysis revealed that one version was relatively hard and another version was relatively easy (see Appendix **??** for details). Note, however, that since treatments were randomly assigned independently of the quiz version, differences in quiz success rates do not affect our results.

[6]Because of a coding error, we lost approximately one third of our belief-elicitation data. Information on which author was responsible for the mistake is available upon request.

Second, we randomly assigned one of four different types of hands at the round level: with 10, 15, 23 or 28 total number of sets. Each hand was randomly selected, without replacement, from a collection of 20 possible hands, 5 of each type. Subjects did not get any information about the hand and were told only that they would "see a 12-card hand that was randomly selected from a larger list of 12-card hands".

4.3.2 Results

Table 2 reports regression results for the first-stage experiments. For each outcome of interest, we report random-effects regressions of the outcome of interest on treatment dummies.[7]

Table 4.2: Number Of Sets Found And Belief Gaps For First-Stage Experiments

	Dependent Variable			
	Sets Found		Belief Gaps	
	(1)	(2)	(3)	(4)
Constant	3.05*** (.20)	3.85*** (.19)	24.49*** (2.83)	30.72*** (2.70)
15-set hand	−.04 (.25)		6.50* (3.97)	
23-set hand	.93*** (.24)		12.71*** (3.82)	
28-set hand	.70*** (.26)		13.76*** (4.01)	
$0.35 piece rate		−.75*** (.26)	2.60 (3.69)	
N	492	492	331	331

Note: Figures are derived from random-effects regressions of dependent variable on treatment dummies (*10-set hand* and *$0.10 piece rate* are the omitted categories).

We are mostly interested in three outcomes. First, we ask if subjects find more sets in hands with a larger number of possible sets. The answer is "yes", but not much more. Subjects find on average an extra 0.9 and 0.7 sets for hands with 23 or 28 total sets, respectively, when compared to 10-set hands; there is no difference between 10 and 15-set hands. Second, we ask if performance in the task is sensitive to monetary incentives. Indeed it is, but *negatively* so: subjects in the 10-cent treatment find significantly more sets than subjects in

[7]We also estimated models (3) and (4) using random-effects Tobit regressions, with similar results.

the 35-cent treatment. The last measure we study is the belief gap: the difference between the believed and actual share of total sets found. Belief gaps are positive and large for *all* treatments, and increase monotonically with the total number of sets in the hand seen in model (3). As per column (4), and there is no direct impact of the piece rate on belief gaps.

Finally, Figure 4.3 provides evidence that the time constraint in the set-finding task is not binding. It depicts the total number of sets found, per round and combining all treatments, for each 10-second block. If the time limit was binding, we would expect to see either a uniform distribution or a spike closer to the end. We instead observe a larger number of sets at the beginning and a decline towards the end of the 120 seconds.

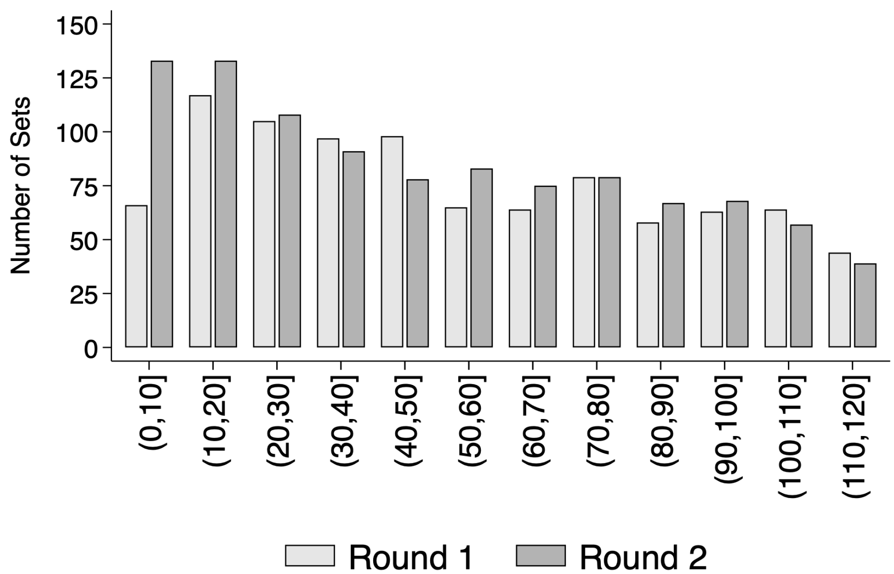


Figure 4.3: Timing Of Sets Found

Note: Figure plots the total number of sets found per 10-second block for rounds 1 and 2.

Results from the first-stage experiments informed the design of our main experiment, which is discussed in next section.

4.4 Main Experiment

4.4.1 Experimental Design

We ran four sessions of the main experiment in May and October of 2019 on Amazon Mechanical Turk. Sessions were advertised as a "short decision-making experiment" with a participation fee of $0.50 and the chance for an additional payment.[8] As in the first-stage experiment, we started by describing the cards and the characteristics of a set, and moved on to a comprehension quiz. Again, only subjects that answered 4 or more questions correctly were allowed to continue; the others were automatically eliminated and paid the $0.50 participation fee. Subjects approved in the quiz proceeded with the experiment, which consisted of two tasks.

Task 1 was a set-finding exercise. Subjects were presented with a 12-card hand and given 120 seconds to find sets. As before, they incurred a 5-second penalty for selecting a combination of cards that either did not constitute a set or that had already been found. The piece rate for each correct set was fixed at $0.10, and all of the 12-card hands had 28 sets in total.[9] Immediately after completing task 1, subjects were asked to estimate the share of the total number of sets they believed they had found. This elicitation was not incentivized.

The piece rate and number of possible sets contained in the hand were chosen based on the results of the first-stage experiment. First, we observed only a small difference between the number of sets found under the $0.10 or $0.35 piece rates; indeed, subjects found significantly *more* sets under the lower piece rate. We opted for the $0.10 incentive to obtain a larger sample size given budgetary constraints. Second, even though subjects found more sets from hands that contained more sets, belief gaps—the difference between subjects *beliefs* about and the *actual* share of total sets found—were largest for the 28-set hands. We anticipated using variation in this gap to study the role of subjects' contemplation of their own unawareness (see Section 4.4.2.3 for details), and hence decided to use only 28-set

[8] See Appendix ?? for complete screenshots and instructions. Interested readers can complete the main experiment by accessing `https://faep2.herokuapp.com/` and typing as a username any combination of 5 or more alphanumeric characters.

[9] To minimize the scope for cheating in the online environment, each subject was shown a randomly selected hand drawn from a collection of 20 different 12-card hands. The instructions informed subjects that they would "see a 12-card hand that was randomly selected from a larger list of 12-card hands".

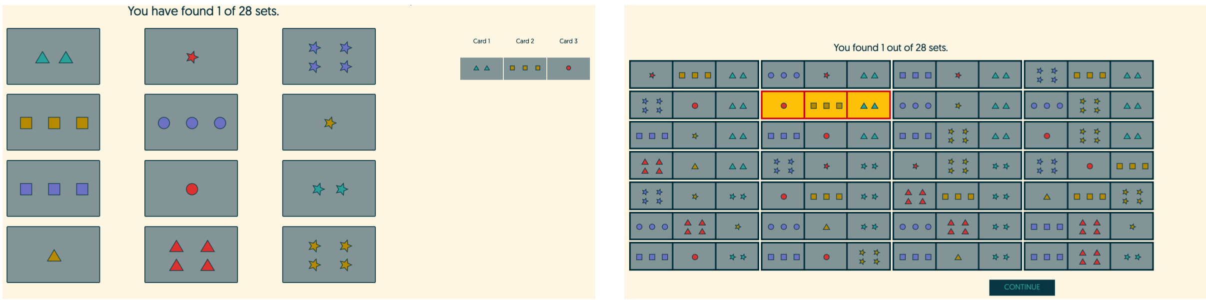
hands.

After completing task 1, subjects moved on to task 2, which elicited risk preferences using a version of the investment game [Gneezy and Potters, 1997]. At the start of task 2 subjects were endowed with an additional \$1 and asked to decide how much to invest in a risky option. Specifically, subjects were asked how much they wanted to "keep safe" and how much they were willing to "bet" on a risky lottery. Subjects could not lose the amount kept safe, but there were two possible outcomes for the amount bet: (i) the *good* outcome, where the money bet was multiplied by 3; or (ii) the *bad* outcome, where the money bet was lost. Each outcome occurred with 50% probability, so the investment had a positive expected rate of return.

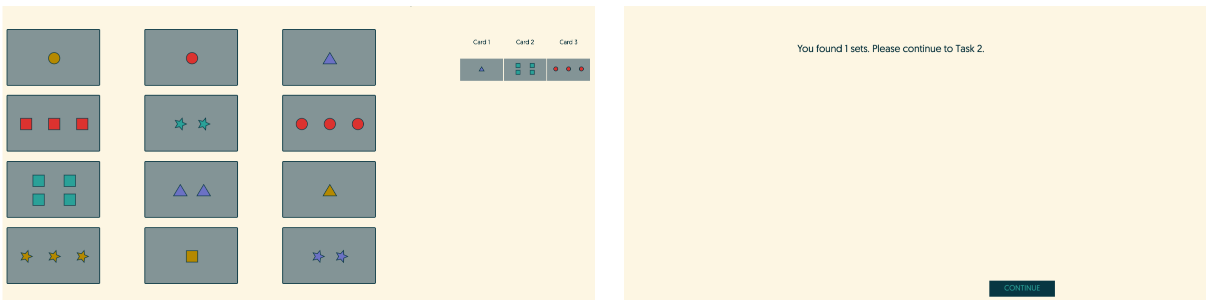
We implemented three treatments varying the nature and extent of subjects' knowledge about the set-finding task, and two treatments varying the context in which the betting decisions were made. Thus, we have a 3×2, between-subject experimental design. We next describe the differences between the knowledge treatments. Figure 4.4 displays those differences using screenshots of the experimental interface.

- **Full Information:** This is our control treatment, where we shut down the unawareness channel by informing subjects of the number of available sets prior to completing task 1. Subjects saw a counter at the top of the set-finding screen stating "You found n of 28 sets", where n was the number of sets found at any given point in time. After task 1 and the belief elicitation stage, subjects were shown an array of all sets contained in their hand, with the ones they had found highlighted.
- **Unawareness:** Subjects were not informed about the number of possible sets. After completing task 1 and the belief elicitation, subjects were simply reminded of the number of sets they had found.
- **Unawareness-Info:** Subjects were not informed about the total number of possible sets prior to completing task 1, as in the *unawareness* treatment. However, after completing task 1 and the belief elicitation, subjects were shown an array of all sets contained in their hand, with the ones they had found highlighted.

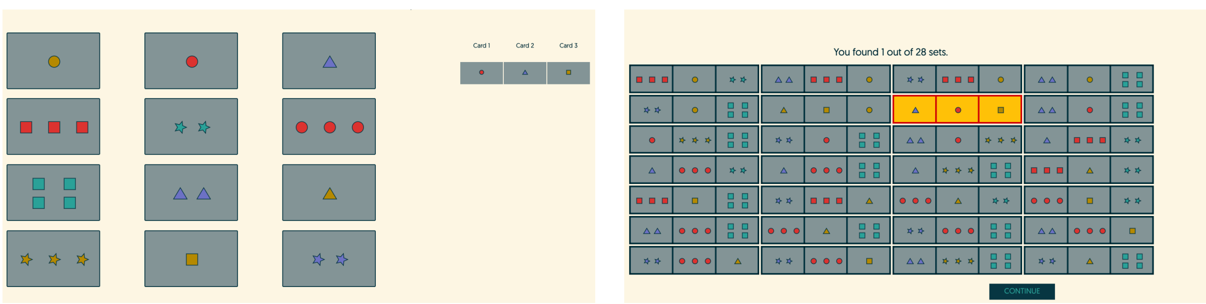
Next, we describe the treatments varying the *context* of the risky decision. The only



(a) Full Information



(b) Unawareness



(c) Unawareness-Info

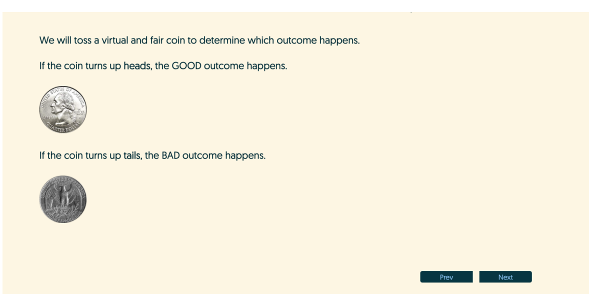
Figure 4.4: Screenshots Of Set-Finding And Feedback Screens

treatment variation is related to the description of the objective risk; the actual probabilities of the *good* and *bad* outcomes, as well as the amount available to invest and the payoff structure, were invariant across treatments. Figure 4.5 presents screenshots of the investment stage for both treatments.

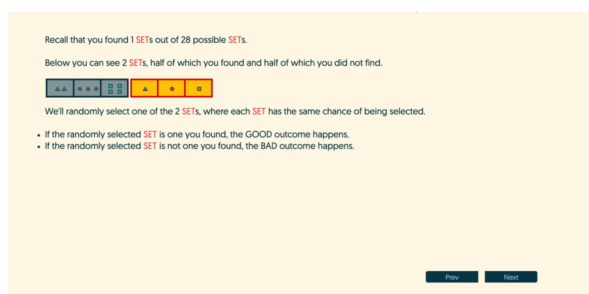
– **Neutral:** The chances of the good and bad outcomes were framed as the outcome of tossing a fair coin. Subjects saw pictures of the two sides of a coin and were told that we would "toss a virtual and fair coin to determine which outcome happens". If the coin turned up heads, the good outcome occurred; and if the coin turned up tails, the bad outcome was realized.

- **Context:** The framing of the random outcome was done using the sets found and *not* found by the subject. Specifically, subjects saw an array of sets half of which was populated by the sets found and half by the sets not found. The sets found by the subject were highlighted. Subjects were informed that we would "randomly select one of the n sets, where each set has the same chance of being selected". If the randomly selected set was among those found by the subject, the good outcome occurred; otherwise, the bad outcome realized.

Note that if a subject did not find *any* sets, she could not be assigned to the *context* treatment, and hence was automatically assigned to the *neutral* version. For this reason all of the analysis in this section restricts attention to subjects that found at least 1 set—92% of the sample. Lastly, to increase confidence in our instructions, the actual randomization was implemented using public lottery outcomes from the Pennsylvania Lottery. This was explained to the subjects prior to the betting decisions and described in the study's pre-registration.[10]



(a) Neutral Framing



(b) Context Framing

Figure 4.5: Screenshots of Different Framings for Risky Decision

[10] AEARCTR-0004145

4.4.2 Results

We will present the results of our main experiment in three parts. First, we explore the differences between the knowledge treatments with respect to (i) the number of sets found, (ii) beliefs gaps, and, most importantly, (iii) the share of the endowment bet. Second, we investigate risky attitudes between the context treatments. Third, we examine how the role of context interacts with subjects' degree of salience of unawareness, providing evidence for the channel by which unawareness changes subsequent choice behavior.

4.4.2.1 Knowledge Treatments Table 4.3 presents regression results for three dependent variables: (i) the number of sets found, (ii) belief gaps, and (iii) the share of the endowment bet. We report models with and without controls, which include hand fixed-effects,[11] demographics (age group, gender, and education level) and, for some models, the number of sets found. Treatment *full information* is the omitted category in all columns.

Note first, from columns (1) and (2), that subjects found the same number of sets, on average, across the three knowledge treatments. Being fully aware of the number of available sets prior to completing task 1 did not impact subjects' productivity, which speaks to the inherent complexity of our task. Second, although there are no differences in the number of sets found, there are stark differences in belief gaps, i.e., the difference between subjects beliefs about and the actual share of total sets found—columns (3) and (4). The belief gap is positive and significant for all treatments. The gap is, however, almost two and a half times larger for the *unawareness* and *unawareness-info* treatments. In Section 4.4.2.3 we explore this gap further when studying the role of context. Finally, columns (5) and (6) of Table 4.3 reveal no treatment effect with respect to risk preferences. On average, subjects bet close to 40% of their endowment in the risky lottery.

4.4.2.2 The Role of Context Suppose you just read a magazine article about the discovery of a new super virus that is afflicting your country. You have thus suddenly become aware of a hitherto unforeseen risk to your health. Does this increased awareness

[11] Not all of the 20 different 12-set hands had similar levels of difficulty. The average number of sets found was 3.65 with a standard deviation of 1.14. The lowest average was 1.84 and the maximum 5.63.

Table 4.3: Number of Sets, Belief Gaps, and Share Bet

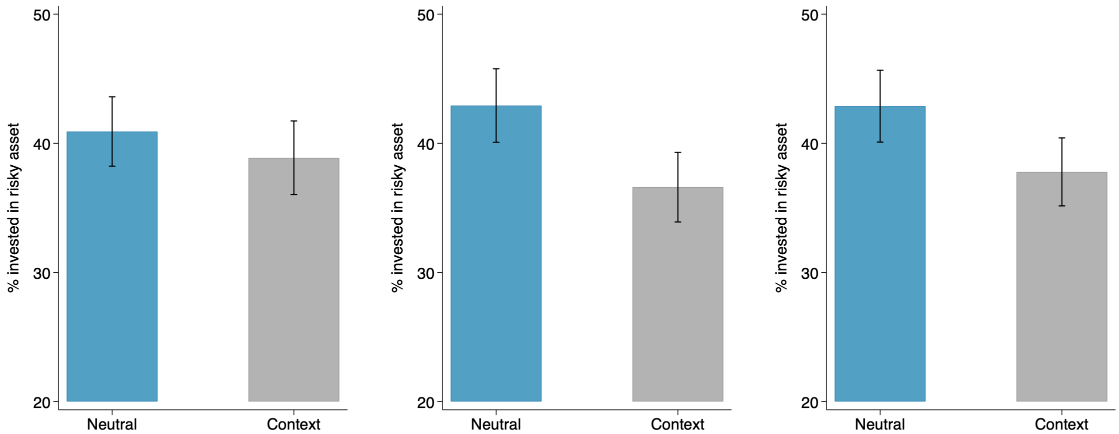
	Dependent Variable					
	Sets Found		Belief Gaps		% Bet	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	4.11***	4.02***	12.55***	13.30**	39.86***	38.95***
	(.15)	(.59)	(1.40)	(5.34)	(1.96)	(7.11)
Unawareness	−.20	−.19	17.90***	17.41***	−.21	.10
	(.21)	(.19)	(2.15)	(2.15)	(2.78)	(2.82)
Unawareness-Info	−.07	−.08	17.42***	17.60***	.53	.34
	(.21)	(.20)	(2.06)	(2.06)	(2.75)	(2.81)
Controls		X		X		X
N	890	887	890	887	890	887

Note: Figures are derived from OLS regressions of dependent variable on treatment dummies (*full information* is the omitted category). Only includes subjects who found at least 1 set (92% of total). Controls include hand fixed-effects and dummies for age group (18-35, 36-45, 46-65, and 65+), education level (high-school, some college, college, and masters+) and gender. For models (3) to (6) controls also include the number of sets found. We also run Tobit regressions for models (5) and (6), with similar results.

affect your risk attitudes? And if it does, is the effect restricted to the health domain or does it spill over to other areas (financial, relationships, etc.)? This is the primary motivation for our second research question: are changes in risk taking behavior induced by unawareness domain dependent? In other words, are changes in risk taking behavior dependent on the risky decision being contextually linked to the domain in which the decision maker was exposed to unawareness?

Within each knowledge treatment, subjects saw one of two randomly chosen descriptions of the risky asset. In the *neutral* condition, the outcome was determined by a (virtual) coin flip, whereas in the *context* condition it was determined by randomly drawing a set among a list of sets contained in the subject's hand, half of which the subject had found. In both cases, the chance of the good outcome was identical: 50%. We ask if the framing affects the amount bet in the risky lottery.

Figure 4.6 depicts, for each knowledge treatment, the average share of the endowment bet for each framing. An interesting pattern emerges: the context of the risky decision is relevant for treatments *unawareness* and *unawareness-info*, but not for the control treatment. In line with our earlier example, it's as if experiencing unawareness about the virus causes people to become more risk averse in their health behaviors, an attitude that doesn't carry over to other domains.



(a) Full Information (b) Unawareness (c) Unawareness-Info

Figure 4.6: Risky Investment and Context

Note: Average share of the endowment invested in risky asset for each combination of knowledge (*full information*, *unawareness*, and *unawareness-info*), and framing (*neutral* and *context*) treatments. Error bars indicate standard errors of the mean.

Table 4.4 reports the same results using linear regressions, both with and without controls. As before, the set of controls include hand fixed-effects, demographics (age group, gender, education level) and the number of sets found.

Columns (1) and (2) report a significant negative coefficient for the *context* treatment. On average, subjects bet 4.5 percentage points (p.p) less compared to the *neutral* lotteries. This corresponds to a decrease of a little over 10% in the amount bet. Models (3) and (4) break down the effect by knowledge treatment. It's clear from those regressions that the impact of context in the amount bet is coming from the treatments *unaware* and *unaware-info*. Due to the lower sample sizes, however, coefficients are not significant for each separate treatment.

Columns (5) and (6) combine data from treatments *unawareness* and *unawareness-info*. Recall that both treatments are identical up to the end of task 1. Moreover, both treatments have similar results with respect to the share of the endowment bet and number of sets found. With the resulting increase in statistical power, models (5) and (6) find a negative and statistically significant effect of context on our measure of risky behavior. Framing

Table 4.4: Share of Endowment Bet in Risky Lottery

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	42.18***	41.46***	40.92***	39.98***	40.92***	40.02***
	(2.23)	(7.24)	(2.69)	(7.40)	(2.68)	(7.40)
Unaware	−.21	0.04	2.02	2.65		
	(2.77)	(2.81)	(3.91)	(4.05)		
Unaware-Info	.40	0.20	1.96	1.92		
	(2.74)	(2.80)	(3.86)	(3.98)		
Context	−4.48**	−4.25*				
	(2.24)	(2.27)				
Full Information × Context			−2.04	−1.49	−2.04	−1.50
			(3.92)	(4.01)	(3.92)	(4.01)
Unaware × Context			−6.32	−6.47		
			(3.92)	(4.04)		
Unaware-Info × Context			−5.09	−4.75		
			(3.83)	(3.89)		
Unaware-All					1.98	2.26
					(3.34)	(3.48)
Unaware-All × Context					−5.71**	−5.59**
					(2.74)	(2.78)
Controls		X		X		X
N	890	887	890	887	890	887

Note: Figures are derived from OLS regressions of dependent variable on treatment dummies (*full information* is the omitted category). Only includes subjects who found at least 1 set (92% of total). Controls include hand fixed-effects, number of sets found, and dummies for age group (18-35, 36-45, 46-65, and 65+), education level (high-school, some college, college, and masters+) and gender. We also run Tobit regressions for all models, with very similar results.

the betting scenario in the same domain as the unawareness-inducing task causes a drop of approximately 13% in the amount bet in the risky lottery.

There are two possible channels by which subjects become more risk-averse: first, that their risk preferences change, i.e., they become less tolerant of risk; second, that their perception of probabilities, albeit intended to be objective, change. Do subjects assess the objective risk differently in each context treatment?[12] These could be caused, for example, by an unfamiliarity with the description: a coin toss is more readily understandable than selecting one set out of a group of n sets. Or maybe the experienced unawareness itself might affect individuals' ability to assess the objective risk component.

With that in mind, we added a second belief elicitation for the latter half of our sessions. After describing task 2, we elicited subjects' beliefs about the chances of the good outcome

[12] We thank David Huffman for raising this possibility.

happening. We incentivized this stage by paying an additional \$0.50 if the question was correctly answered.

Indeed subjects in the *context* treatment were less likely to answer correctly: while 87% of subjects responded correctly for the *neutral* treatment, only 75% did so for the *context* treatment ($p < 0.00$). Also, the average guess for *neutral* and *context* were, respectively, 51.0% and 47.1% ($p = 0.01$, two-sided t-test). Note, however, that the overall impact of the assessed probability on betting behavior is small: 0.15p.p per additional percentage point assigned to the good outcome (see Appendix **??** for details).

Hence, if we assume the estimated coefficient would be similar for the entire sample (recall that we only elicited beliefs about probability for $\sim 45\%$ of the sample), the observed gap of 3.9p.p in the assessment of the objective probability ($51\% - 47.1\%$) translates into an extra $0.15 \times 3.9 = 0.59$ percentage points of the endowment invested in the risky asset. That is only about 10% of the observed difference in the amount bet by context treatment (5.59p.p per column (6) of Table 4.4). We conclude that the main channel by which risk taking behavior changes is via changes in risk tolerance rather than changes in assessment of objective probabilities.

4.4.2.3 Salience of Unawareness We next dig deeper into the role of context by examining the interaction between the context and subjects' degree of *salience of unawareness* (henceforth SOU). We are motivated by the fact that, as discussed in Section 4.4.2.2, treatments *unawareness* and *unawareness-info* have very similar results. This suggests a limited role of information disclosure after subjects experience the unawareness-inducing task. As such, we conjecture that the most important driver of differences between the full information and the unawareness treatments are differences in the salience of unawareness, i.e., subject's introspective understanding of the task before the state space is revealed.

Our measure of SOU is the belief gap: the difference between the share of sets a subject thought she had found and the actual share of sets found. The higher (lower) the difference, the lower (higher) the degree of SOU. For example, suppose subject 1 guessed he had found 30% of the sets, but had actually found only 15%, while subject 2's guess were 45% and 40%, respectively. The difference for subject 1 is thus +15% and for subject 2 it is +5%. We then say subject 2 is *more* aware of her own unawareness compared to subject 1. Figure 4.7 compares the share of endowment bet, by treatment and context, for subjects with above-median and below-median SOU.

As Figure 4.7 makes clear, the difference between investment choices in the *neutral* versus *context* treatments comes almost exclusively from subjects with above-median salience of unawareness and only for the treatments that induce unawareness. Intuitively, subjects that are more *aware* of their own unawareness are more susceptible to having their risk preferences affected by making a risky decision within the same context as the unawareness-inducing task.

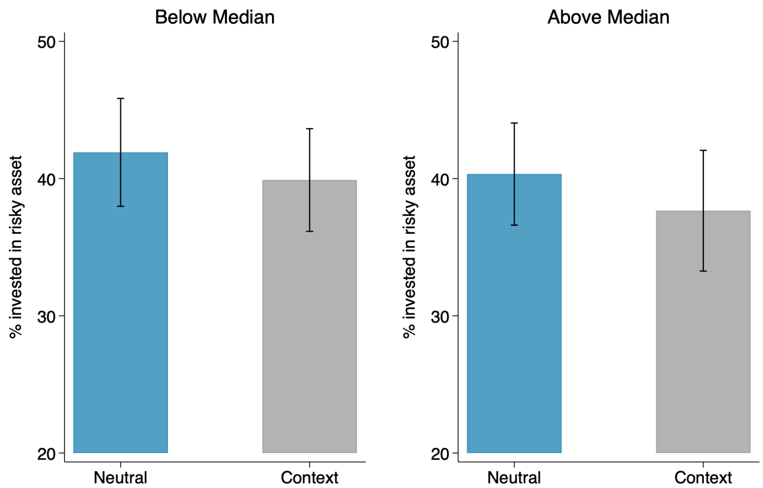
4.5 A Simple Model of Preference Change Following Discoveries

In this section, we outline a simple model that can help explore the channel by which the salience of unawareness can dictate preferences. In an effort to convey meaning with the minimal notational and technical investment, we introduce the abstract model alongside a running example. Also, we gloss over many technical details but, when doing so, make sure to point to the relevant part of the theoretical literature on awareness.

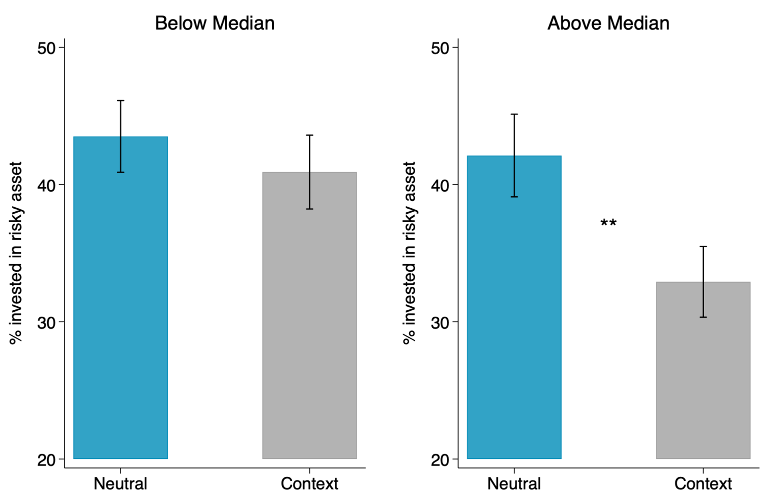
An agent faces a decision problem which depends on the truth or falsity of a set of uncertain statements, $\mathcal{L}$. The agent is not aware of all statements, but rather of a subset $\mathcal{A} \subseteq \mathcal{L}$. The uncertainty faced by the agent is modeled by a state-space, Ω. Unlike in the standard state-space model where a state, $\omega \in \Omega$, represents the resolution of *all* statements, here each state ω is associated with a set of propositions $\mathcal{L}(\omega)$ and determines the truth only of $\varphi \in \mathcal{L}(\omega)$. The true of the propositions is given by the family of truth functions

$$\left\{t_\omega : \mathcal{L}(\omega) \to \{\mathbf{T}, \mathbf{F}\}\right\}_{\omega \in \Omega}.$$

The interpretation is that the agent might consider possible different states in which different objects, concepts, properties, etc. *exist*: the statement $q =$ "there is a quantum algorithm breaking protocol x," makes sense only in those states in which (the concept of) quantum computers exist. Allowing the set of statements to vary over the state space, while adding



(a) Full Information



(b) Unawareness-All

Figure 4.7: Salience Of Unawareness And Investment In The Risky Asset

Note: Figure (A): Share of endowment bet in the *full information* treatment, by context, for subjects above and below the corresponding median of the *salience of unawareness* measure. Figure (B): Share of endowment bet in the *unawareness* treatments, by context, for subjects above and below the corresponding median SOU. Error bars indicate standard errors of the mean.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

complexity to the model, is requisite in capturing agents who are unsure of how unaware they are.

Critically, the statements in $\mathcal{L}$ discuss not only physical aspects of the world, but also the agent's own perception. For example, it is perfectly reasonable to entertain the statement "the agent is fully aware" which would be true exactly in the states where $\mathcal{A} = \mathcal{L}(\omega)$.[13] This allows us to model how an agent reasons about her own awareness (and knowledge), and in particular allows for introspective unawareness—when the agent considers it possible that she might be unaware and takes this into account when making decisions.

Example. *Neelay Mehta is deciding whether to use the new cryptographic protocol x to secure financial records for his business. He will only change protocols if x is sufficiently likely, say 75% likely, to be secure. In addition, being somewhat of a computer guy, Neelay knows how the cryptographic algorithm underlying protocol x works and that it is secure in the context of his understanding of how a computer works. Of course, he admits that it is possible that there exists models of computation outside of his understanding which might render x vulnerable.*

There are three relevant statements,

$p =$ *"cryptographic protocol x is secure,"*

$q =$ *"there is a quantum algorithm breaking protocol x," and*

$a =$ *"the agent is initially fully aware,"*

and Neelay's awareness is $\mathcal{A} = \{p, a\}$.

We can represent his subjective uncertainty with three states: $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Let $\mathcal{L}(\omega_1) = \{p, a\}$ and $\mathcal{L}(\omega_2) = \mathcal{L}(\omega_3) = \{p, q, a\}$ and the truth functions

$$t_{\omega_1} : \begin{cases} p \mapsto \boldsymbol{T} \\ a \mapsto \boldsymbol{T} \end{cases} \qquad t_{\omega_2} : \begin{cases} p \mapsto \boldsymbol{T} \\ q \mapsto \boldsymbol{F} \\ a \mapsto \boldsymbol{F} \end{cases} \qquad t_{\omega_2} : \begin{cases} p \mapsto \boldsymbol{F} \\ q \mapsto \boldsymbol{T} \\ a \mapsto \boldsymbol{F} \end{cases}$$

[13]That statements regarding the awareness and beliefs of the agent can themselves be *consistently* assigned truth values over Ω is not obvious, but it is true. The proof of this is tangential to our aims here, but can be found in various forms in Fagin and Halpern [1987], Halpern and Rêgo [2013], Piermont [2019a].

Notice: he knows that if he is fully aware, then p is true, but if he is not fully aware then p might be true or false. Further notice that a is true exactly when he is fully aware. Notice also, that although Neelay contemplates the contingency that he is unaware (the states where a is false), he does not know what *he is unaware of.* $\square$

The agent's awareness defines a partition of the state-space via the $\mathcal{A}$-*indistinguishability* equivalence relation $\omega \sim_{\mathcal{A}} \omega'$. Two states, ω and ω', are $\mathcal{A}$-indistinguishable whenever they yield the same truth values for all statements the agent is aware of. Formally:

$$\omega \sim_{\mathcal{A}} \omega' \iff (t_\omega(\varphi) = t_{\omega'}(\varphi)) \text{ for all } \varphi \in \mathcal{A}.$$

If $\mathcal{A} \subseteq \mathcal{A}'$, then the partition generated by $\sim_{\mathcal{A}'}$ is a refinement of that generated by $\sim_{\mathcal{A}}$: as the agent becomes more aware she can distinguish between more states so her perception of the state space becomes finer.

The agent also entertains a probability measure, π, over the cells in the partition generated by $\sim_{\mathcal{A}}$. For consistency, we must require $\pi(E) > 0$ implies $\mathcal{A} \subseteq \mathcal{L}(\omega)$ for all $\omega \in E$, so that the agent rules out events incompatible with her current awareness. This prohibits the nonsensical state-of-affairs in which the agent is aware of some statement p (for example "quantum computers are fast") but also believes it possible that concept of p does not exist.

Example (continued)**.** *Since Neelay's awareness is $\mathcal{A} = \{p, a\}$, he can distinguish between all states: $\sim_{\mathcal{A}}$ generates the partition $\{\{w_1\}, \{w_2\}, \{w_3\}\}$. Indeed, in state ω_1 a is true but false everywhere else, and states ω_2 and ω_3 can be distinguished by the truth of p. Thus, his probabilistic beliefs assign a probability to each state. Perhaps, in a bout of overconfidence, he thinks it's exceedingly likely that he fully understands computation, embodied by the probability $\pi = (\frac{8}{10}, \frac{1}{10}, \frac{1}{10})$. Thus his evaluation of the truth of p, is*

$$\pi(\{\omega \mid t_\omega(p) = \boldsymbol{T}\}) = \pi(\{\omega_1, \omega_2\}) = \frac{9}{10}$$

exceeds his threshold and he will adopt protocol x. $\square$

Of course, to discuss the results of our experiment, we must move past the static model, and examine what happens when the agent learns or becomes more aware. This model accommodates the standard information acquisition problem: when the agent learns that φ is true, her beliefs update via Bayes' rule, and her new beliefs are given by

$$\pi^{new}(E) = \frac{\pi(E \cap \{\omega \in \Omega \mid t_\omega(\varphi) = \mathbf{T}\})}{\pi(\{\omega \in \Omega \mid t_\omega(\varphi) = \mathbf{T}\})}.$$

It can also accommodate the discovery of a novel statement, which represents the expansion of awareness. Often, the discovery of a novel statement is accompanied by learning that that statement is true, but this can be seen as the combination of first becoming aware of φ and then learning that φ is true—the latter belief change being handled by conditioning in the usual way.

The expansion of awareness is captured as a movement from $\mathcal{A}$ to $\mathcal{A}' \supset \mathcal{A}$. As mentioned above, this means her perception will become finer, so that her new beliefs, π^{new}, may be defined on a different, finer partition. Even when the agent becomes aware of a novel statement φ without learning the truth of φ, her new beliefs will assign a probability to the event that $\{\omega \in \Omega \mid t_\omega(\varphi) = \mathbf{T}\}$. Note that this event was not (necessarily) in measured by $\sim_{\mathcal{A}}$ and therefore not assigned a probability by her initial probability measure.

Despite not learning the truth of φ, the agent's beliefs *do* change. Why? Because $\mathcal{L}$ concerned not only statements about the physical world, but also statements regarding the epistemology of the agent. As a simple example, if the agent initially put positive probability on a state ω in which she was fully aware (so that $\mathcal{A} = \mathcal{L}(\omega)$), then after discovering φ she *must* disregard ω (since now $\mathcal{A} \not\subseteq \mathcal{L}(\omega)$). Therefore, the very act of becoming aware requires conditioning on the event $\{\omega \in \Omega \mid \omega$ was compatible with discovering $\varphi\}$.[14]

Notice that discovering a novel statement implies that the agent learns that $a =$"the agent was fully aware" was false instigating a belief update in a traditional manner. Of course, the agent need not discover any particular statement to learn the statement a. For

[14]Again, that there exists a consistent notion of what it means for a state to be *compatible with discovering* a statement is not self-evident. This is also true and also outside of the scope of this outline, but a formal theory is put forth in Halpern and Piermont [2020]. Notice also that the conditioning event in question might not in measured by $\sim_{\mathcal{A}}$ so standard Bayesian updating cannot be applied as there is no ex-ante probability by which to condition. This can be circumvented by appealing to a slight generalization of Bayesian Updating, the behavioral theory of which is outlined in Piermont [2019b].

example, by interacting with a game and observing inexplicable results, a player could learn that she does not know all the rules without explicitly discovering any additional rules.

Example (continued)**.** *Neelay reads a blog that alludes to the 'computational paradigm outside of the classical computer.' Although he continues to be unaware of the specific reference the author makes, he nonetheless learns that a is false. His updated beliefs are given by* $\pi^{new} = (0, \frac{1}{2}, \frac{1}{2})$.

Notice in this contrived example, the resulting beliefs are exactly the same as when he instead reads a blog post explaining quantum computing, discovering the existence q without learning its truth. Even though he learns nothing about the truth of q, only states ω_2 and ω_3 are compatible with discovering anything at all! Hence his new beliefs are constructed exactly by conditioning on a being false.

By either learning he is unaware, or discovering what he is unaware of, Neelay moves from a position of wanting to adopt the new protocol to one of avoiding it. □

Notice how, in the example, there is correlation between unawareness/complexity and beliefs about other aspects of the world. In the event where Neelay is not fully aware of the intricacies of cryptography, it is more likely that protocol x is insecure—that is, he considers the protocol breakable, but only if it is breakable in a way he does not currently understand.

More generally, it is this correlation between unawareness and the payoffs associated with various actions that induces behavior change in response to the salience of unawareness. That the change in actions following discovery are generally more cautious is not dictated by the model, but seem to be an empirical trend akin to uncertainty aversion. A reasonable explanation is that subjects correlate being unaware with the environment being more complicated, and thus inducing more uncertainty about payoffs; the standard motivations for uncertainty aversion, then, induce cautious behavior.

The final piece of the puzzle is that we can reason, loosely, about what types of things we are unaware of. Thus the correlation between unawareness and payoffs is context dependent. Learning that there are novel computational models that might be able to break a cryptographic protocol will change the agent's perception of the uncertainty regarding the efficacy of data storage methods, but likely not on the risks associated with climate change.

4.6 Conclusion

We introduce a novel experimental task designed to induce unawareness in a natural way, namely, a task whose complexity is not easily grasped. We then manipulate subjects' exposure to unawareness and study the impact of unawareness on risk taking when the risky domain is either described in a neutral fashion or framed in the same context as the unawareness-inducing task. First, we find strong evidence that the unawareness manipulation works, as the gap between self-assessed and actual performances is much larger for the unawareness treatments compared to the control treatment. Second, even though we find no average treatment effect on risky behavior, we find evidence of a differential impact of context: subjects are shown to be more risk averse in the contextual domain, but *only* in treatments that induce unawareness. Finally, these effects are mediated by the individual's salience of unawareness, as only subjects that are aware of their own unawareness are impacted by the context of the risky decision.

Our results underscore the importance of considering the context of the decision when evaluating the behavioral effects of unawareness. This insight can potentially have implications for future work, both empirical and theoretical, on the interaction of unawareness on information disclosure and learning.

Appendix A Appendix for Chapter 2

A.1 Survey Instrument

This appendix contains an English translation of the questions used in the survey with undergraduate students. The survey was conducted over two weeks on November, 2018. The order of the questions on public and private sector wage expectations and preferences was randomized for each individual. The (optional) open questions at the end were individualized depending on the answer to the initial questions on sector preference.

Hello!

This research is part of a study about the labor market in Brazil and takes, on average, 17 minutes to complete. Your participation is anonymous and voluntary. At the end of the survey, you can enter to win one of 75 gift-cards from Saraiva Bookstore worth R$ 100 each. If you decide to participate, we will ask you for your email in order to send the gift-card if you are one of the winners.

Felipe A. Araujo is the researcher responsible for this study. Felipe graduated from UFJF in 2008, and is currently pursuing a PhD in economics at the University of Pittsburgh, USA. He can be reached by email [f.araujo@pitt.edu] or by phone [xxx-xxxxxx].

Q1 After graduating, you plan to:

- ○ Find a job
- ○ Go to graduate school
- ○ Open your own business
- ○ Other (please specify) ____________________

Q2 If you decide to look for a job, would you prefer to get a job in the public or private sector?

- ○ Public sector
- ○ Private sector
- ○ I am indifferent

Q3 What is the chance that you will take at least one public-sector exam after graduation?

- ◯ Very low
- ◯ Low
- ◯ Neither high nor low
- ◯ High
- ◯ Very high

Q4 How old are you?

- ◯ 16
- ◯ …
- ◯ 39
- ◯ 40 or more

Q5 What's your gender?

- ◯ Male
- ◯ Female
- ◯ Other

Q6 What year are you attending?

- ◯ First
- ◯ Second
- ◯ Third
- ◯ Fourth
- ◯ Fifth or higher

Q7 What level are you currently attending?

- Undergraduate
- Specialization
- Masters
- Doctorate
- Post-doc

Q8 What is your major?

- Business Administration
- Public Administration
- Architecture and Urbanism
- Arts
- Visual Arts
- Cinema and Audiovisual
- Design
- Fashion
- Arts and Design
- Humanities
- Computer Science
- Religion
- Biological sciences
- Accounting
- Economics

- Natural Sciences
- Social Sciences
- Communication
- Law
- Nursing
- Environmental and Sanitary Engineering
- Civil Engineering
- Computational Engineering
- Production Engineering
- Electrical Engineering
- Electrical Engineering - Robotics and Industrial Automation
- Electrical Engineering - Energy
- Electrical Engineering - Power Systems
- Electrical Engineering - Electronic Systems
- Electrical Engineering - Telecommunications
- Mechanical Engineering
- Statistics
- Pharmacy
- Philosophy
- Physics
- Physiotherapy

- Geography
- History
- Journalism
- Literature
- Mathematics
- Medicine
- Veterinary Medicine
- Music
- Normal Degree
- Nutrition
- Dentistry
- Pedagogy
- Psychology
- Chemistry
- Radio, TV and Internet
- Social Work
- Information Systems
- Tourism

Q9 What is your GPA?

____

Q10 Which campus do you study at?

- ◯ Juiz de Fora
- ◯ Governador Valadares

Next, I would like to ask you some questions about your future earnings expectations. In all of the following questions, the term *salary* refers to a **monthly** and **pre-tax salary**.

Q11 First consider the **private** sector jobs you would consider applying to after graduation, and think of the starting salary you could get.

What is the **lowest** starting salary you believe you could possibly be offered for a **private** sector job?

____

Q12 And what is the **highest** starting salary you believe you could possibly be offered for a **private** sector job?

____

Q13 Still thinking about the starting salary you would possibly be offered for a **private** sector job:

What do you think is the percent chance that your **private** sector starting salary offer will be:

	Use the slider to select a chance between 0% and 100%. 0 100
R$ [value 1] or **more** per month	
R$ [value 2] or **more** per month	
R$ [value 3] or **more** per month	
R$ [value 4] or **more** per month	

Q14 Now consider a job in the **public** sector obtained via a public-sector exam. Consider a job you would apply for, and think about the starting salary you could get after you graduate.

What is the **lowest** starting salary you believe you could possibly receive in the **public** sector?

____

Q15 And what is the **highest** starting salary you believe you could possibly receive in the **public** sector?

____

Q16 Still thinking about the starting salary you could possibly receive in the **public** sector after graduating:

What do you think is the percent chance that your starting salary in the **public** sector will be:

	Use the slider to select a chance between 0 and 100%. 0 100
R$ [value 1] or **more** per month	
R$ [value 2] or **more** per month	
R$ [value 3] or **more** per month	
R$ [value 4] or **more** per month	

The next questions relate to your salary expectations after 10 years of experience. In all of these questions, please ignore the effects of inflation, that is, consider that R$ 1 in 10 years has the same purchasing power as R$ 1 today. As in the previous questions, the term *salary* refers to a **monthly** and **pre-tax salary**.

Q17 Please think again about your salary expectations in a **private** sector job, and consider the salary you could receive after 10 years of experience.

What is the **lowest** salary you believe you could possibly get in a **private** sector job after 10 years of experience?

____

Q18 And what is the **highest** salary you believe you could possibly get in a **private** sector job after 10 years of experience?

____

Q19 Still thinking about the salary you could possibly get from a **private** sector job after 10 years of experience:

What do you think is the percent chance that your salary after 10 years of **private** sector experience will be:

	Use the slider to select a chance between 0% and 100%. 0 100
R$ [value 1] or **more** per month	
R$ [value 2] or **more** per month	
R$ [value 3] or **more** per month	
R$ [value 4] or **more** per month	

Q20 Finally, I would like to ask you one last time about your **public** sector salary expectations. Consider a job obtained via a public-sector exam and please reflect on the salary you could receive after 10 years of experience.

What is the **lowest** salary you believe you could possibly receive in the **public** sector after 10 years of experience?

____

Q21 And what is the **highest** salary you believe you could possibly receive in the **public** sector after 10 years of experience?

____

Q22 Still thinking about the salary you could possibly receive in the **public** sector after 10 years of experience:

What do you think is the percent chance that your salary after 10 years of experience in the **public** sector will be:

Use the slider to select a chance between 0% and 100%.

	0 — 100
R$ [value 1] or **more** per month	
R$ [value 2] or **more** per month	
R$ [value 3] or **more** per month	
R$ [value 4] or **more** per month	

Q23 What is the **lowest** salary you would be willing to accept for a job in the **private** sector after you graduate? Please answer in terms of a monthly and pre-tax salary.

____

Q24 What is the **lowest** salary you would be willing to accept for a job in the **public** sector, obtained via a public-sector exam, after you graduate? Please answer in terms of a monthly and pre-tax salary.

____

Q25 Imagine you have just started a job in the **private** sector. What do you think is the chance, in percentage terms, that you will lose your job in the first **3 years**?

Use the slider to select a chance between 0% e 100%

0 100

Q26 Now imagine that you have just started a job in the **public** sector. What do you think is the chance, in percentage terms, that you will lose your job in the first **3 years**?

Use the slider to select a chance between 0% e 100%

0 100

Q27 In case you decide to pursue a career in the public sector, what do you think is the chance, in percentage terms, that you will be approved in a public-sector exam for a job that is attractive to you?

Use the slider to select a chance between 0% e 100%

0 100

Q28 Imagine that you have signed-up for a public-sector exam in your field of study. Suppose this public-sector exam has 1,000 candidates for a total of 5 openings, and that you had 6 months to study for the tests.

Consider a measure that splits the candidates in 5 different groups according to the final ranking in the exam. For example, the group [201,400] refers to the group of candidates that were classified between the 201st e 400th positions. What do you think is the chance, in percentage terms, that you would be part of each of the groups?

For example, the value 30 in one of the groups means you believe there is a 30% chance that you belong to that group. A value of 0, on the other hand, means you believe there is a 0% chance you would belong to that group, while a value of 100 means you believe you'd part of that group with 100% chance. Use the fields below to indicate your chances of being part of each of the groups, and note that the sum of all fields should be equal to 100.

[1, 200]: ________
[201, 400]: ________
[401, 600]: ________
[601, 800]: ________
[801, 1000]: ________
Total: ________

Q29 Now suppose the public-sector exam has taken place, and that you were classified in the [randomly selected] group. Imagine you decide to take another exam with the same characteristics (1,000 candidates, 5 openings, your field of study, and 6 months to study), what do you think are the chances, in <u>percentage</u> terms, that you would belong to each of the groups? Use the fields below to indicate your chances of being part of each of the groups, and note that the sum of all fields should be equal to 100.

[1, 200]: ________
[201, 400]: ________
[401, 600]: ________
[601, 800]: ________
[801, 1000]: ________
Total: ________

We are approaching the end of the survey!
The next questions are about your personal preferences and characteristics.

Q30 How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? Please indicate your answer using a scale from 0 to 10, where 0 means "not at all willing" and 10 means "very willing".

0 = Not at all willing   10 = Very willing

0 1 2 3 4 5 6 7 8 9 10

Q31 In general, how willing are you to take risks? Please indicate your answer using a scale from 0 to 10, where 0 means “not at all willing” and 10 means “very willing”.

0 = Not at all willing   10 = Very willing

	0	1	2	3	4	5	6	7	8	9	10

Q32 Do you agree with the following statement? “Job stability is higher in the public sector than in the private sector”. Please indicate your answer using a scale from 0 to 10, where 0 means “totally disagree” and 10 means “totally agree”.

0 = Totally disagree   10 = Totally agree

	0	1	2	3	4	5	6	7	8	9	10

Q33 Do you agree with the following statement? “The benefits (retirement, health plan, work hours, etc.) are better in public-sector jobs than in private-sector jobs? Please indicate your answer using a scale from 0 to 10, where 0 means “totally disagree” and 10 means “totally agree”.

0 = Totally disagree   10 = Totally agree

	0	1	2	3	4	5	6	7	8	9	10
1 ()											

Q34 Will you graduate at the end of the year?

- ○ Yes
- ○ No

Q35 Have you ever taken a public-sector exam?

- Yes
- No

Q36 Approximately how many people do you known that either having been approved at, or are currently studying for, a public-sector exam?

- 1
- …
- 9
- 10 or more
- I don’t know anyone

Q37 Have you ever worked in the private sector?

- Yes
- No

Q38 Are you currently working?

- Yes, I have been approved in a public-sector exam
- Yes, I work in the private sector
- Yes, I have my own business
- No

Q39 [OPTIONAL] Please, describe in more detail your reasons to prefer to have your own business instead of finding a job in the private or public sectors.

____

____

____

Q40 [OPTIONAL] Please, describe in more detail your reasons to prefer to go to graduate school instead of looking for a job or opening you own business.

____

____

____

Q41 [OPTIONAL] Please, describe in more detail your reasons to prefer a job in the private sector instead of a job in the public sector.

____

____

____

Q42 [OPTIONAL] Please, describe in more detail your reasons to prefer a job in the public sector instead of a job in the private sector.

____

____

____

Q43 [OPTIONAL] Lastly, please describe in more detail what your decision process would be if you were considering jobs in both the public and private sectors. For example, what would be your strategy to decide between taking public-sector exams or applying for positions in the private sector?

____

____

____

A.2 Tables and Figures

Table A.1: Wage Regressions for Sample of Test-Takers

	Full Sample	High-School Graduates	College Graduates
Coefficient on public-sector	5.157***	4.873***	5.456***
	(0.031)	(0.023)	(0.003)
% Public-Sector	0.40	0.32	0.51
Mean Hourly Wage	20.18	13.36	27.93
Median Hourly Wage	14.63	10.77	21.94
R^2	0.125	0.129	0.071
N	1,532,248	685,786	672,546

Note: OLS regression on a 30% random sample of test-takers who held a formal job in 2017. Controls include years of education, experience, experience^2, gender, and state or region fixed-effects. Sample further divided into test-takers with a high-school degree or a college degree as of 2017. Mean hourly wages (BRL) is the dependent variable and is restricted to be at least 75% of the 2017 federal minimum and at most equal to the public-sector legal maximum (245 BRL), for public-sector workers, or two times the public-sector legal maximum for private-sector workers. Age is restricted to be between 18 and 55.

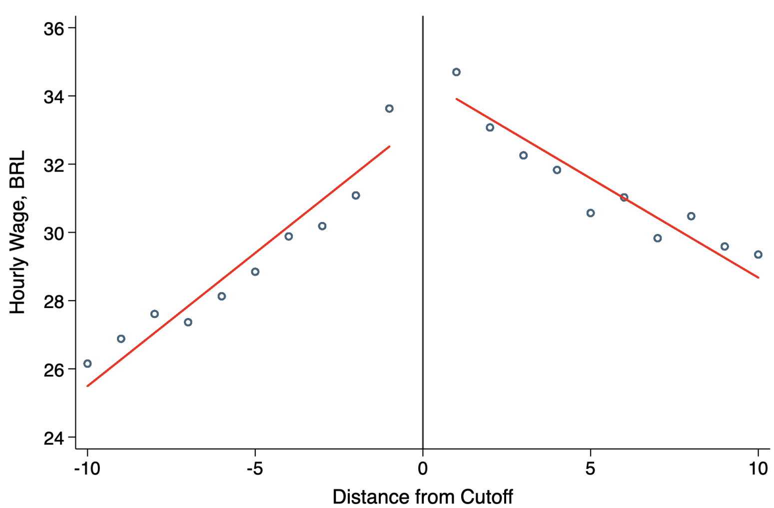
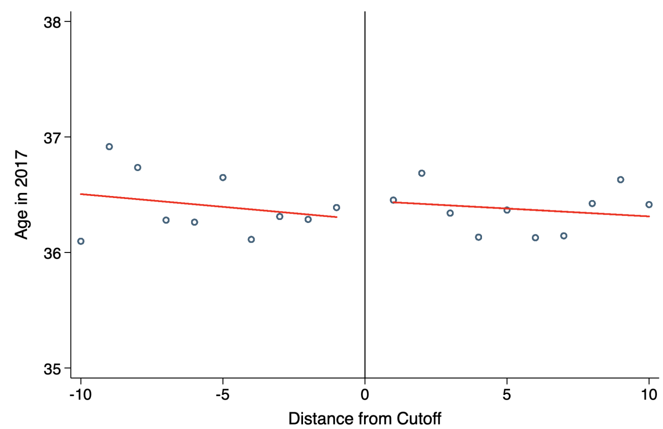
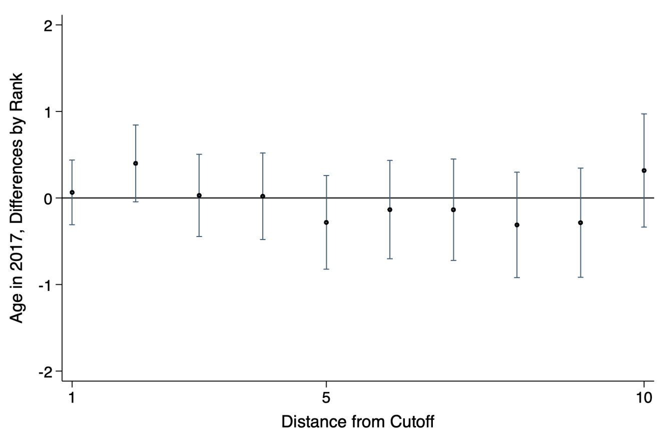


Figure A.1: Discontinuity in Hourly Wage

Note: Discontinuity on restricted sample: 1.196***(0.249), N=63,441. Linear regression with hourly wage as dependent variable. Robust standard errors clustered at the running variable.



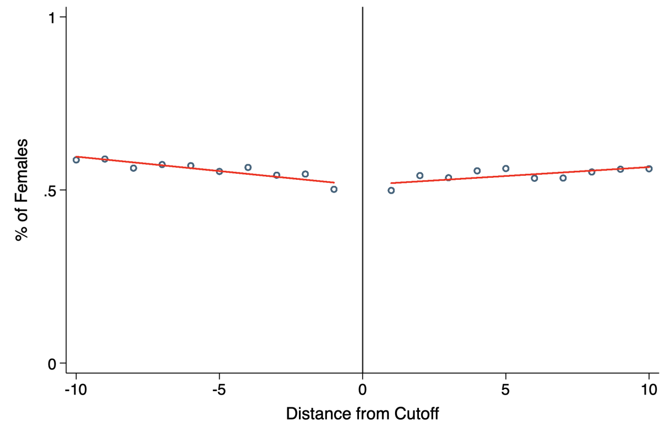
(a) Discontinuity in Age



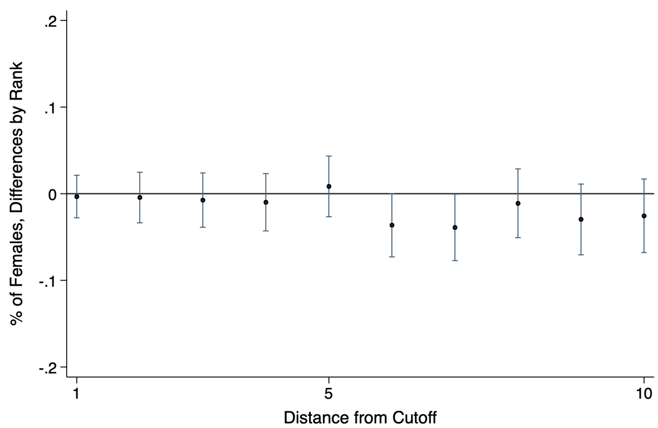
(b) Differences in Age, Ranks X and -X

Figure A.2: Discontinuity in Age

Note: Figure (A): Discontinuity on restricted sample: 0.163 (0.132), N=63,718. Linear regression with age in 2017 as dependent variable. Robust standard errors clustered at the running variable. Figure (B): Differences in age by values of running variable with opposite signs. Confidence intervals (95%) from regressions with age in 2017 as dependent variable and dummy indicating positive value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$



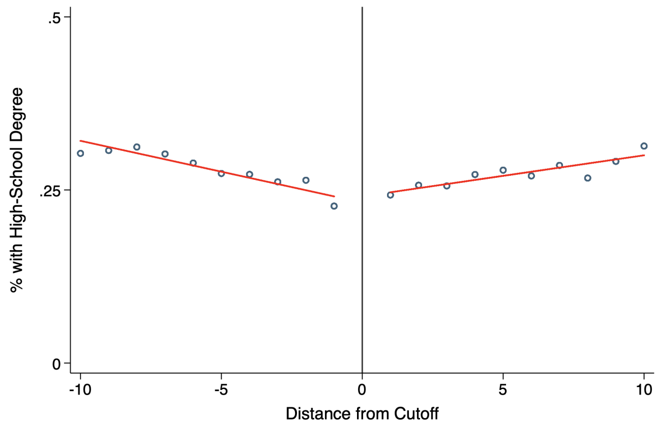
(a) Discontinuity in Gender Composition



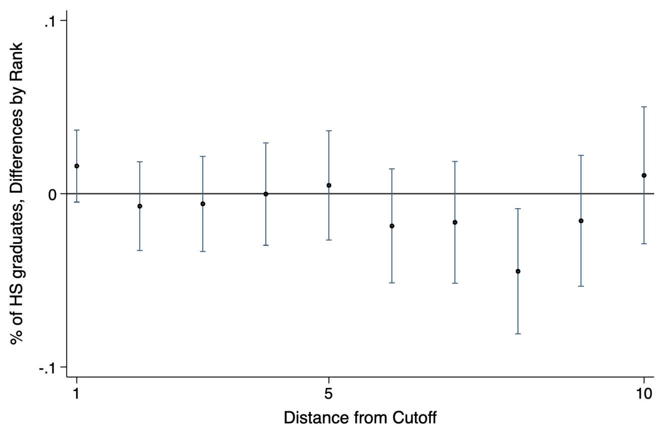
(b) Differences in Gender Composition, Ranks X and -X

Figure A.3: Discontinuity in Gender Composition

Note: Figure (A): Discontinuity on restricted sample: 0.002 (0.002), N=63,720. Linear regression with indicator for female as dependent variable. Robust standard errors clustered at the running variable. Figure (B): Differences in % of females by values of running variable with opposite signs. Confidence intervals (95%) from regressions with indicator for female as dependent variable and dummy indicating positive value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$



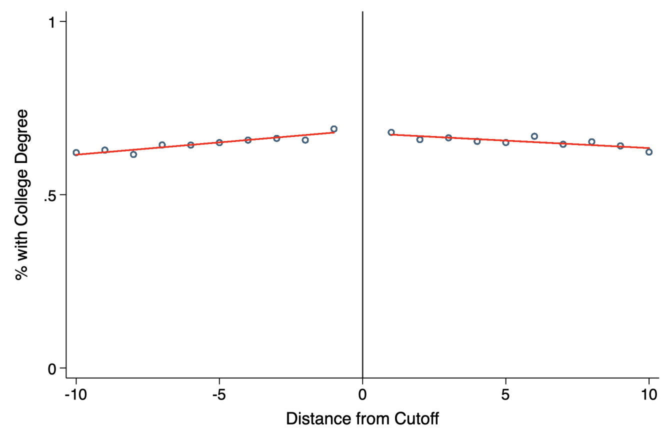
(a) Discontinuity in High-School Education



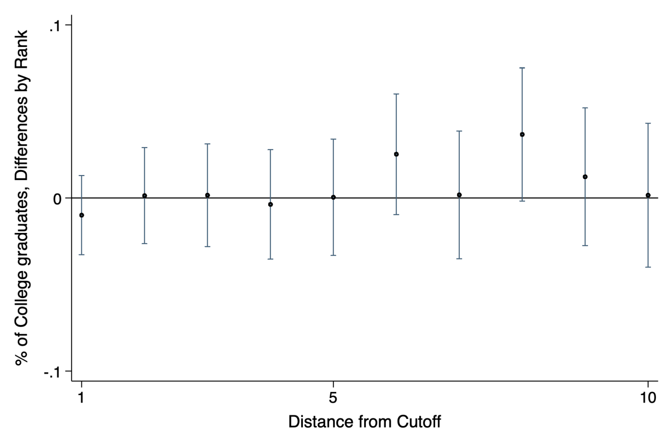
(b) Differences in High-School Education, Ranks X and -X

Figure A.4: Discontinuity in Share with High-School Degree

Note: Figure (A): Discontinuity on restricted sample: 0.009 (0.010), N=63,720. Linear regression with indicator for high-school education as dependent variable. Robust standard errors clustered at the running variable. Figure (B): Differences in % of high-school degree by values of running variable with opposite signs. Confidence intervals (95%) from regressions with indicator for high-school education as dependent variable and dummy indicating positive value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$



(a) Discontinuity in College Education



(b) Differences in College Education, Ranks X and -X

Figure A.5: Discontinuity in Share with College Degree

Note: Figure (A): Discontinuity on restricted sample: -0.008 (0.009), N=63,720. Linear regression with indicator for college education as dependent variable. Robust standard errors clustered at the running variable. Figure (B): Differences in % of college degree by values of running variable with opposite signs. Confidence intervals (95%) from regressions with indicator for college education as dependent variable and dummy indicating positive value of the running variable.
$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

Appendix B Appendix for Chapter 3

B.1 Supplementary Tables and Figures

B.1.1 Invariance to Quantitative Beliefs

Figure B.1 shows the effect of the cutoffs in round one for the second and third mover on the best-response cutoffs for the first mover in round two. In particular the figure indicates the difference in best response between the first and second round, so $51 - \mu_{1,2}(\mu_{2,1}, \mu_{2,3})$. For most second- and third-mover responses close to the equilibrium the best response is identical. Even with substantially different beliefs about the cutoffs, the first mover should still have a difference in their first and second cutoffs of at least 10. The sole exceptions are those cases where the second and third movers are believed to use boundary cutoffs, either always or never switching.

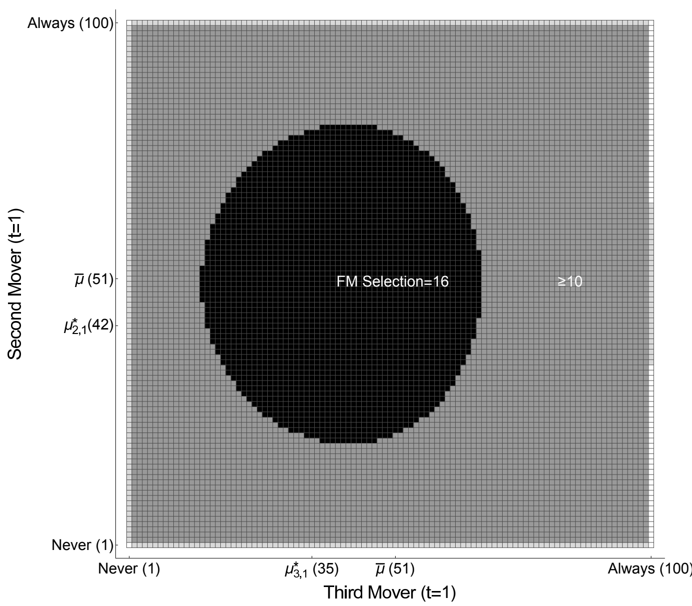


Figure B.1: Differences for First-Mover Cutoffs (Round 1 to 2) as a Function of Beliefs

B.1.2 Experimental Design

Overall allocation of subjects to session is given in Table B.1.

Table B.1: Experimental Design Table

Treatment	Type	Sessions	Subjects	Supergames
Selection	Game	4	66	1,386
No Selection	Decision	2	33	693
S-Across	Game	4	60	1,260
S-Within	Game	4	69	1,446
S-Explicit	Decision	2	36	756
S-Peer	Game	4	72	1,512
S-Simple-Random	Game	4	72	2,952
S-Simple-Fixed	Game	4	72	2,952
All		28	480	12,960

Note: Table presents number of sessions and subjects per treatment, which are classified by type (game versus decision problem).

B.1.3 Robustness

This appendix contains tables and figures that extend the analysis in the main text. Table B.2 reports the same regressions as in Table 1 in the main text, but using data from supergames 6 to 20 instead of supergames 11 to 20. Table B.3 extends the same analysis for players in the role of second- and third-movers, and finds very similar results. Table B.4 replicates the analysis from Table 1 for the *S-Across* treatment, and Table B.5 extends the analysis to include data from supergames 6 to 20. Tables B.6 and B.7 do the same for the *S-Explicit* treatment.

Table B.8 report results from the *S-Within* treatment, in which subjects are informed about the actions taken by other players during the course of the supergame. In the *S-Within* treatment, the relevant conditioning variable should be the information that other people switched, and the passing of time per se does not convey direct actionable information. Table

B.9 extends the analysis to include data from supergames 6 to 20.

Tables B.10 and B.11 report results of supergame 21 behavior for subjects in the *S-Within* and *S-Peer* treatments, respectively. For both cases, we cannot reject the hypothesis that cutoffs are similar to the *Selection* treatment. We conclude that neither receiving strategic feedback during the path of play, nor discussing the optimal strategy with peers significantly affects behavior in supergame 21.

Table B.2: Average Cutoff for *No Selection* and *Selection* Treatments, First-Movers

Treatment	Game Round	Theory	Supergame 6 to 20			Supergame 21		
			Estimate	Test (p-Value)		Estimate	Test (p-Value)	
		$\mu^\star$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \mu_t^{\star j}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \mu_t^{\star j}$
No Selection (Control)	Round 1, $\hat{\mu}_1^{NS}$	[51]	54.13 (2.40)	–	0.192	52.76 (2.48)	–	0.479
	Round 2, $\hat{\mu}_2^{NS}$	[51]	53.53 (2.42)	0.249	0.295	52.88 (2.48)	0.907	0.449
	Round 3, $\hat{\mu}_3^{NS}$	[51]	53.01 (2.46)	0.100	0.414	53.27 (2.48)	0.620	0.360
	Joint Tests:		0.197‡		0.201§	0.875‡		0.817§
Selection (Treatment)	Round 1, $\hat{\mu}_1^S$	[51]	46.92 (1.20)	**0.007**	**0.001**	43.45 (3.04)	**0.018**	**0.013**
	Round 2, $\hat{\mu}_2^S$	[35]	43.28 (1.23)	**0.000**	**0.000**	39.91 (3.04)	**0.001**	0.107
	Round 3, $\hat{\mu}_3^S$	[28]	38.92 (1.30)	**0.000**	**0.000**	36.32 (3.04)	**0.000**	**0.006**
	Joint Tests:		**0.000**‡		**0.000**§	**0.000**‡		**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 171/138/33 Total/*Selection*/*No Selection* first-mover subjects across supergames 6-20, and 55/22/33 in supergame 21. *Selection* treatment exclude subjects in the second- and third-mover roles. †–Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t) and the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Table B.3: Average Cutoff per Round for *Selection* Treatments, Second- and Third-Movers

Treatment	**Game Round**	**Theory**	**Supergame 11 to 20**			**Supergame 21**		
			Estimate	**Test (p-Value)**		**Estimate**	**Test (p-Value)**	
		$\mu^\star$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \mu_t^{\star j}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \mu_t^{\star j}$
Selection, Second-Movers	Round 1, $\hat{\mu}_1^S$	[42]	44.64 (0.68)	**0.000**	**0.000**	46.09 (2.80)	0.066	0.145
	Round 2, $\hat{\mu}_2^S$	[31]	41.32 (0.98)	**0.000**	**0.000**	40.36 (2.80)	**0.001**	**0.001**
	Round 3, $\hat{\mu}_3^S$	[25]	38.00 (1.40)	**0.000**	**0.000**	35.18 (2.80)	**0.000**	**0.000**
	Joint Tests:		**0.000**‡		**0.000**§	**0.024**‡		**0.000**§
Selection, Third-Movers	Round 1, $\hat{\mu}_1^S$	[35]	43.79 (1.31)	**0.000**	**0.000**	45.50 (2.57)	**0.030**	**0.000**
	Round 2, $\hat{\mu}_2^S$	[28]	40.57 (1.34)	**0.000**	**0.000**	41.95 (2.57)	**0.001**	**0.000**
	Round 3, $\hat{\mu}_3^S$	[23]	37.63 (1.40)	**0.000**	**0.000**	36.82 (2.57)	**0.000**	**0.000**
	Joint Tests:		**0.000**‡		**0.000**§	**0.059**‡		**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 133/138 second-/third-mover subjects across supergames 11-20, and 22 second- and third-movers in supergame 21. †−Univariate significance tests columns examine differences from either the first-round coefficient from the control (H_0 :$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t) and the theoretical prediction (H_0 :$\hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Table B.4: Average Cutoff for *S-Across* Treatment, First-Movers (Supergames 11 to 20)

Treatment	**Game Round**	**Theory**	**Supergame 11 to 20**				**Supergame 21**			
			Estimate	**Test (*p*-Value)**			**Estimate**	**Test (*p*-Value)**		
		$\mu^\star$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\mu_t^j = \mu_t^{\star j}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\hat{\mu}_t^j = \mu_t^{\star j}$
S-Across	Round 1, $\hat{\mu}_1^{S(A)}$	[51]	49.26	0.082	0.255	0.375	51.80	0.807	0.068	0.809
			(1.93)				(3.31)			
	Round 2, $\hat{\mu}_2^{S(A)}$	[35]	45.37	**0.034**	0.320	**0.000**	49.45	0.458	**0.037**	**0.000**
			(2.00)				(3.31)			
	Round 3, $\hat{\mu}_3^{S(A)}$	[28]	41.30	**0.000**	0.390	**0.000**	45.35	**0.097**	**0.048**	**0.000**
			(2.11)				(3.31)			
	Joint Tests:		**0.000**‡			**0.000**§	**0.001**‡			**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 60 first-mover subjects across supergames 11-20, and 20 first-movers in supergame 21. †−Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t), the first-mover coefficients from the *Selection* treatment ($H_0 : \hat{\mu}_t^j = \hat{\mu}_t^S$ for treatment j, round t) and the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Table B.5: Average Cutoff for *S-Across* Treatment, First-Movers (Supergames 6 to 20)

Treatment	**Game Round**	**Theory**	**Supergame 6 to 20**				**Supergame 21**			
			Estimate	**Test (p-Value)**			**Estimate**	**Test (p-Value)**		
		$\mu^{\star}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\mu_t^j = \mu_t^{\star j}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\hat{\mu}_t^j = \mu_t^{\star j}$
S-Across	Round 1, $\hat{\mu}_1^{S(A)}$	[51]	49.59	0.150	0.239	0.459	51.80	0.807	0.068	0.809
			(1.90)				(3.31)			
	Round 2, $\hat{\mu}_2^{S(A)}$	[35]	46.40	**0.015**	0.182	**0.000**	49.45	0.458	**0.037**	**0.000**
			(1.94)				(3.31)			
	Round 3, $\hat{\mu}_3^{S(A)}$	[28]	41.66	**0.000**	0.261	**0.000**	45.35	**0.097**	**0.048**	**0.000**
			(2.02)				(3.31)			
	Joint Tests:		**0.000**‡			**0.000**§	**0.001**‡			**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 60 first-mover subjects across supergames 6-20, and 20 first-movers in supergame 21. †−Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t), the first-mover coefficients from the *Selection* treatment ($H_0 : \hat{\mu}_t^j = \hat{\mu}_t^S$ for treatment j, round t) and the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Table B.6: Average Cutoff per Round for *S-Explicit* Treatment (Supergames 11 to 20)

Treatment	**Game Round**	**Theory**	**Supergame 11 to 20**				**Supergame 21**			
			Estimate	**Test (*p*-Value)**			**Estimate**	**Test (*p*-Value)**		
		$\mu^\star$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\mu_t^j = \mu_t^{\star j}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\hat{\mu}_t^j = \mu_t^{\star j}$
S-Explicit	Round 1, $\hat{\mu}_1^{S(E)}$	[51]	55.52	0.844	**0.001**	0.067	59.58	0.199	**0.004**	0.059
			(2.47)				(4.55)			
	Round 2, $\hat{\mu}_2^{S(E)}$	[34]	48.46	0.076	**0.054**	**0.000**	53.58	0.877	**0.016**	**0.000**
			(2.49)				(4.55)			
	Round 3, $\hat{\mu}_3^{S(E)}$	[17]	42.57	**0.000**	0.238	**0.000**	50.08	0.615	**0.015**	**0.000**
			(2.55)				(4.55)			
	Joint Tests:		**0.000**‡			**0.000**§	**0.000**‡			**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 36 subjects in the *S-Explicit* treatment across supergames 11-20, and 12 first-movers in supergame 21. †−Univariate significance tests columns examine differences from either the first-round coefficient from the control (H_0 :$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t), the first-mover coefficients from the *Selection* treatment (H_0 :$\hat{\mu}_t^j = \hat{\mu}_t^S$ for treatment j, round t) and the theoretical prediction (H_0 :$\hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Table B.7: Average Cutoff per Round for *S-Explicit* Treatment (Supergames 6 to 20)

Treatment	**Game Round**	**Theory**	**Supergame 6 to 20**				**Supergame 21**			
			Estimate	**Test (*p*-Value)**			**Estimate**	**Test (*p*-Value)**		
		$\mu^{\star}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\mu_t^j = \mu_t^{\star j}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\hat{\mu}_t^j = \mu_t^{\star j}$
S-Explicit	Round 1, $\hat{\mu}_1^{S(E)}$	[51]	55.49 (2.42)	0.698	**0.002**	0.063	59.58 (4.55)	0.199	**0.004**	0.059
	Round 2, $\hat{\mu}_2^{S(E)}$	[34]	48.88 (2.44)	0.134	**0.043**	**0.000**	53.58 (4.55)	0.877	**0.016**	**0.000**
	Round 3, $\hat{\mu}_3^{S(E)}$	[17]	43.84 (2.48)	**0.004**	0.083	**0.000**	50.08 (4.55)	0.615	**0.015**	**0.000**
	Joint Tests:		**0.000**‡			**0.000**§	**0.000**‡			**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 36 subjects in the *S-Explicit* treatment across supergames 6-20, and 12 first-movers in supergame 21. †–Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t), the first-mover coefficients from the *Selection* treatment ($H_0 : \hat{\mu}_t^j = \hat{\mu}_t^S$ for treatment j, round t) and the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Table B.8: Average Cutoff per In-Group Switches, *S-Within* (Supergames 11 to 20)

Treatment	In-Group Switches	Theory	Supergame 11 to 20			
			Estimates			
		$\mu^{\star}$	Round 1, $\hat{\mu}_{1s}^{S(W)}$	Round 2, $\hat{\mu}_{2s}^{S(W)}$	Round 3, $\hat{\mu}_{3s}^{S(W)}$	**Joint Tests:**
S-Within	Zero	[51]	42.91 [**0.000**]† (1.80)	41.51 [**0.000**]† (1.87)	39.35 [**0.000**]† (2.03)	**0.001**‡
	One	[14]	35.66 [**0.000**]† (1.88)	34.44 [**0.000**]† (1.87)	34.09 [**0.000**]† (1.95)	0.206‡
	Two	[3]	30.77 [**0.000**]† (2.76)	29.95 [**0.000**]† (2.25)	31.18 [**0.000**]† (2.26)	0.813‡
	Joint Tests:		**0.000**‡	**0.000**‡	**0.000**‡	

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 69 subjects across supergames 11-20. †−P-Value of univariate significance tests of the theoretical predictions (H_0 :$\hat{\mu}_{ts}^{j} = \mu_{ts}^{\star j}$). ‡–Join test: Cutoffs are stationary across the treatment (H_0 : $\hat{\mu}_{1s}^{S(W)} = \hat{\mu}_{2s}^{S(W)} = \hat{\mu}_{3s}^{S(W)}$ across rounds, and H_0 : $\hat{\mu}_{t1}^{S(W)} = \hat{\mu}_{t2}^{S(W)} = \hat{\mu}_{t3}^{S(W)}$ across number of in-group switches).

Table B.9: Average Cutoff per In-Group Switches, *S-Within* (Supergames 6 to 20)

Treatment	**In-Group Switches**	**Theory**	**Supergame 6 to 20**			
			Estimates			
		$\mu^\star$	Round 1, $\hat{\mu}_{1s}^{S(W)}$	Round 2, $\hat{\mu}_{2s}^{S(W)}$	Round 3, $\hat{\mu}_{3s}^{S(W)}$	**Joint Tests:**
S-Within	Zero	[51]	43.42 [**0.000**]† (1.68)	42.15 [**0.000**]† (1.72)	40.93 [**0.000**]† (1.86)	**0.006**‡
	One	[14]	37.04 [**0.000**]† (1.73)	37.04 [**0.000**]† (1.72)	35.60 [**0.000**]† (1.79)	0.192‡
	Two	[3]	34.67 [**0.000**]† (2.45)	31.26 [**0.000**]† (2.05)	32.53 [**0.000**]† (2.06)	0.291‡
	Joint Tests:		**0.000**‡	**0.000**‡	**0.000**‡	

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 69 subjects across supergames 6-20. †−P-Value of univariate significance tests of the theoretical predictions ($H_0 : \hat{\mu}_{ts}^{j} = \mu_{ts}^{\star j}$). ‡–Join test: Cutoffs are stationary across the treatment ($H_0 : \hat{\mu}_{1s}^{S(W)} = \hat{\mu}_{2s}^{S(W)} = \hat{\mu}_{3s}^{S(W)}$ across rounds, and $H_0 : \hat{\mu}_{t1}^{S(W)} = \hat{\mu}_{t2}^{S(W)} = \hat{\mu}_{t3}^{S(W)}$ across number of in-group switches).

Table B.10: Average Cutoff in *S-Within* Treatment, First-Movers (Supergame 21)

Treatment	Game Round	Theory	Supergame 21			
			Estimate	**Test (*p*-Value)**		
		$\mu^{\star}$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\hat{\mu}_t^j = \mu_t^{\star j}$
S-Within	Round 1, $\hat{\mu}_1^{S(W)}$	[51]	39.57 (3.32)	**0.001**	0.413	**0.001**
	Round 2, $\hat{\mu}_2^{S(W)}$	[35]	37.00 (3.32)	**0.004**	0.540	0.547
	Round 3, $\hat{\mu}_3^{S(W)}$	[28]	34.09 (3.32)	**0.001**	0.639	0.067
	Joint Tests:		**0.005**‡			**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 23 first-mover subjects in supergame 21. †−Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t), the first-mover coefficients from the *Selection* treatment ($H_0 : \hat{\mu}_t^j = \hat{\mu}_t^S$ for treatment j, round t) and the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

Table B.11: Average Cutoff in *S-Peer* Treatment, First-Movers (Supergame 21)

Treatment	Game Round	Theory	Supergame 21			
			Estimate	**Test (p-Value)**		
		$\mu^\star$	$\hat{\mu}$	$\hat{\mu}_t^j = \hat{\mu}_1^{NS}$	$\hat{\mu}_t^j = \hat{\mu}_t^S$	$\hat{\mu}_t^j = \mu_t^{\star j}$
S-Peer	Round 1, $\hat{\mu}_1^{S(A)}$	[51]	46.13 (3.31)	0.187	0.577	0.140
	Round 2, $\hat{\mu}_2^{S(A)}$	[35]	41.5 (3.31)	**0.035**	0.739	**0.049**
	Round 3, $\hat{\mu}_3^{S(A)}$	[28]	37.17 (3.31)	**0.005**	0.859	**0.001**
	Joint Tests:		**0.000**‡			**0.000**§

Note: Figures derived from a single random-effects least-squares regression for all chosen cutoffs against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). There are 24 first-mover subjects in supergame 21. †−Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t), the first-mover coefficients from the *Selection* treatment ($H_0 : \hat{\mu}_t^j = \hat{\mu}_t^S$ for treatment j, round t) and the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

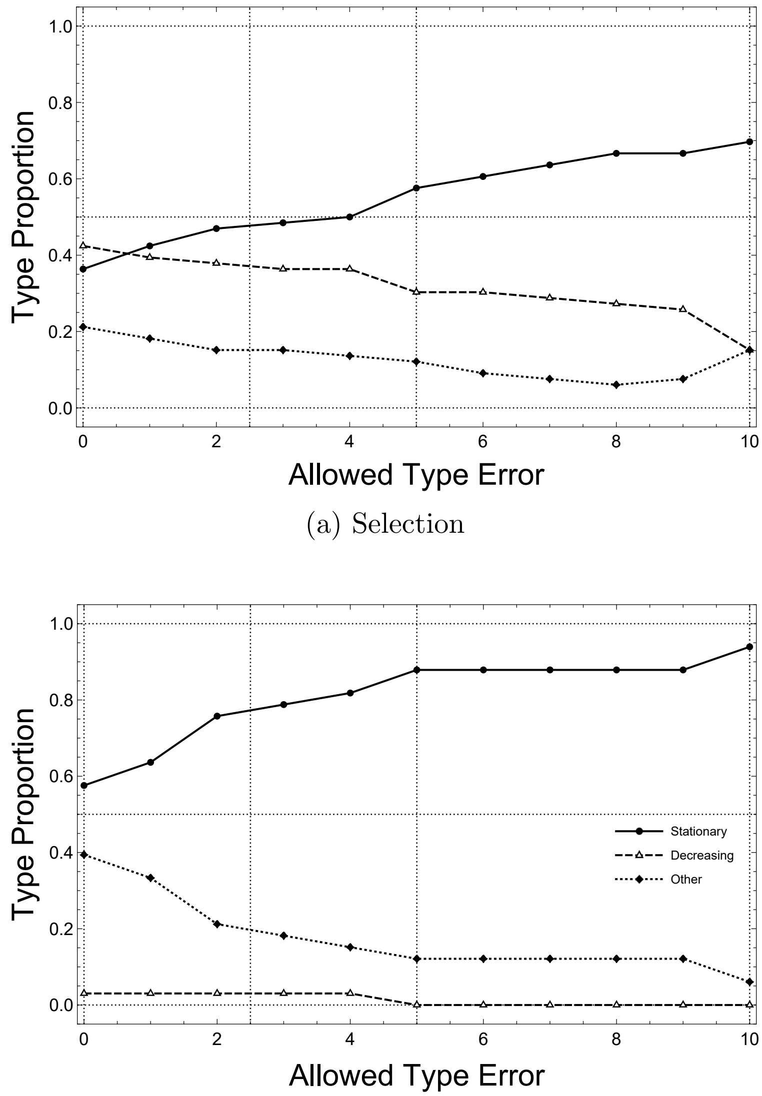
Table B.12: Average Cutoff in Last Round for *S-Simple* Treatments (Supergame 40)

Treatment	**Game Round**	**Theory**	**S-Simple-Random**			**S-Simple-Fixed**		
			Estimate	***p*-values**		**Estimate**	***p*-values**	
		$\mu^\star$	$\hat{\mu}$	$H_0:\hat{\mu}_j^W=\hat{\mu}_1^W$	$H_0:\hat{\mu}_t^j=\mu_t^{\star j}$	$\hat{\mu}$	$H_0:\hat{\mu}_t^j=\hat{\mu}_1^{NS}$	$H_0:\hat{\mu}_t^j=\mu_t^{\star j}$
S-Simple	Round 1, $\hat{\mu}_1^{NS}$	[51]	44.88 (2.85)	–	**0.033**	50.20 (2.85)	–	0.770
	Round 2, $\hat{\mu}_2^{NS}$	[32]	41.62 (2.85)	0.421	**0.001**	43.00 (2.85)	**0.077**	**0.000**
	Round 3, $\hat{\mu}_3^{NS}$	[22]	38.00 (2.85)	**0.090**	**0.000**	30.67 (2.85)	**0.000**	**0.003**
	Joint Tests:		**0.000**‡		**0.000**§	**0.000**‡		**0.000**§

Note: Figures derived from a single OLS regression ($N = 144$ subjects each making a single choice) against treatment-round dummies. Standard errors in parentheses, risk-neutral predicted cutoffs in square brackets (switch ball if value is lower than cutoff). †−Univariate significance tests columns examine differences from either the first-round coefficient from the control ($H_0 : \hat{\mu}_t^j = \hat{\mu}_1^{NS}$ for treatment j, round t), the first-mover coefficients from the *Selection* treatment ($H_0 : \hat{\mu}_t^j = \hat{\mu}_t^S$ for treatment j, round t) and the theoretical prediction ($H_0 : \hat{\mu}_t^j = \mu_t^{\star j}$). ‡–Joint test of stationary cutoffs across the supergame ($H_0 : \hat{\mu}_1^j = \hat{\mu}_2^j = \hat{\mu}_3^j$ for treatment j); §–Joint test of PBE cutoffs in supergame ($H_0 : 0 = \hat{\mu}_1^j - \mu_1^{\star j} = \hat{\mu}_2^j - \mu_2^{\star j} = \hat{\mu}_3^j - \mu_3^{\star j}$).

B.1.4 Type Classification Robustness

In Figure B.2 we indicate the three type categories as we vary the bandwidth parameter ϵ from 0 to 10.



(a) Selection

(b) No Selection

Figure B.2: Type Robustness to ϵ

Table B.13 reports on the proportion of types as classified in the last five partial strategy supergames for each treatment. Rather than the full strategy method final supergames, we here take averages across supergames to assess the strategy cutoffs.

Following the paper we focus on the type specifications with an error bandwidth $\epsilon = 2.5$

(though we also provide data on $\epsilon = 0$ and $\epsilon = 5$).

For treatment *S-Simple-Random* a subject is *decreasing* if the difference in average cutoffs for *all* pairwise comparisons (first minus second-mover, first minus third-mover, and second minus third-mover) are strictly positive; likewise, subject is ϵ-*decreasing* if all such differences are (weakly) greater than ϵ. For treatment *S-Within* subject is *decreasing* if the difference in average cutoffs for *all* pairwise comparisons (no switches minus 1 switch, no switches minus 2 switches, and 1 switch minus 2 switches) are strictly positive; and subject is ϵ-*decreasing* if all such differences are (weakly) greater than ϵ. For all other treatments subject is *decreasing* if minimum of (round 1 - round 2) cutoffs, irrespective of type, is strictly positive, and is ϵ-*decreasing* if minimum is larger than ϵ. Treatment *Selection* includes data from *S-Deliberation* as both treatments are identical up until cycle 21.

Table B.13: Type Classifications on Last Five Cycles

	N_S	**Decreasing**			**Non-Decreasing**		
		Exact	$\epsilon = 2.5$	$\epsilon = 5$	Exact	$\epsilon = 2.5$	$\epsilon = 5$
No Selection	33	12.1%	**3.0%**	3.0%	87.9%	**97.0%**	97.0%
Selection	138	29.7%	**25.4%**	23.9%	70.3%	**64.6%**	66.1%
S-Across	60	21.7%	**20.0%**	18.3%	78.3%	**80.0%**	81.7%
S-Explicit	36	50.0%	**36.1%**	36.1%	50.0%	**63.9%**	63.9%
S-Within	69	65.6%	**49.2%**	44.3%	34.4%	**50.8%**	55.7%
S-Simple-Random	72	47.9%	**39.4%**	33.8%	52.1%	**60.6%**	66.2%

Note: Type classification based on cutoffs chosen in cycles 36-40 for *S-Simple-Random* and cycles 16-20 for all other treatments.

B.1.5 Effects from Experience and Risk Aversion

Table B.14: Individual Regressions

Explanatory Variables		**(I)** **Sel. Only**		**(II)** **Sel.+NoSel.**		**(III)** **Sel.+NoSel.**	
Experienced Rematching	$\hat{\nu}^{G}_{1,10}$	0.037		0.063			
	$\hat{\nu}^{G}_{11,20}$	0.055		0.051			
Experienced Outcomes	$\hat{\nu}^{H}_{1,10}$	**0.363**	$\star\star$ 0.030	**0.277**	$\star\star$ 0.049	**0.341**	$\star\star\star$ 0.003
	$\hat{\nu}^{H}_{11,20}$	0.015		0.018			
Risk Aversion	$\hat{\rho}$	**-2.24**	$\star\star$ 0.046	**-2.12**	$\star\star$ 0.021	**-2.11**	$\star\star$ 0.017
No Selection	δ_{NoSel}	–		5.52		**7.23**	$\star\star$ 0.020
Constant		21.6		25.8		26.9	
N		66		91		91	
$\overline{R}^2$		0.141		0.198		0.214	

Note: Statistically significant variables indicated by stars ($\star$–10%;$\star\star$–5%; $\star\star\star$–1%) with p-values beneath them. $\overline{R}^2$ is the adjusted R-squared measure of fit.

For each ϵ-stationary subject we generate their experienced rematching ($\hat{\nu}^{G}_{i}$) and final ($\hat{\nu}^{H}_{i}$) outcomes, averaging across supergames 1–10 and supergames 11–20. Despite facing the same generating process in each treatment there is substantial subject variation in rematching and final outcomes in these blocks of ten supergames. For each individual subject variation in the average experiences can be used to validate the mechanic in the hypothesized learning model. In Table B.14, we regress each stationary subject's supergame-21 cutoff choice on the observed experiences ($\hat{\nu}^{G}_{i}$ and $\hat{\nu}^{H}_{i}$) and elicited risk preference ($\hat{\rho}_i$). In the first regression (Column I), we present results for the 66 stationary-type subjects in our *Selection* treatments (the baseline and *S-Across*). In the second (Column II) we pool in the additional 25 stationary subjects from the *No Selection* treatment (including a treatment dummy). In the third regression, we remove the experienced rematching (both) and late-session experienced final outcomes as independent variables (which can be motivated by the adjusted R-squared measures provided in the last row).

Across specifications we find that variation in final *outcomes* in the early-session supergames, $\hat{\nu}^{H}_{1,10}$, has a strong effect in the predicted direction. However, subjects' learning seems to occur quite quickly, as we do not find a significant final cutoff response to final-outcome variation in the second block of ten supergames. While the signs on both rematching variables and the late-session final-outcome variable are positive, the size of the estimated effects are far smaller than for the early-session final outcomes. Separate from experience, subject-variation in elicited risk preferences is a significant predictor for subjects' final supergame cutoffs in all of the regressions, reflecting the other requirement for the aggregate match made by the behavioral model.

B.2 Theoretical Model

We model a dynamic setting with adverse selection occurring over time. To do this we set up a finite population of objects $\mathcal{O} = \{o_1, \ldots, o_M\}$ (with the interpretation of long-side participants or durable goods, etc.). Each object has a common value, an iid draw over $V \subset \mathbb{R}$ according to a commonly known distribution F so that the average value is $\bar{v}$. These objects are initially matched to a group of individuals $\mathcal{I} = \{1, \ldots, N\}$ (with the interpretation of short-side participants, consumers, etc.). Each individual is randomly assigned to one of the objects at the beginning of the game through a one-to-one matching function, so that individual i has the initial object μ_0^i. In our setting there are more objects than individuals ($N < M$) and so some of the objects $\hat{\mathcal{O}}_0 \subset \mathcal{O}$ are initially unassigned. The set of unassigned objects will function here as a rematching population, and it is adverse selection over this population that is our main focus.

Choices take place over a time variable $t \in \{1, 2, \ldots, T \cdot N\}$, where individuals take turns receiving an opportunity to see their object's value. In particular each individual sees their object's value in period $t_i^\star = i + \tau_i N$, where τ_i is an iid random variable from 1 to T.[1] At the exogenously determined time τ_i the individual i faces a choice: keep the initially assigned object μ_0^i which has a known value v_0, or instead rematch to an object μ_t^i from the unmatched population $\hat{\mathcal{O}}_t$. Importantly, rather than the rematching population being some fixed outcome or a draw from a stationary distribution, the rematching population after agent i's choice is given by $\hat{\mathcal{O}}_{t+1} = \{\mu_0^i\} \cup \hat{\mathcal{O}}_t \setminus \{\mu_t^i\}$.

To see that there is adverse-selection in our environment it suffices to solve the first few periods of the model. Define the event that the participant who moves in period t observes their value that period as $\mathcal{I}_t$ and the event that they observe their value and switch as $\mathcal{S}_t$, with complementary event $\overline{\mathcal{S}}_t$.

$t = 1$ Suppose individual 1 observes their object's value in period 1 (the event $\mathcal{I}_1 = \{t_1^\star = 1\}$). Because no-one else can have moved yet, each object in the rematching pool is an iid draw from the distribution $G_1(v\,|\mathcal{I}_1) = F(v)$. The optimal choice is therefore to rematch

[1] The interpretation given is that there are T periods, but agents move in turns within the period, so 1 is the first-mover, 2 the second-mover, etc. An alternative interpretation is that draws where two agents choose at the same time are resolved in preference to the first-mover, then the second-mover, and so on.

if $\mu_0^1 < \bar{v} =: v_1^\star$, which happens with probability $p_1^\star = \Pr\{t_1^\star = 1\} \cdot \Pr\{\mu_0^1 < v_1^\star\}$.

$t = 2$ Suppose 2 observes their object's value. A draw from the rematching pool is therefore distributed as $G_2(v\,|\mathcal{I}_2) = G_2(v) = \lambda \Pr\{\mathcal{S}_1\,|\mathcal{I}_2\} \cdot F(v\,|v < v_1^\star) + (1 - \lambda p_1^\star) \cdot G_1(v)$ where $\lambda = \frac{1}{M-N}$ is the probability of drawing any particular object from the rematching pool. The distribution $G_2(v)$ therefore has an expected value $v_2^\star < v_1^\star$, and the optimal choice by 2 is therefore to rematch if $\mu_0^2 < v_2^\star$, which happens with probability $p_2^\star = \Pr\{t_2^\star = 2\} \cdot \Pr\{\mu_0^2 < v_2^\star\}$

$\vdots$

t Were i to see their value at t, the distribution of the rematching population is $G_t(v\,|\mathcal{I}_t) = \lambda p_{t-1}^\star \cdot F(v\,\middle|v < v_{t-1}^\star) + (1 - \lambda p_{t-1}^\star) \cdot G_{t-1}(v\,\middle|\overline{\mathcal{S}}_{t-1}, \mathcal{I}_t)$ which has expectation $v_i^\star$. Individual i therefore switches with probability $p_t^\star = \Pr\{\mathcal{I}_t\} \cdot \Pr\{\mu_0^i < v_t^\star\}$.

The optimal solution proceeds inductively, with the additional complication that after period N agents must condition their choices on their own information, and that of other agents.[2]

Define the two sets of time periods

$$\mathcal{X}_t(s, s') = \{s, s+1, \dots s'-1, s'\} \cap \{r\,|J(r) = J(t)\},$$
$$\mathcal{Y}_t(s, s') = \{s, s+1, \dots s'-1, s'\} \cap \{r\,|J(r) \neq J(t)\},$$

the time periods between s and s' where the player who moves at t is or is not the relevant decision maker, respectively. The optimal decision rules according to :

$$v_t^\star := \sum_{s \in \mathcal{Y}_t(1,t-1)} \frac{q_s \cdot F(v_s^\star) \cdot \prod_{r \in \mathcal{Y}_t(s+1,t-1)} (1 - q_r \cdot F(v_r^\star))}{1 - \sum_{r \in \mathcal{X}_s(s+1,t-1)} q_r F(v_r^\star)} \mathbb{E}\,[v\,|v < v_s^\star]$$

This forms a decreasing sequence $\{v_t^\star\}_{t=jN+1}^{(j+1)N}$ across each sequence of player turns $j \in \{1, \dots, T\}$, while for any individual i the optimal decision rule $\left\{v_{jN+i}^\star\right\}_{j=1}^{T}$ is decreasing.

That the first sequence of turns has a decreasing optimal decision rule is illustrated above, as each participant faces a rematching pool that is a convex combination of the previous participant's distribution F_{t-1}, but with a positive probability of $F(v\,|v < \mathbb{E}\nu_{t-1})$ mixed in. For the first result we need to nest information. Define the event that a participant

[2]For instance, agent i observing the value of μ_o^1 in period $N+i$ knows that they did not see their value in period i. Similarly, the inductive step must incorporate the nested condition that player $i-1$ did not switch when working backwards.

i observes the object at time $t = j \cdot N + i$ and switches as $\mathcal{S}_t$ and its complement of not switching as $\mathcal{N}_t$. Essentially this event encodes information that j did not switch in periods $(j-1)N+i$, $(j-2)N+i$, etc. The rematching distribution of a participant who is thinking about switching at time t is given by:

$$G_t(v) = \lambda \cdot \Pr\{\mathrm{S}_{t-1}\} \cdot F(v \,|v < v_1^\star) + (1 - \lambda \cdot \Pr\{\mathrm{S}_t\}) \cdot G_{t-1}(v \,|\mathcal{N}_{t-1}).$$

For the recursive calculation, the event $\mathcal{N}_{jN+i}$ only contains information that is relevant to the probability that person i saw their object in a previous period, and so for any fixed sequence of turns $\{v_t^\star\}_{t=jN+1}^{(j+1)N}$ the same reasoning as before holds, given any distribution at the start of the sequence $G_{jN}(v)$.

B.3 Instructions

B.3.1 Instructions to Part 1 [Supergames 1-5]

B.3.1.1 No Selection (Control) and S-Explicit

Introduction[3]

Thank you for participating in our study. Please turn off mobile phones and other electronic devices. These must remain turned off for the duration of the session.

This is an experiment on decision making. The money you earn will depend on both your decisions and chance. The session will be conducted only through your computer terminal; please do not talk to or attempt to communicate with any other participants during the experiment. If you have a question during this instruction phase please raise your hand and one of the experimenters will come to where you are sitting to answer your question in private.

During the experiment, you will have the opportunity to earn a considerable amount of money depending on your decisions. At the end of the experiment, you will be paid in private and in cash. On top of what you earn through your decisions during the experiment, you will also receive a $6 participation fee.

Outline

Your interactions in this experiment will be divided into "Cycles".

- In each cycle you will be holding one of four balls, called Balls A to D.
- Each ball has a value between 1 and 100, and your payoff in each cycle will be determined by the value of the ball you are holding at the cycle's end.
- Initially you will not know any of the four ball's values, and will only know which of the four balls you are holding. Each cycle is divided into three rounds, and in one of these rounds you will see the value of your ball.
- At the point when you see your ball's value you will be asked to make your only choice for the cycle:

[3] Some passages are different depending on treatment, and are indicated by highlighted text and the name of the treatment in brackets. Everything else is the same.

  - either keep the ball you are holding.
  - or instead let go of your current ball and take hold of one of the other three balls.

Main Task

In more detail, a cycle proceeds as follows:

- In each new cycle and for each participant, the computer randomly draws four balls. Each ball's value is chosen in an identical manner:
  - With 50% probability the computer rolls a fair hundred-sided die: so the ball has an equal probability of being any number between 1 and 100.
  - With 25% probability the ball has value 1.
  - With 25% probability the ball has value 100.
- After drawing the values for the four balls, the computer randomly shuffles them into positions A to D. Once in place, the four balls' positions are fixed for the entire cycle. So whatever the value on Ball A, this is its value for the entire cycle. Only at the end of each cycle are four new balls drawn for your next cycle.
- Once the balls are in position, the computer randomly matches you to a ball, so you start out holding one of the balls A to D. There are therefore three leftover balls which are held by the Computer.
  - For example, one possible initial match might be that you hold Ball B. So because Ball B is held in this example, the Computer starts the cycle holding Balls A, C and D.
- Your outcome each cycle will depend only on the ball you are holding at the end of the cycle.
  - In each cycle you will know which of the four balls you start out holding.
  - You will not know any of the four balls' values to start with.
  - Every cycle you will make just one decision. At some random point in the cycle you will be told your ball's value. You will then be asked to make a choice after learning your ball's value: either keep holding your ball or give your ball to the computer, and instead take one of the three balls the computer is holding.

- The point in the cycle when you see your ball's value is random. Each cycle is divided into three rounds where you are given a chance to see your ball.
- The round in which you will see your held ball's value and make your choice for the cycle is random:
  - In round one, the computer flips a fair coin. If the coin lands Heads, you will see your ball's value. So you have a 50% chance of seeing your ball's value in the first round.
  - In round two, if you did not see you ball in round one you get another 50% chance of seeing its value: another coin flip.
  - Finally, in round three, if you did not see your ball's value in either round one or round two you will see its value for sure in round three and make your choice.
- Whenever you do see your held ball's value—either in round one, two or three— you will make your only decision for the cycle. The two options you have are:
  1. Keep hold of your ball until the end of the cycle.
  2. Switch balls: Give your ball to the computer to hold, and instead take one of the balls it is currently holding.

[S-Explicit]

- *If you do choose to switch balls with the computer, the procedure the computer uses to select a ball to give you in exchange varies with the round:*
  - *In round one, the computer will randomly select one of the three available balls, choosing between each of the three balls it is holding with equal probability.*
  - *In round two, the computer will randomly select a ball from the two lowest value balls of the three it is holding, choosing each of the two lowest-value balls with equal probability. So, in round two the computer will never offer you the highest-value ball of the three.*
  - *In round three, the computer will only offer you the lowest-value ball of the three it is holding.*

- Your cycle payoff is \$0.10 multiplied by the number on the ball you are holding at the end of the cycle. So a ball with value 1 at the end of a cycle has a payoff of \$0.10, a ball with value 50 has a payoff of \$5.00, while a ball with value 100 has a payoff of \$10.00.

Cycle Summary

1. Each participant is given four balls A to D, where each ball has a random value between 1 and 100.
2. Each participant is then assigned one of the four balls to hold, with the leftover balls held by the computer.
3. Across three rounds the participants are given the chance to see the value of the balls they are holding.
   - Whenever you see the value of the ball you are holding you must decide whether to keep holding it, or trade it with the computer.
   - In rounds one and two you have a 50% probability of seeing the held ball's values. Any participant that reaches round three without seeing their ball's value will always see its value in round three, and are then given the option to trade it for one of the computer balls.

   [S-Explicit]
   - *The procedure the computer uses to choose the ball it is willing to exchange with you changes across the cycle. In round one it will randomize across all three balls. In round two, it randomizes over the two lowest-value balls. In round three, it will offer the lowest-value ball with certainty.*

Experiment Organization

There will be three parts to this experiment. The first part will last for 5 cycles. After this you will get instructions for the second part which will last for another 15 cycles, where the task is very similar. Part 3 will last for a single cycle. Following part 3, we will conclude the experiment with a number of survey questions for which there is the chance for further payment.

Payment

- Monetary payment for Parts 1 and 2 will be made on two randomly chosen cycles, where each of the 20 cycles in the first two parts are equally likely to be selected for payment.
- You will be given the opportunity for further earnings in Part 3 and the survey at the end of the experiment (which we will explain once the preceding parts end).
- All participants will receive a $6 participation fee added to total earnings from the other parts of the experiment.

B.3.1.2 Selection, S-Across, S-Within, and S-Peer

Introduction[4]

Thank you for participating in our study. Please turn off mobile phones and other electronic devices. These must remain turned off for the duration of the session.

This is an experiment on decision making. The money you earn will depend on both your decisions and chance. The session will be conducted only through your computer terminal; please do not talk to or attempt to communicate with any other participants during the experiment. If you have a question during this instruction phase please raise your hand and one of the experimenters will come to where you are sitting to answer your question in private.

During the experiment, you will have the opportunity to earn a considerable amount of money depending on your decisions. At the end of the experiment, you will be paid in private and in cash. On top of what you earn through your decisions during the experiment, you will also receive a $6 participation fee.

Outline

Your interactions in this experiment will be divided into "Cycles".

- In each cycle you will be in a group of three, with each participant holding one of four balls, called Balls A to D.
- Each ball has a value between 1 and 100, and your payoff in each cycle will be determined by the value of the ball you are holding at the cycle's end.

[4]Some passages are different depending on treatment, and are indicated by highlighted text and the name of the treatment in brackets. Everything else is the same.

- At the start of each cycle you will see which ball each of the three participants are holding. However, you will NOT know any of the four ball's values. Each cycle is divided into three rounds, and in one of these rounds you will see the value of your ball.
- At the point when you see your ball's value you will be asked to make your only choice for the cycle:
  - either keep the ball you are holding.
  - or instead let go of your current ball and take hold of whichever ball is not being held by another group member.

Main Task

In more detail, a cycle proceeds as follows:

- At the start of each cycle the computer randomly divides all of the participants in the room into groups of three. Each player will randomly be given one of three roles: either First Mover, Second Mover or Third Mover.
  - The groups of three and specific roles assigned are fixed for each cycle.
  - In each new cycle you will be randomly matched into a new group of three.
  - In each new cycle you will be randomly assigned to either be the First, Second or Third Mover.
- In each new cycle and for each separate group of three, the computer randomly draws four balls. Each ball's value is chosen in an identical manner:
  - With 50% probability the computer rolls a fair hundred-sided die: so the ball has an equal probability of being any number between 1 and 100.
  - With 25% probability the ball has value 1.
  - With 25% probability the ball has value 100.
- After drawing the values for the four balls, the computer randomly shuffles them into positions A to D. Once in place, the four balls' positions are fixed for the entire cycle. So whatever the value on Ball A, this is its value for the entire cycle. Only at the end of each cycle are four new balls drawn for your next cycle and next group of three.
- Once the balls are in position, the computer randomly gives a different ball to each of the three group members. Each group member therefore starts out holding one of the

balls A to D. But because the three group members are each holding one of the four balls there is one leftover ball. This leftover ball is held by the Computer.

  - For example, one possible initial match might be that Ball A is held by the Third Mover; Ball B by the First Mover; and Ball D by the Second Mover. So because Balls A, B and D are all held in this example, the Computer starts the cycle holding the leftover Ball C.

[Selection, S-Across, and S-Peer]

- *Your outcome each cycle will depend only on the ball you are holding at the end of the cycle.*
  - *In each cycle you will know which of the four balls you start out holding*
  - *You will not know any of the four balls' values to start with, nor which balls the other two group members are holding.*

[S-Within]

- *Your outcome each cycle will depend only on the ball you are holding at the end of the cycle.*
  - *You will not know any of the four ball's values to start with.*
  - *You will know which ball is being held by each participant, and will also know if a ball was previously held by another participant.*
  - Every cycle you will make just one decision. At some random point in the cycle you will be told your ball's value. You will then be asked to make a choice after learning your ball's value: either keep holding your ball or give your ball to the computer, and instead take whichever ball the computer is holding.
- The point in the cycle when you see your ball's value is random. Each cycle is divided into three rounds where each group member is given a chance to see their ball. Each round is further divided into a sequence of turns, dictated by your role:
  1. The first mover gets the first opportunity to see their ball's value. If they see it, they make their one choice for the cycle, if not they must wait until the next round for another opportunity to see their ball's value.

2. After the first mover, the second mover gets an opportunity to see their ball. Again, if they see it, they make their one choice for the cycle, otherwise they must wait until the next round.
3. Finally, after both the first and second mover, the third mover gets an opportunity to see their ball. As before, if they see their ball they make their one choice for the cycle, otherwise they must wait until the next round.

- The round in which you will see your held ball's value and make your choice for the cycle is random:
  - In round one, the computer flips a fair coin once for each group member. If the coin lands Heads, the group member sees their ball's value. So each group member has a 50% chance of seeing their ball's value in the first round.
  - In round two, any group members who did not see their ball in round one get another 50% chance of seeing its value: another coin flip.
  - Finally, in round three, any group members who did not see their ball in either round one or round two see their ball's value for sure in round three and make their choice.
- Whenever you do see your held ball's value—either in round one, two or three—you will make your only decision for the cycle. The two options you have are:
  1. Keep hold of your ball until the end of the cycle.
  2. Switch balls: Give your ball to the computer to hold, and instead take the ball it is currently holding.
- The cycle ends after every participant within a group sees the ball's value and makes a decision.

[S-Across]

- *At the end of the cycle you will get feedback on what happened. You will be told:*
  - *The balls each group member and the computer started with.*
  - *Choice 1/2/3: The identity of the group member who was 1st/2nd/3rd to see their ball's value; the round they saw their ball's value; their choice (keep or switch); and which ball the computer was holding after their choice.*

- Your cycle payoff is \$0.10 multiplied by the number on the ball you are holding at the end of the cycle. So a ball with value 1 at the end of a cycle has a payoff of \$0.10, a ball with value 50 has a payoff of \$5.00, while a ball with value 100 has a payoff of \$10.00.

Cycle Summary

1. The computer randomly forms the participants in the room into groups of three.
2. Each group is given four balls A to D, where each ball has a random value between 1 and 100.
3. Each of the three participants are given one of the four balls to hold, with the leftover ball held by the computer.
4. Across three rounds the group members move in sequence according to their roles, and each are given the chance to see the value of the ball they are holding.
   - Whenever you see the value of the ball you are holding you must decide whether to keep holding it, or trade it with the computer.
   - In rounds one and two each group member has a 50% probability of seeing their held ball's value. Any group members that reach round three without seeing their ball's value will always see it in round three, and are then given the option to trade it for the current computer ball.

Experiment Organization

There will be three parts to this experiment. The first part will last for 5 cycles. After this you will get instructions for the second part which will last for another 15 cycles, where the task is very similar. Part 3 will last for a single cycle. Following part 3, we will conclude the experiment with a number of survey questions for which there is the chance for further payment.

Payment

- Monetary payment for Parts 1 and 2 will be made on two randomly chosen cycles, where each of the 20 cycles in the first two parts are equally likely to be selected for payment.
- You will be given the opportunity for further earnings in Part 3 and the survey at the end of the experiment (which we will explain once the preceding parts end).

- All participants will receive a $6 participation fee added to total earnings from the other parts of the experiment.

B.3.2 Instructions to Part 2 [Supergames 6-20]

We will now pause briefly before continuing on to the second part of the experiment. The task for the next 15 cycles of the experiment is very similar to the last 5. In fact, there is only one difference from part one.[5]

So far, if you flipped a head you have been told the value of the ball you are holding prior to deciding whether or not to trade it for the computer's ball. For the remaining cycles you instead will be asked to provide a cutoff rule in case you see your ball.

This cutoff is the minimum value you would need to keep the ball you are holding. In every round of a cycle, you will be asked to provide a cutoff for trading your ball should you see its value that round.

You will be asked to choose your cutoff value by clicking on the horizontal bar at the bottom of your screens per the projected slide. You can click anywhere on the bar to change your cutoff, and you can always adjust your minimum cutoff by plus or minus one by clicking on the two buttons below the bar.

In the projected example I selected a minimum cutoff of 80.

After you submit your cutoff the computer will then flip the coin if you are in rounds one or two to determine if you see your ball's value, similar to part one.

If the coin flip is tails, nothing happens, and you will have to wait to decide until at least the next round, where you will repeat this procedure and provide another minimum cutoff.

If instead the coin flip is Heads, or you are making your decision in round three where you are guaranteed to see your ball, the computer will show you the value of your ball. The computer will automatically keep it or trade your ball according to the minimum cutoff you selected.

If your ball's value is LOWER than your selected minimum, you will automatically

[5]The experimenter read the instructions aloud after Part 1 had ended. Slides were used to show screenshots and emphasize important points (see section B.3.3). The text was identical for all treatments, and the accompanying slides differ only for treatment S-Within.

trade your ball for the computer's ball, which you will keep until the end of the cycle. In the projected example I had selected 80 as my minimum cutoff. In the above example, it shows what would happen if I saw my held ball, and its value was 75. Because this is lower than my selected minimum value of 80, the computer uses my selected cutoff to automatically trade my ball for the computer's ball, rather than keeping it.

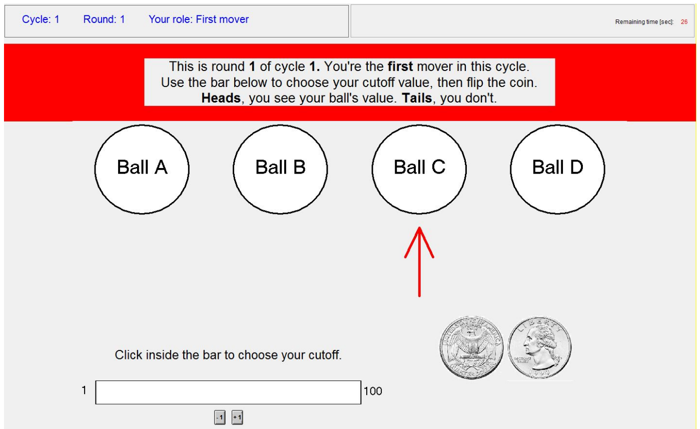
The next example shows what happens if the coin flip is heads, and your held ball is equal to or greater than your selected minimum cutoff. In this case, because the ball's value is greater than my selected minimum value of 80, the computer uses my selected cutoff to automatically keep my ball until the end of the cycle, rather than trading it. The projected example illustrates what would happen if my held ball had a value of 85. Because 85 is above my selected cutoff of 80, I would keep my ball until the end of the cycle.

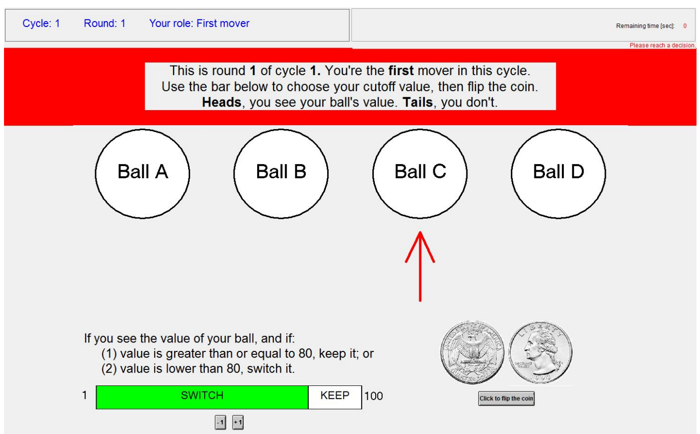
Because of this procedure, you will maximize your potential earnings by selecting the chosen cutoff value to answer the following question: What is the smallest value X for which I would keep my ball right now, where for any balls lower than X, I would rather trade them for the computer's ball?

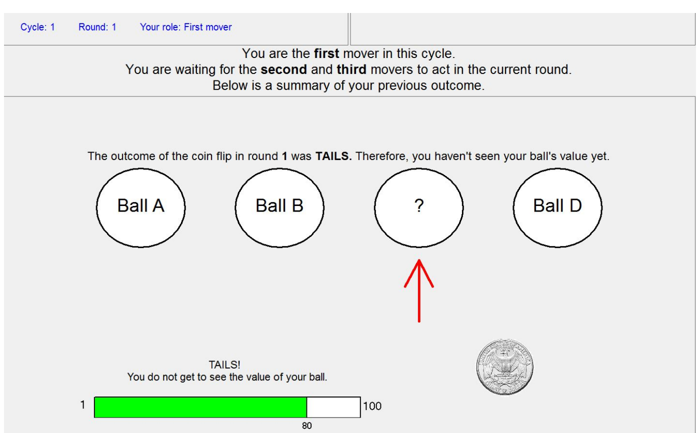
The computer will now ask you three questions to make sure you understand this cutoff. At the top of your screens the computer will indicate a ball you are holding. Just for these question we will also tell you the ball the computer is holding. For each question we will give you a selected cutoff, and the value of the ball you are holding. Given this information, we would like you to select what happens.

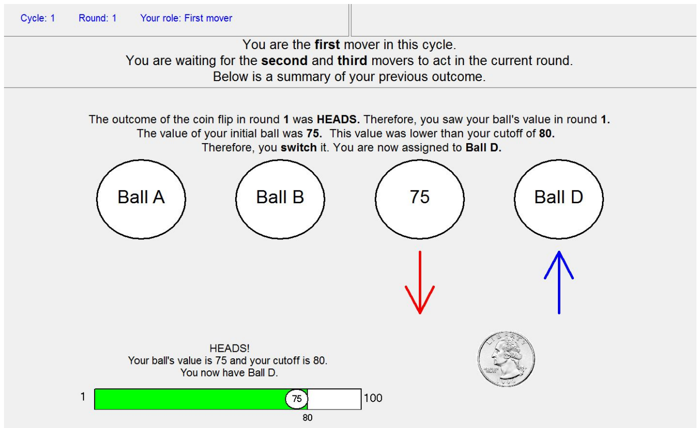
You must answer all three questions correctly for the experiment to proceed.

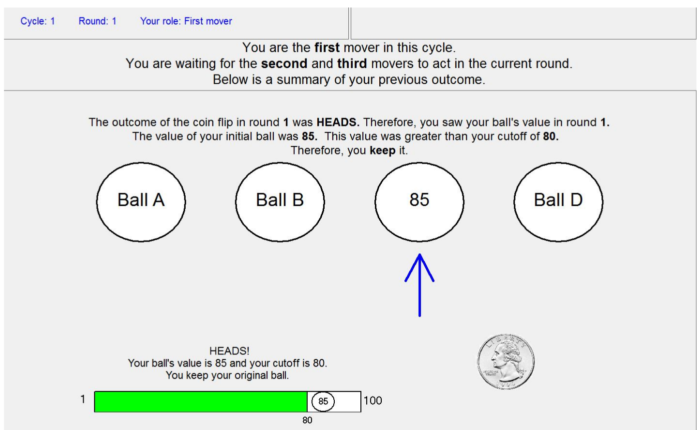
B.3.3 Slides for Instructions to Part 2








Cycle: 1
Round: 1
Your role: First mover
You are the **first** mover in this cycle.
You are waiting for the **second** and **third** movers to act in the current round.
Below is a summary of your previous outcome.
The outcome of the coin flip in round **1** was **HEADS.** Therefore, you saw your ball's value in round **1.**
The value of your initial ball was **75.** This value was lower than your cutoff of **80.**
Therefore, you **switch** it. You are now assigned to **Ball D.**
Ball A
Ball B
75
Ball D
HEADS!
Your ball's value is 75 and your cutoff is 80.
You now have Ball D.
1
75
100
80


Cycle: 1
Round: 1
Your role: First mover
You are the **first** mover in this cycle.
You are waiting for the **second** and **third** movers to act in the current round.
Below is a summary of your previous outcome.
The outcome of the coin flip in round **1** was **HEADS.** Therefore, you saw your ball's value in round **1.**
The value of your initial ball was **85.** This value was greater than your cutoff of **80.**
Therefore, you **keep** it.
Ball A
Ball B
85
Ball D
HEADS!
Your ball's value is 85 and your cutoff is 80.
You keep your original ball.
1
85
100
80

B.3.4 Handouts for Part 2 [Supergames 6-20]

B.3.4.1 No Selection (Control) and S-Explicit

Part Two[6][7]

- Everything in part two cycles will be the same as part one, with one exception.
- In every round, you will be asked to provide a *cutoff*.
- The *cutoff* you provide is the *minimum* value required for you to keep your ball.
- After you have confirmed your cutoff X between 1 and 100 the computer will determine if you see your ball's value this round:
  - If you see your ball's value this round a choice will be made according to your minimum-value cutoff.
  - If you do not see your ball's value this round, you will provide another minimum-value cutoff in the next round.
- In the round where you see your ball, the computer will use your minimum cutoff X as follows:
  - *IF* your ball's value is equal to or greater than the minimum cutoff X, then you will choose to keep your ball.

  [No Selection (Control)]
  - *OTHERWISE if your ball's value is less than the minimum cutoff X, the computer will choose to trade your ball for one of the computer's balls.*

  [S-Explicit]
  - *OTHERWISE if your ball's value is less than the minimum cutoff X, the computer will choose to trade your ball for the one chosen by the computer that round.*

  [Selection, S-Across, S-Within, and S-Peer]
  - *OTHERWISE if your ball's value is less than the minimum cutoff X, the computer will choose to trade your ball for the one held by the computer that round.*
- Because of this procedure, you should choose your cutoff value to answer the following question:

[6] The experimenter distributed the handouts before showing the slides and reading the script. They served as a consultation material for subjects.

[7] Some passages are different depending on treatment, and are indicated by highlighted text and the name of the treatment on brackets. Everything else is the same.

[No Selection (Control)]

What is the smallest value X from 1 to 100 for which I would like to keep my ball right now, rather than give it to the computer in exchange for one of the balls the computer is holding?

[S-Explicit]

What is the smallest value X from 1 to 100 for which I would like to keep my ball right now, rather than give it to the computer in exchange for its selected ball for this round?

[Selection, S-Across, S-Within, and S-Peer]

What is the smallest value X from 1 to 100 for which I would like to keep my ball right now, rather than give it to the computer in exchange for the ball the computer is holding?

B.3.5 Instructions to Part Three

We will now pause briefly before continuing on to the third part of the experiment. The task for the final cycle of the experiment is very similar to the last 20. However, where we paid two random cycles from the last 20, we will pay you whatever you earn in this last cycle for sure. So for this one cycle you will earn between \$0.10 and \$10.00 depending on your final ball.[8]

In this cycle we will randomly assign you to be either a first, second or third mover, and will tell you your role. Like the preceding rounds, we will ask you for a minimum cutoff value to keep your ball, and the computer will automatically keep or switch your ball depending on your selected cutoff if you see your balls value that round.

[No Selection, Selection, S-Explicit, and S-Across]

The only difference in part 3 is that we will only tell you which round you saw your ball's value in at the end of the cycle.

[S-Within]

There are two differences in Part 3. First, we have removed the information on which balls the other participants are holding. All you will know is which ball you are currently

[8]The experimenter read the instructions aloud after Part 2 had ended.
Slides were used to show screenshots and emphasize important points (see section B.3.6).

holding. Second, we will now only tell you which round you saw your ball's value and how you decided at the very end of this cycle.

You will submit cutoffs in rounds 1 to 3, as before. In the round where you see your ball's value, you will use your selected cutoff for that round to make a decision.

[No Selection, Selection, S-Across, and S-Within]

If your balls' value is lower than your minimal cutoff you will exchange your ball with the one the computer is holding at that point.

[S-Explicit]

If your ball's value is lower than your minimal cutoff you will exchange your ball with the one the computer has selected to exchange for that round.

If your balls' value is equal to or greater than your minimum cutoff you will keep your ball for the cycle. Because of this procedure, nothing in the structure of the task has changed from Part 2. So you should make your decisions exactly as before. The new procedure allows us to collect information on the cutoffs you would select in all three rounds of the cycle.

[No Selection, Selection, S-Across, and S-Explicit]

Effectively, the only thing that has changed from part two is the point at which we tell you you have made your choice, and that your payoff from this cycle will always be added to your final cash payoff for the experiment.

[S-Within]

Recall that, in contrast to the previous cycles, (i) you will now not receive any information about which balls the other two participants and the computer are currently holding in any period, and (ii) you will choose a cutoff in all three rounds.

In terms of payment, remember that the outcome from this cycle will always be added to your final cash payoff for the experiment, so that the ball you are holding at the end of this cycle will add between $0.10 and $10.00 to your final payoff, depending on the ball you are holding at the end.

Please now make your choices for the final cycle.

B.3.6 Slides for Instructions to Part 3

B.3.6.1 No Selection, Selection, S-Across, S-Explicit, and S-Within

- Part Three will consist of a single cycle
- Whatever you earn in this cycle will be added to the two random cycles selected from Parts 1 and 2
  - So you have the chance to earn between $0.10 and $10.00 for this cycle depending on your final ball

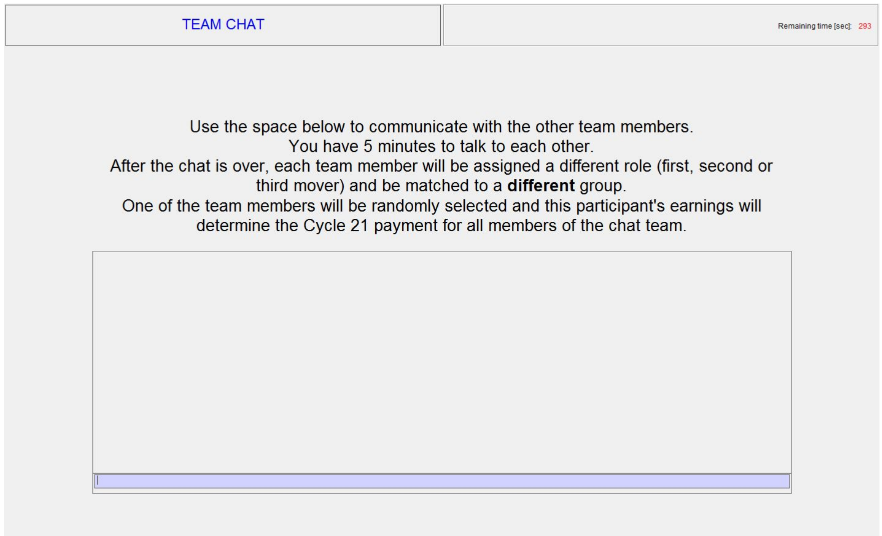
- You will be told whether you are the First, Second or Third mover
- We will ask you for your minimal cutoff in each new round as before
- If your coin flip is heads, you will choose an action according to your cutoff as before
- The only difference is that we will not tell you when and if you have made a choice until the end of the cycle

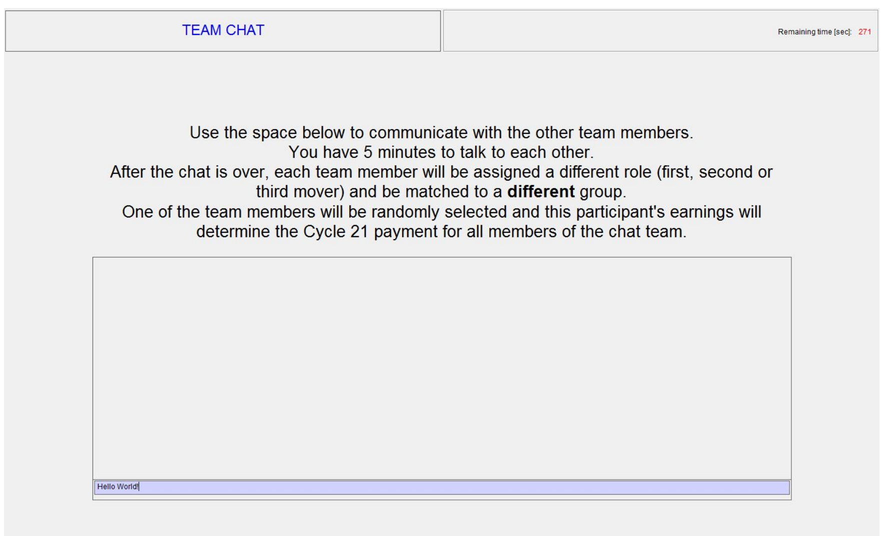
- You will submit cutoffs in rounds 1 to 3, as before
- In the round where you see your ball's value, you will use your selected cutoff to make a decision
  - If your balls' value is lower than your minimal cutoff you will exchange your ball with the one the computer is holding
  - If your balls' value is equal to or greater than your minimal cutoff you will keep your ball
- Because of this procedure, nothing in the structure of the task has changed
  - The new procedure allows us to collect information on the cutoff you would select in all three rounds
  - All that has changed from part two is the time at which we inform you on your choice for the cycle

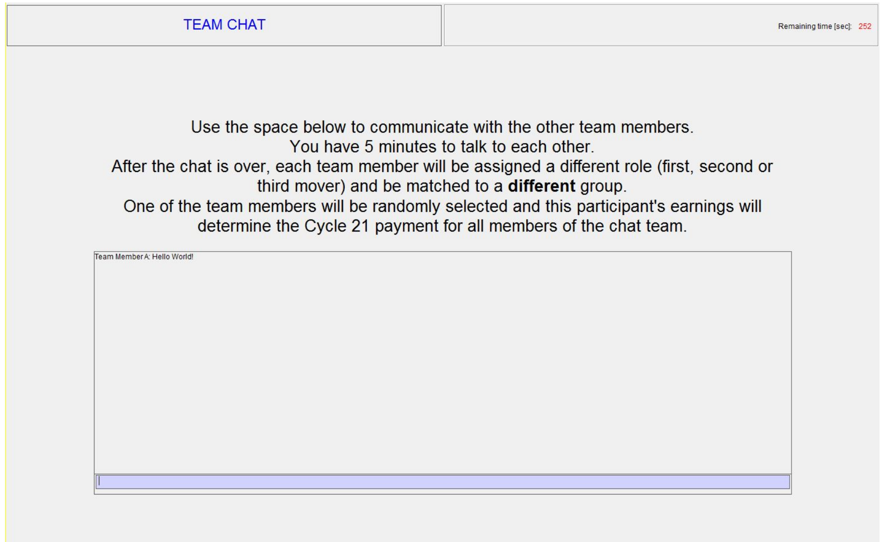
B.3.6.2 *S-Peer*

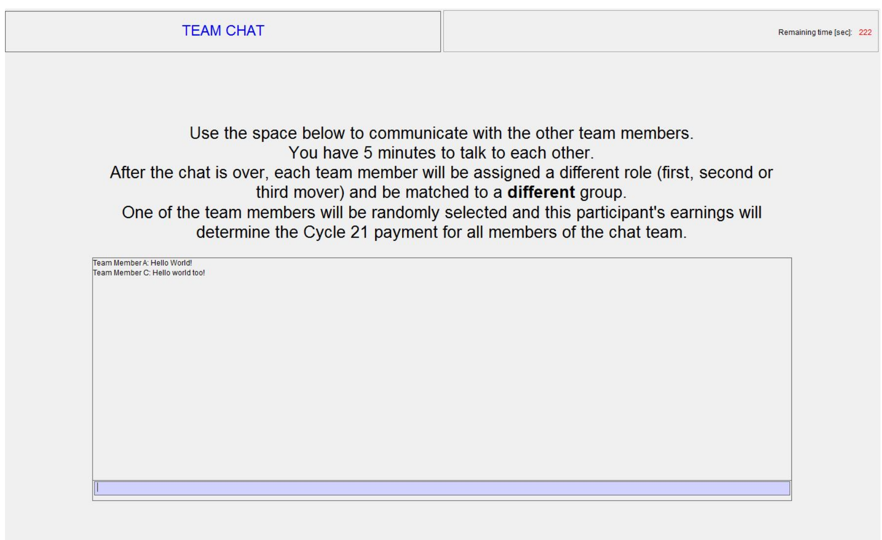
- Part Three will consist of a single cycle
- Whatever you earn in this cycle will be added to the two random cycles selected from Parts 1 and 2
  - So you will earn between $0.10 and $10.00 for this cycle
- We will ask you for your minimal cutoff in each new round as before
- If your coin flip is heads, you will choose an action according to your cutoff as before
- The only difference is that we will not tell you when and if you have made a choice until the end of the cycle

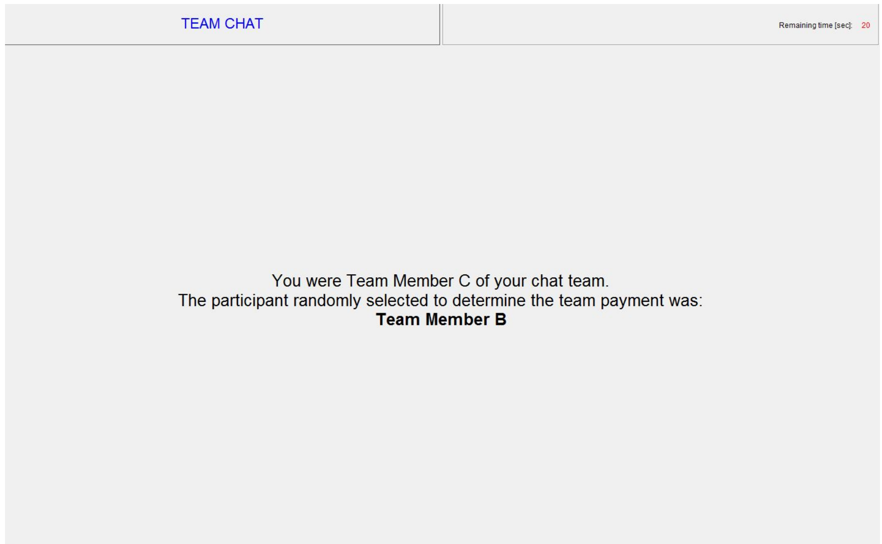
- You will submit cutoffs in rounds 1 to 3, as before
- In the round where you see your ball's value, you will use your selected cutoff to make a decision
  - If your balls' value is lower than your minimal cutoff you will exchange your ball with the one the computer is holding
  - If your balls' value is equal to or greater than your minimal cutoff you will keep your ball
- Because of this procedure, nothing in the structure of the task has changed.
  - The new procedure allows us to collect information on the cutoff you would select in all three rounds
  - All that has changed from part two is the time at which we inform you on your choice for the cycle

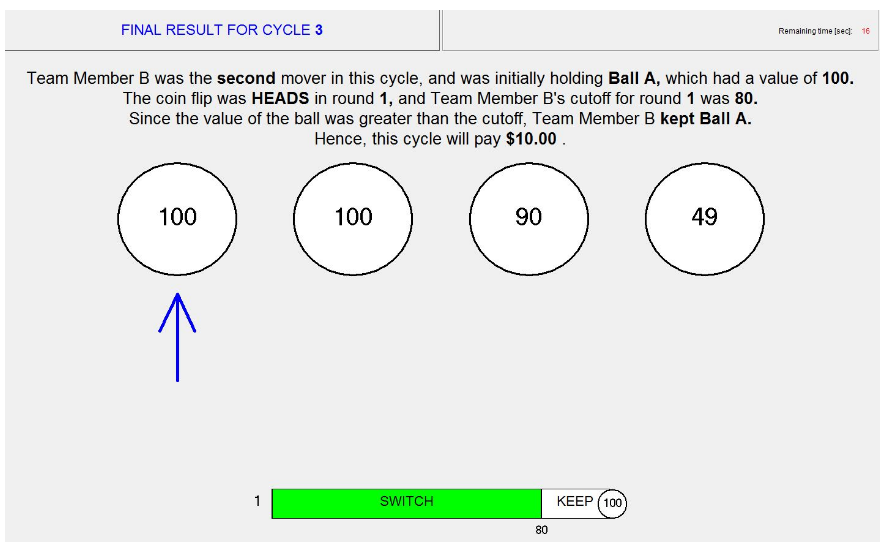

TEAM CHAT
Remaining time [sec]: 293
Use the space below to communicate with the other team members.
You have 5 minutes to talk to each other.
After the chat is over, each team member will be assigned a different role (first, second or third mover) and be matched to a **different** group.
One of the team members will be randomly selected and this participant's earnings will determine the Cycle 21 payment for all members of the chat team.


TEAM CHAT
Remaining time [sec]: 271
Use the space below to communicate with the other team members.
You have 5 minutes to talk to each other.
After the chat is over, each team member will be assigned a different role (first, second or third mover) and be matched to a **different** group.
One of the team members will be randomly selected and this participant's earnings will determine the Cycle 21 payment for all members of the chat team.
Hello World!


TEAM CHAT
Remaining time [sec]: 252
Use the space below to communicate with the other team members.
You have 5 minutes to talk to each other.
After the chat is over, each team member will be assigned a different role (first, second or third mover) and be matched to a **different** group.
One of the team members will be randomly selected and this participant's earnings will determine the Cycle 21 payment for all members of the chat team.
Team Member A: Hello World!


TEAM CHAT
Remaining time [sec]: 222
Use the space below to communicate with the other team members.
You have 5 minutes to talk to each other.
After the chat is over, each team member will be assigned a different role (first, second or third mover) and be matched to a **different** group.
One of the team members will be randomly selected and this participant's earnings will determine the Cycle 21 payment for all members of the chat team.
Team Member A: Hello World!
Team Member C: Hello world too!


TEAM CHAT
Remaining time [sec]: 20
You were Team Member C of your chat team.
The participant randomly selected to determine the team payment was:
Team Member B


FINAL RESULT FOR CYCLE 3
Remaining time [sec]: 16
Team Member B was the second mover in this cycle, and was initially holding Ball A, which had a value of 100.
The coin flip was HEADS in round 1, and Team Member B's cutoff for round 1 was 80.
Since the value of the ball was greater than the cutoff, Team Member B kept Ball A.
Hence, this cycle will pay $10.00 .
100
100
90
49
1
SWITCH
KEEP
100
80

B.3.7 Instructions for Part 4

Finally we will conduct a number of survey questions for which there will be the chance of an additional payment.

Part four consists of three sets of questions, which will have a series of possible prizes. One participant in the room will be randomly selected for payment on these additional questions.

The first question is a decision making task. You will be presented with three balls. One of these balls is worth $10, while the other two are worth $0. The computer has shuffled the three balls, and fixed their locations.

1. We will ask you to choose one of the the three balls.
2. After you have chosen a ball, we will reveal one zero dollar ball from the two balls that you did not choose.
3. We will then make you an offer:
   - Would you like the ball you initially chose plus $5?
   - Or would you instead like to switch to the remaining ball plus X times $0.10?

- X will vary between 1 and 100.
- We would like you to tell us the minimum value of X for which you would like to swap.
  - If you swap you get the value of the remaining unchosen ball (either 0$ or $10) plus $0.10 times X (so between $0.10 and $10)
  - If you keep your ball, you get its value (either 0$ or $10) plus $5.

Please make your choices for this task now.

[Wait while subjects complete task]

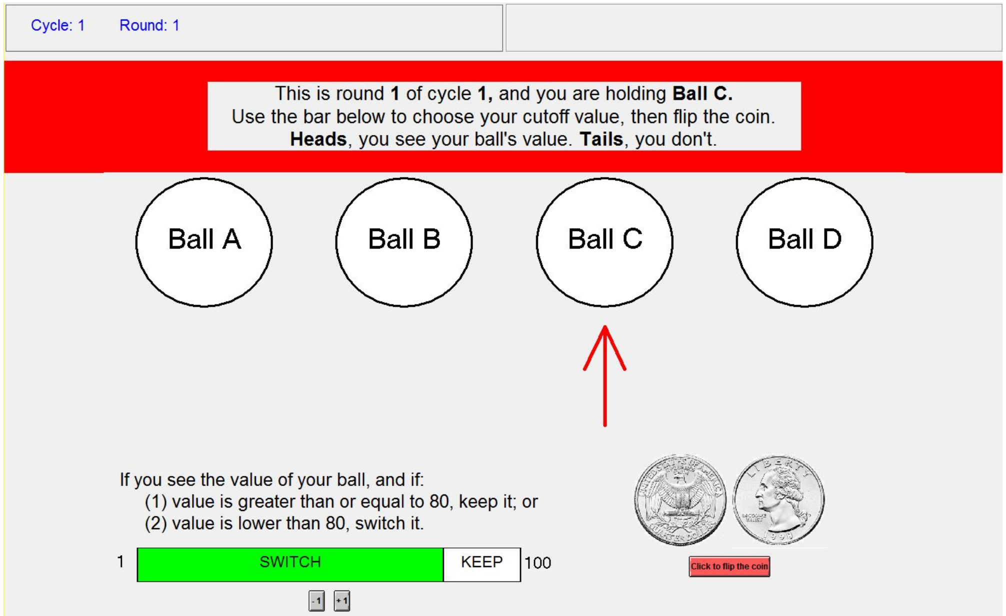
The next task will ask you to answer three numerical questions within a 15 second time-limit. Whoever is selected for payment in part 4 will receive $1 per correct answer.

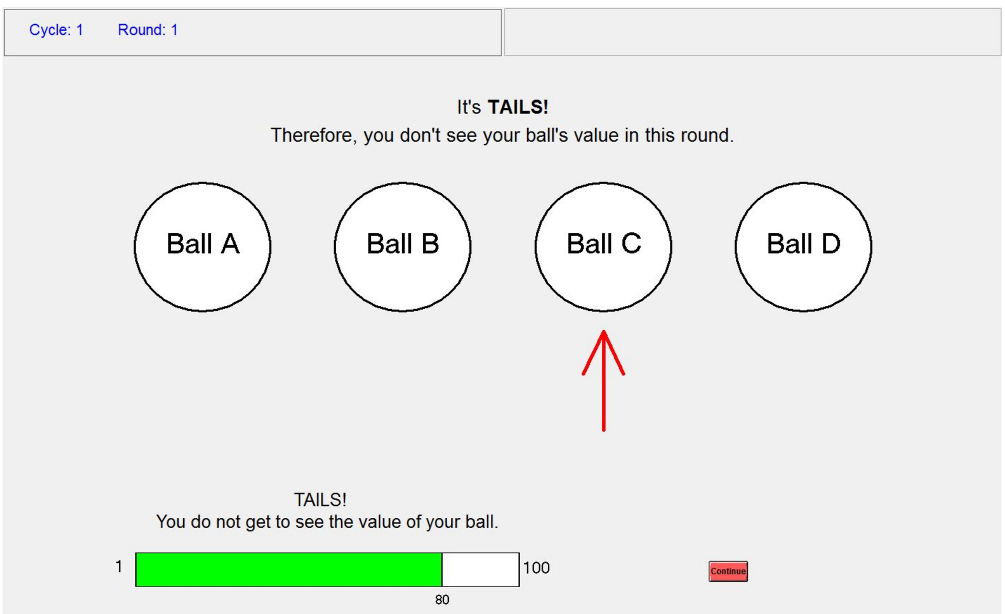
[Wait while subjects complete task]

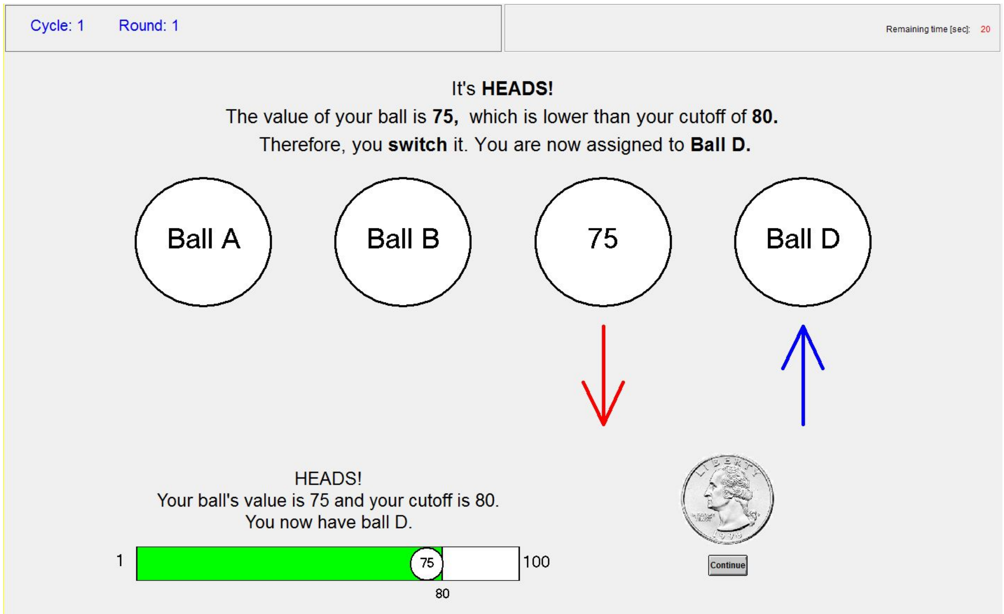
Finally, we would like you to make a series of choices between lotteries. In each choice you will be asked to pick either Lottery A or Lottery B, where each offers a probability over two monetary prizes.

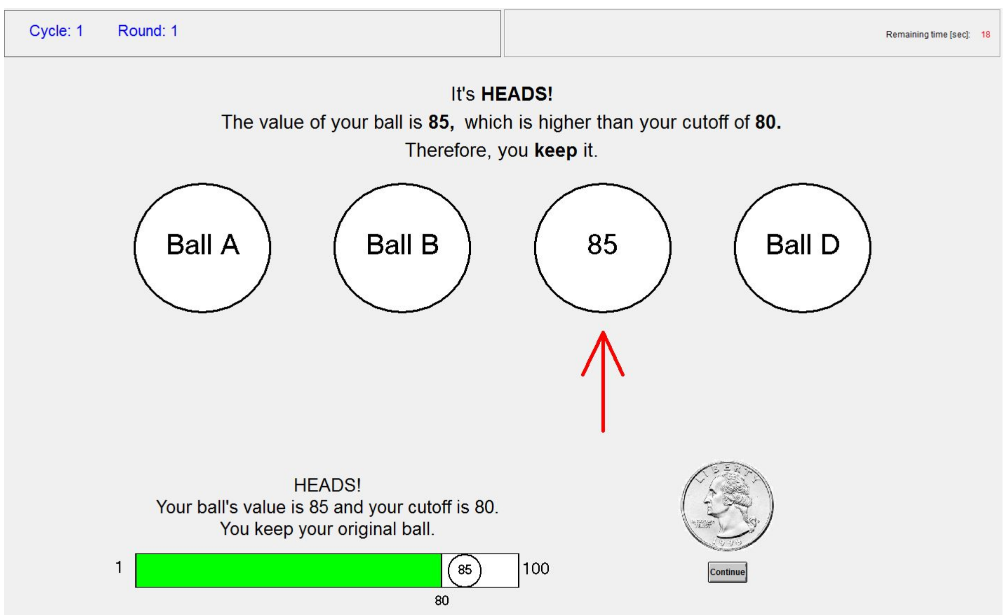
One of your four choices from these lotteries will be selected for payment, and the outcome added to your total earnings if you are selected for payment in part four.

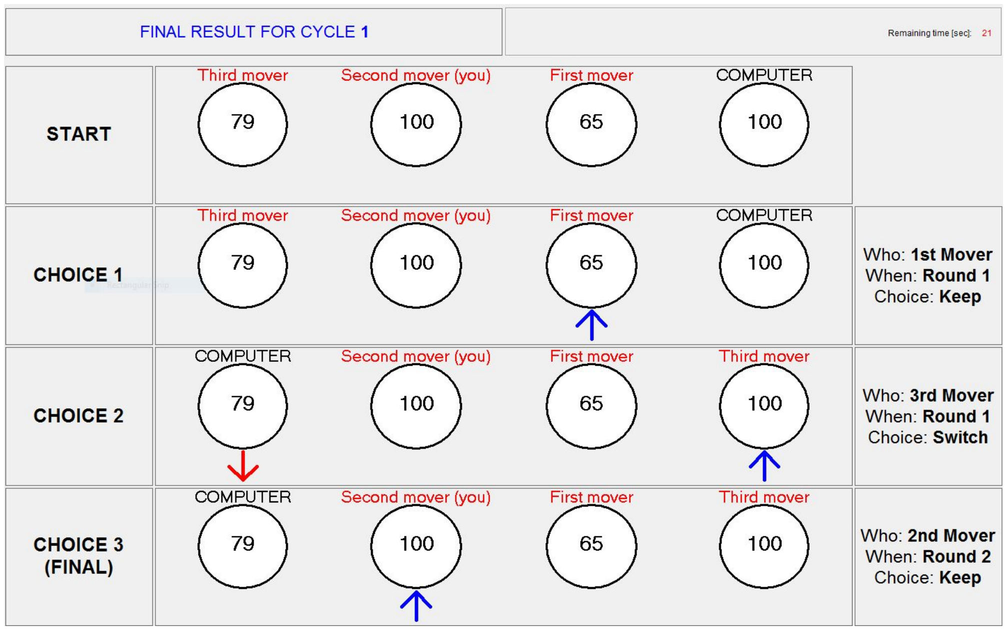
B.4 Screenshots






Cycle: 1
Round: 1
Remaining time [sec]: 20
It's **HEADS!**
The value of your ball is **75,** which is lower than your cutoff of **80.**
Therefore, you **switch** it. You are now assigned to **Ball D.**
Ball A
Ball B
75
Ball D
HEADS!
Your ball's value is 75 and your cutoff is 80.
You now have ball D.
1
75
100
80
Continue


Cycle: 1
Round: 1
Remaining time [sec]: 18
It's **HEADS!**
The value of your ball is **85,** which is higher than your cutoff of **80.**
Therefore, you **keep** it.
Ball A
Ball B
85
Ball D
HEADS!
Your ball's value is 85 and your cutoff is 80.
You keep your original ball.
1
85
100
80
Continue


FINAL RESULT FOR CYCLE 1
Remaining time [sec]: 21
START
Third mover
79
Second mover (you)
100
First mover
65
COMPUTER
100
CHOICE 1
Third mover
79
Second mover (you)
100
First mover
65
COMPUTER
100
Who: 1st Mover
When: Round 1
Choice: Keep
CHOICE 2
COMPUTER
79
Second mover (you)
100
First mover
65
Third mover
100
Who: 3rd Mover
When: Round 1
Choice: Switch
CHOICE 3 (FINAL)
COMPUTER
79
Second mover (you)
100
First mover
65
Third mover
100
Who: 2nd Mover
When: Round 2
Choice: Keep

B.4.1 Chat Transcripts

Session-17, Group-1

A: Hey guys (*t=6*)

C: hello (*t=14*)

B: Hi, ideally we want the bot to have the lowest value (*t=20*)

A: true (*t=36*)

C: i think a decent cutoff is somewhere between 20-50 (*t=51*)

B: this way when one of us 3 get selected at random we will always not have the lowest value chosen (*t=56*)

B: I think a cut off between those values is ideal as well (*t=69*)

A: i usually put my cut-offs lower if i'm third mover, around 5-20 (*t=91*)

A: because the people before me are likely switching out a 1 value (*t=107*)

C: very true (*t=116*)

B: (*t=127*)

A: but i agree with 20-50 for first or second mover (*t=128*)

C: so if we are a first mover go between 20-50 and lover it if we are the later mover (*t=141*)

A: progressively getting lower with each round (*t=141*)

B: I think that is a good plan (*t=144*)

B: Also don't forget the coin flip (*t=150*)

A: but we won't know the outcome of it :((*t=158*)

B: we can communicate if we get heads or tails (*t=167*)

B: for example (*t=171*)

B: if you are first mover (*t=181*)

B: quickly select your cutoff value at 40 or something around there (*t=190*)

B: oh shit nevermind (*t=204*)

A: my understanding is that this round, we won't be told which round we're making the decision, therefore we won't know where we're making our decision (*t=216*)

C: yeah, i think we just have to go in with the cutoffs we want (*t=236*)

B: Yea I made a mistake, I guess we will just have to follow the formula of putting in low cutoffs later on (*t=243*)

C: sounds good (*t=251*)

A: sounds good (*t=253*)

B: sounds good (*t=262*)

B: (___/) (*t=281*)

B: (= ' . ' =) (*t=286*)

B: ()__() (*t=289*)

A: wow you're good at that (*t=297*)

Session-17, Group-2

A: hey (*t=9*)

C: hello there (*t=18*)

B: hello (*t=22*)

A: so honeslty i've just been making like 60 my cut-off? (*t=37*)

C: Ive been doing about 60-65 for the first one (*t=51*)

B: ive been using 40 just to be safe (*t=61*)

B: what are u guys going to use on this round (*t=74*)

C: and then once Ive gotten to the 2nd and 3rd rounds Ive decreased it to like 40 and then 20 just to be safer (*t=79*)

A: uhhh, i'm probably going to stick to 60ish (*t=95*)

A: idk really haha (*t=99*)

C: I think I'm going to do 60 too fir the first round (*t=111*)

B: lets hope it gives us all 100s (*t=116*)

A: yeah, forreal lol (*t=121*)

B: *all get 1s' line up for dimes (*t=141*)

A: i've been just trying to calculate it so i atleast come out with $20 when i can (*t=148*)

A: wow, no don't even joke (*t=155*)

A: haha (*t=158*)

B: lol (*t=159*)

C: did you guys decrease your cutoff as the rounds go on or no? (*t=166*)

A: the lowest i've gone is 50 (*t=175*)

B: i left mine at 40 worked good, either get 60+ or 1 (*t=184*)

A: because we've had some rounds where the values were really low, but not too many (*t=191*)

A: I usually got pretty lucky I think (*t=199*)

C: yea i havent had too many in the 20-50 range (*t=205*)

A: all right, 60s it is haha (*t=255*)

B: if we all have all put a high cut off do you think its better for this round (*t=258*)

A: uhhhh I think that's why I like being in the middle because it's kinda a catch all. (*t=281*)

A: like you might not make a lot but you probs won't get 0.01 either (*t=291*)

A: or .10 (*t=295*)

Session-17, Group-3

C: Hi! (*t=9*)

B: Hello! (*t=11*)

A: hi! (*t=16*)

C: How is everyone? (*t=18*)

A: kinda tired (*t=23*)

C: I hear ya. Me too (*t=30*)

B: Yeah (*t=33*)

C: Anyways, what would you guys like to do? (*t=46*)

A: Im not really sure (*t=71*)

C: Does anyone have any opinions about the study cut offs. (*t=75*)

B: my cutoff was 50, what are yours? (*t=95*)

C: I am not really sure either. I had lower cut off during my game (1-10) (*t=100*)

A: mine was 35 (*t=109*)

B: 1-10 was kinda low, i think (*t=118*)

C: How did 50 work? Or 35? (*t=119*)

B: 50 give a lot of chance to get 100 (*t=127*)

C: Did it give you higher chances of having a better ball? (*t=132*)

B: yeah, like i have 5 times 100 (*t=151*)

A: I dont necesarily think so (*t=155*)

C: Good. My lower numbers did not work well for me (*t=166*)

B: how many 100 did 35 get? (*t=172*)

A: mm maybe like 5 or 6 (*t=186*)

C: I did it because at the beginning of the experiment I had several 1s going back to 1s (*t=197*)

C: Yeah, I had like 4 100s, and a few 80s (*t=211*)

C: You did better (*t=214*)

C: So it sounds like a slightly higher cut off gives better odds (*t=238*)

C: Thanks for the good information (*t=252*)

A: okay so you want to go with the 50 (*t=253*)

C: I do (*t=259*)

C: Want to go with 50, sorry (*t=264*)

A: lol its cool (*t=279*)

C: It seems like it yields better probability in the outcome (*t=280*)

C: Cool (*t=282*)

C: Good luck to everyone n the rest of the study (*t=297*)

Session-17, Group-4

C: I've been using 60 as my minimum cutoff (*t=22*)

B: 50 seems to be the safest bet. Worked well so far. (*t=28*)

A: ive used 48 (*t=33*)

B: so lets averaged them? (*t=53*)

A: 49???? (*t=64*)

C: Around 52 or 53? (*t=96*)

C: If we average the 3 (*t=102*)

B: yeah it would be 52 and change (*t=108*)

C: Ok so how about 52 (*t=117*)

B: works for me (*t=122*)

A: hmmm (*t=130*)

A: what about a little lower (*t=161*)

B: why lower? (*t=169*)

B: 51? (*t=184*)

A: what if it's 50 (*t=186*)

B: ok (*t=189*)

C: Works for me (*t=192*)

A: and then we get a 1 (*t=193*)

A: ok so 50 (*t=198*)

B: yes, 50 (*t=203*)

A: :) (*t=221*)

Session-17, Group-5

C: soo what should we make the cutoff (*t=14*)

B: As the rounds go on, the chances that the ball the computer is holding has a really small value increases (*t=31*)

C: yea (*t=38*)

B: because in previous rounds, if someone had a small value they probably switched and gave it to the computer (*t=55*)

B: so what I've been doing is decreasing my cutoff values each round (*t=66*)

C: so as the rounds go on we should increase the cutoff (*t=67*)

A: go down by 5 for cutoff in each round? (*t=67*)

B: decrease (*t=74*)

C: decrease yea (*t=80*)

C: my bad haha (*t=83*)

B: I've been doing 50/40/20 (*t=93*)

B: idk what you guys think haha (*t=99*)

C: ive been doing 30/25/20 (*t=107*)

C: because im cool with $3 (*t=121*)

C: instead of potentially giving up close to $5 (*t=141*)

B: yeah that's true (*t=147*)

A: ive been 45/35/25 (*t=148*)

A: doesn't matter to me (*t=156*)

C: how about we do 40/30/20? (*t=163*)

B: yeah sounds good (*t=169*)

A: perfect (*t=173*)

C: alright cool (*t=176*)

C: lets make some money (*t=182*)

B: we get matched with other people though anyway right haha (*t=185*)

C: oh idk whats the point of this then (*t=200*)

B: just to share strategies I guess lol (*t=213*)

C: true haha (*t=220*)

A: yeah it says above each team member will be assigned a different group (*t=241*)

C: well good luck (*t=242*)

B: yep to you guys too haha (*t=258*)

C: 8^)-+((*t=278*)

Session-17, Group-6

A: Hello World! (*t=12*)

C: Hello (*t=16*)

B: hello (*t=18*)

B: what cutoff number do you guys feel is best/ (*t=64*)

A: What have you folks been using as your cutoff? (*t=65*)

A: I've been doing 20 (*t=82*)

B: ive been using everything above 60 (*t=85*)

C: So here are my thoughts: The chance of you getting a low # that someone else switched out is based on which mover you are and what round it is. Typically I go with ~50 if I am mover 1 or 2 on the first round (*t=96*)

B: so is 50 the best option then? (*t=129*)

A: Alright so mover 1 and 2 we want to do 50 plus (*t=131*)

B: that sounds good (*t=152*)

C: Then drop down for each subsequent round. Because you get stuck with what you switch too and as time goes on that is much more likely to be a low # (*t=154*)

A: So thats round 1 (*t=161*)

A: Do the same thing for round 2?? (*t=169*)

B: yeah we should (*t=191*)

A: Alright want to drop to 40? (*t=197*)

B: it feels better biting doin (*t=202*)

B: yeah we should drop to 40 (*t=218*)

C: I'd say if you are mover 1 or 2, you do 50/40/20. Mover 3 40/30/20 (*t=230*)

A: Follow that exactly (*t=240*)

C: Or something like that (*t=243*)

B: ok gotcha (*t=255*)

A: ;) (*t=273*)

C: Go team! :) (*t=291*)

A: Let (*t=296*)

Session-18, Group-1

B: wazzzzzzzzzzzzzzup any ideas? (*t=9*)

C: noooope (*t=16*)

B: ive just put 40 everytime haha (*t=33*)

C: i did 49 everytime (*t=46*)

A: i did 10 everytime (*t=58*)

B: soooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo should we all put 10 orrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr (*t=101*)

C: hmmmm (*t=104*)

C: my brain hurts i just wanna leave (*t=129*)

A: haha (*t=136*)

B: if i walk outta here with 20 bucks ill be happy (*t=153*)

C: ya if it's less i'll feel like i wasted my time (*t=177*)

C: but does 10 work/? (*t=185*)

B: i think imma stick with 40, it worked for me nice (*t=216*)

C: same w 49 hehe (*t=228*)

B: alright so basically we all just wasted 5 min here (*t=245*)

B: GO TEAM (*t=248*)

C: other people r typin so much like damn what are they plannin out here (*t=257*)

B: world domination (*t=266*)

A: no idea (*t=267*)

C: may the odds be ever in your favor (*t=288*)

C: good luck yallllll (*t=294*)

Session-18, Group-2

B: hi (*t=10*)

C: I always put 42 (*t=14*)

C: every time (*t=17*)

A: I went 70 pretty much everytime (*t=26*)

B: im cool w/ 42 (*t=35*)

A: It was pretty successful (*t=42*)

C: 42 was also successful (*t=52*)

A: I got 100 a good amount of times (*t=54*)

C: i think they had higher values for the last part than the first (*t=66*)

C: is this the team or do we get new teams? (*t=78*)

A: We are collectively deciding on one value right? (*t=79*)

B: i also went in the 40's and got 100 6-10 times (*t=82*)

C: me too (*t=92*)

B: this is our team (*t=94*)

B: idt we all have to say the same amount tho (*t=111*)

A: I know but are we deciding on one value collectively? (*t=113*)

B: i dont belive so, just talking about a gameplan (*t=128*)

B: i just dont want .10 tbh (*t=167*)

A: ok so who ever gets to decide, you guys want to go in the 40s? (*t=171*)

A: I would prefer to go a little higher (*t=185*)

A: like 60s (*t=193*)

A: ???? (*t=226*)

C: i guess the question is, if we all put the same number, will we get the lower thing that was given up by someone if they accepted the higher one (*t=230*)

B: i think we all play the round with other people like normal and then one of our values in chosen as our payment (*t=261*)

B: so we dont all have to say the same # (*t=272*)

A: Whats the point of the chat then? (*t=287*)

C: if we were all successful, should we just play as we had been (*t=292*)

Session-18, Group-3

B: any of you know what we're supposed to be discussing? (*t=47*)

C: What the minimum should be? (*t=57*)

B: Awesome. Thanks (*t=70*)

C: maybe 60? (*t=83*)

A: so what value are you planning to set (*t=85*)

A: i guess it can be lower, around 50? (*t=128*)

C: that sounds good (*t=137*)

A: ˆ ˆ (*t=153*)

C: so we all set 50 as minimum (*t=212*)

A: yes, i would do this (*t=237*)

B: we'll all be matched to different groups. I'm not quite sure why it would matter whether or not we all picked the same (*t=255*)

A: oh (*t=267*)

A: i thought we will be in the same group... sry (*t=284*)

C: same (*t=288*)

A: good luck (*t=299*)

B: 50 would work if we were all together, but in separate groups, we should probably just do as before (*t=300*)

Session-18, Group-4

C: hello (*t=56*)

A: hello (*t=64*)

B: hey hows it going (*t=81*)

C: oh you know pretty solid (*t=91*)

A: i really have to pee (*t=98*)

C: thanks for the info (*t=107*)

B: Im getting sleepy (*t=110*)

C: (*t=121*)

C: lets all just get 100 on this last level so we're gtd the $10 (*t=121*)

A: are we supposed to be deciding anything (*t=136*)

C: not really just talking about what our cutoffs will be (*t=152*)

A: mine has been 50 the whole time (*t=163*)

C: i set 50 for first mover and less for second & third (*t=178*)

C: but its mostly luck (*t=193*)

B: Mine are 20 if Im the first or second mover; after that I use 2 (*t=220*)

C: the key is to just get a 100 ball (*t=266*)

C: gl (*t=298*)

Session-18, Group-5

C: What's good team (*t=15*)

A: hey team! My strategy has been to choose a low cutoff (between 10 and 25) since there seem be a lot of 1s in each round and it seems better to keep anything above 1 even if it's not that high of a number (*t=58*)

A: sorry that was super detailed (*t=95*)

C: I've been going around 40-50. Mine have had a lot of 100s (*t=121*)

A: i think it is good to choose a higher cutoff if you are mover 1 (*t=154*)

A: i have just by chance been mover 3 a lot, which means I am more likely to get someone's discarded 1 (*t=170*)

B: I choose a cutoff value of 50. Only because there seem to be more numbers above 50 than below (*t=183*)

C: I would agree with that. I sort of assume the later you move the worse the remaining ball is (*t=195*)

B: that's true. It's better to find out early what the value of your ball is. And keep it if it is high. (*t=243*)

A: I guess I am not sure if I should change my strategy of choosing a lower-ish cutoff to choosing a slightly higher one (*t=268*)

C: May the force be with you (*t=298*)

Session-18, Group-6

B: k, so is this supposed to be like economics omegle? (*t=30*)

C: basically (*t=46*)

B: anyone had an econ course yet? (*t=56*)

B: I'm minoring in it (*t=62*)

A: My strategy is to start my cutoff highest in the first round, usually around 50. Then I decrease it as the rounds go on, to 40 then to 30 or something like that. (*t=78*)

B: I figure this: (*t=92*)

B: 25% for both 1 and 100 (*t=98*)

A: But sorry no I have not taken any econ class in college. (*t=98*)

B: after that, 50% for above 50 (*t=116*)

C: Is it just me or is there more like a 75% chance of a 1 (*t=121*)

B: there is a 26% chance for a 1 (*t=134*)

B: 26% for 100 (*t=144*)

C: Right i get that but there's been like 3 ones in each group i've been in (*t=151*)

B: the rest is random (*t=153*)

B: Yea, and I've been denied seeing my number like 30 times (*t=171*)

C: Like multiple 1's so even when I switch a 1 I get a 1 haha (*t=177*)

B: it's just bad luck haha (*t=181*)

B: happened to me to (*t=188*)

B: too* (*t=190*)

A: Same. (*t=194*)

B: yo who's tryna get some food after? lol (*t=213*)

B: okay but just go 50 and then u have a 2/3rds chance of getting above your cutoff (*t=240*)

C: okie will do (*t=256*)

B: econ minor, business major, had stats, those are my qualifications xDD (*t=274*)

A: v hungry but I have class after :(But basically if you haven't seen the number by the third time but like 2 as your cutoff so we dont get 10 cents plz (*t=283*)

Session-19, Group-1

B: So... does anybody have any strategies? (*t=20*)

B: Or is this all up to chance? (*t=49*)

C: i haven't found any strategic way to go about this...just been picking a number i would be ok with getting (*t=95*)

B: Same (*t=115*)

A: ive been picking low numbers and it's been turning out well for me.. (*t=128*)

B: How low? (*t=137*)

A: like single digits (*t=149*)

A: but i feel like theres an element of teamwork here (*t=167*)

A: maybe two people pick low and one pick high? (*t=174*)

A: honestly i have no idea.. (*t=181*)

B: We could try that (*t=187*)

C: ya works for me (*t=196*)

B: So, i'll pick low I guess (*t=206*)

A: haha ok (*t=212*)

C: ill go high (*t=226*)

A: i can pick low (*t=228*)

C: how high do you guys want me to go? (*t=239*)

A: have you guys been picking high or low and how has that been working out for you? (*t=252*)

B: I've been picking 40 and have gotten mixed results (*t=266*)

C: ive been going around 30-40 and its been going pretty good (*t=269*)

A: ok so maybe two high and one low (*t=279*)

C: when we say high what are we talking about (*t=294*)

A: because low has been real good (*t=294*)

Session-19, Group-2

C: What have you all been making your cutoffs? (*t=42*)

B: 30 fam (*t=66*)

A: I have been making mine at 50 (*t=71*)

C: i have been doing 50 too (*t=82*)

B: amateurs lol (*t=276*)

Session-19, Group-3

C: what cutoff's have you guys been using (*t=37*)

A: i used 30 for every one (*t=50*)

C: i used 60 (*t=57*)

B: I've been using 40 every time and it's been working pretty well (*t=60*)

A: yea i feel like the lowerish ones have pretty good outcomes (*t=82*)

C: alrighty so what do you guys suggest we use? 30,40? (*t=116*)

A: 35? (*t=126*)

C: sweet sounds good to met (*t=143*)

C: me* (*t=149*)

B: yeah that works (*t=155*)

C: lets make some cash $$$ (*t=290*)

Session-19, Group-4

C: hello (*t=25*)

B: hi (*t=28*)

A: hi (*t=30*)

B: i say we make cut off 55 (*t=36*)

B: or 60 (*t=56*)

B: idk (*t=57*)

A: i have been doing 60 or 65 (*t=64*)

C: ive been going lower (*t=73*)

B: like what (*t=84*)

C: like 40's (*t=89*)

A: has it been working well (*t=94*)

C: yeah (*t=98*)

C: but we could do 55 (*t=113*)

C: that makes sense (*t=118*)

A: same here so i say 55 (*t=125*)

B: okay perfect (*t=136*)

B: good talk (*t=141*)

A: good luck ppl (*t=150*)

B: thx u too (*t=155*)

B: hopefully we all make $$$$$$$$$$$ (*t=167*)

A: in desperate need of it always (*t=182*)

B: #collegelife (*t=191*)

Session-19, Group-5

A: Hi (*t=19*)

B: hey (*t=23*)

C: Sup (*t=27*)

C: I usually go 75 50 25 (*t=39*)

B: i do 57, 54, 50 (*t=60*)

B: #risky (*t=70*)

C: i guess the goal is to get the lowest possible value into the computers hands (*t=75*)

A: I've been doing (*t=76*)

A: 50 (*t=84*)

A: how can we do that (*t=122*)

B: pray (*t=148*)

C: Basically lol (*t=158*)

C: There's usually one ball thats pretty low (Like sub 10) right? (*t=182*)

A: yeah, but i have seen a couple low balls in one cycle (*t=216*)

B: yeah. i can only remember one instance where it was like 100 100 99 68 (*t=220*)

B: one for me was 1 1 1 1 (*t=240*)

A: i guess the smaller the cutoff we do the better for all of us for the most part? (*t=276*)

B: well, do good guys! maybe 30 as cutoff? (*t=287*)

B: or 20? (*t=295*)

C: 31 (*t=297*)

A: sounds good to me! (*t=298*)

C: :P (*t=299*)

Session-19, Group-6

C: Any ideas? (*t=34*)

B: Initial thoughts? (*t=37*)

A: 50? (*t=67*)

B: 35? (*t=78*)

C: Yeah 35 sounds good, so a lower value is switched out (*t=121*)

A: Cool (*t=129*)

B: It has worked well for me in previous cycles (*t=155*)

A: I've been doing well with 65 (*t=176*)

A: But 35 is more safe (*t=188*)

B: How many .10 vs 100? (*t=193*)

C: I agree with 35 (*t=213*)

B: 35 it is (*t=240*)

A: Sounds good (*t=248*)

Session-20, Group-1

A: hi guys (*t=18*)

B: hello (*t=23*)

C: hi (*t=24*)

A: what do we want to do for cutoffs (*t=25*)

B: what have u guys been doing prior (*t=37*)

B: ive been doing 70 (*t=42*)

C: i've been keeping mine pretty low, around 40 (*t=53*)

A: i think we should start out higher and then go lower in the rounds (*t=61*)

C: yeah i agree (*t=71*)

B: that would work (*t=77*)

A: so first round we should do 70? (*t=98*)

B: im good with that (*t=111*)

C: okay (*t=113*)

B: then have like 40 be our lowest (*t=126*)

A: ok (*t=130*)

C: sounds good (*t=134*)

A: so never go lower than 40? (*t=162*)

C: yeah maybe 40 for the last round? (*t=176*)

B: how many rounds are there again? (*t=185*)

C: 3 (*t=194*)

A: three (*t=194*)

A: so 70 for the first, 40 for the last (*t=206*)

A: what about the middle round? (*t=217*)

B: what about the second round (*t=220*)

B: 55?? (*t=228*)

C: somewhere in the middle (*t=232*)

A: that works (*t=232*)

C: yeah (*t=234*)

B: okay lol (*t=236*)

A: let's hope were lucky haha (*t=245*)

A: or one of us is I guess (*t=257*)

C: haha yeah (*t=263*)

C: good luck guys (*t=289*)

A: same to you (*t=294*)

B: you too! (*t=299*)

Session-20, Group-2

A: Hello world! (*t=12*)

B: hey (*t=37*)

C: Sup (*t=43*)

A: Does anyone have a good strategy for this? (*t=48*)

B: i'm just going to keep doing it how i did before tbh (*t=72*)

A: Yeah thats how I feel too (*t=85*)

C: same (*t=92*)

B: cool cool (*t=108*)

A: since we do not know the order of the chosers, we cannot really make a susinct gameplan (*t=126*)

A: did anyone watch south park on wednesday? (*t=141*)

B: nah i don't like that show (*t=159*)

C: No was it a good episode? (*t=162*)

A: yeah it was pretty funny (*t=178*)

C: I'll have to watch it over the weekend (*t=193*)

A: Im sorry for you member B (*t=194*)

C: Anyone into AHS (*t=203*)

B: nope guess i just live under a rock (*t=227*)

A: I was invited over someones house tonight to watch it but ive never seen it before (*t=228*)

B: t-minus 1 minute thank god (*t=251*)

C: It's really good kind of a creepy show tho (*t=256*)

A: yeah this sucks (*t=258*)

C: yeah I just wanna nap tbh (*t=264*)

A: #dicksoutforharambe (*t=273*)

C: RIP (*t=284*)

B: fuck penn state (*t=289*)

C: ^^^^^ (*t=294*)

Session-20, Group-3

C: greetings (*t=8*)

A: hello (*t=27*)

B: hey (*t=44*)

A: what are you thinking (*t=49*)

A: im not really sure how we help each other (*t=66*)

C: My thought process is to set your first cutoff highest, second cutoff lower than that, and third cutoff lower than that because each successive round there is a greater chance that someone else switched with the computer ball and therefore gave (*t=89*)

A: yeah thats a good idea (*t=109*)

C: so i usually set my fist at 50, second 40, and third about 45 (*t=120*)

C: third about 35*** (*t=126*)

B: well thats good with me (*t=128*)

A: yeah ive been always doing 50 pretty much but thats a better idea (*t=138*)

A: i will do that (*t=150*)

C: have you guys found a different successful strategy (*t=168*)

B: I just kind of eyeball it but If thats been working for you (*t=176*)

C: i dont think we're allowed to say specifically but yes it's been working for me (*t=200*)

A: okay good (*t=208*)

C: best of luck (*t=215*)

A: thanks you too (*t=220*)

Session-20, Group-4

C: if you're the first mover set your threshold pretty high (*t=42*)

C: like around 60-75 (*t=50*)

C: if you're second mover set it around 40-50 (*t=63*)

A: and go down depending on what mover you are (*t=65*)

C: and if you're third set it at like 25 (*t=74*)

C: ˆwhat team member A said (*t=85*)

C: Team Member B you got it? (*t=113*)

B: yes (*t=119*)

C: alright cool (*t=126*)

A: sounds good (*t=149*)

C: about to make some $$$$$$$$$ (*t=276*)

A: let's hope (*t=293*)

C: we got this (*t=297*)

Session-20, Group-5

A: hi (*t=9*)

B: hey (*t=29*)

C: hi (*t=50*)

A: does anyone have a stradegy (*t=54*)

C: what do you think our cutoffs should be (*t=59*)

A: I've just been using 50 (*t=76*)

B: same (*t=87*)

A: do you want to use that or something else (*t=118*)

B: sounds good to me, C are you good with it? (*t=144*)

C: I've actually used a cutoff of 2 for a lot of the rounds, seeing as that anything is better than a $0.10 payoff (*t=166*)

C: but i'm good with 50 if you guys want to do that (*t=179*)

A: has 2 worked for you for the most part (*t=197*)

C: yeah for the most part, because when you use 50 i feel like you end up losing a lot of opportunities to get more than $0.10 (*t=239*)

B: using 50 I dont think I got .10 once and I got about 5 $10.00 (*t=246*)

C: up to you guys though (*t=252*)

A: doesnt matter to me (*t=270*)

B: how did 50 work for you A? (*t=273*)

A: i only got .10 twice (*t=292*)

C: okay we can do 50 (*t=297*)

A: so good i guess (*t=297*)

Session-20, Group-6

C: how should we choose (*t=41*)

B: I have no idea (*t=65*)

C: i guess just make the cut off higher rather than lower? (*t=88*)

A: just make sure we get a number larger than 1 (*t=97*)

B: yeah thats what i was thinking. Make the cutoff a bit higher than usual. (*t=122*)

C: okay sounds good (*t=134*)

C: i feel like its all random anyway (*t=210*)

A: yeah theres no way to predict anything (*t=227*)

B: Lets just hope the person selected gets 100 off the bat (*t=260*)

C: yeah (*t=278*)

B.5 One-Parameter Reinforcement Learning Model

In this section we use a one-parameter reinforcement learning model, based on Erev and Roth (1998) and Luce (1959), to explain the differences in behavior across some of our treatments for subjects classified as *stationary*. Figure F.1 plots the average round-1 cutoff for supergames 6-20 for treatments Selection, S-Across, S-Peer, and S-Explicit. As is clear from the graphs, there is a marked difference in the cutoff choices between S-Explicit and the other treatments. Figure F.2 plots the same averages, but combining the treatments Selection, S-Across and S-Peer.

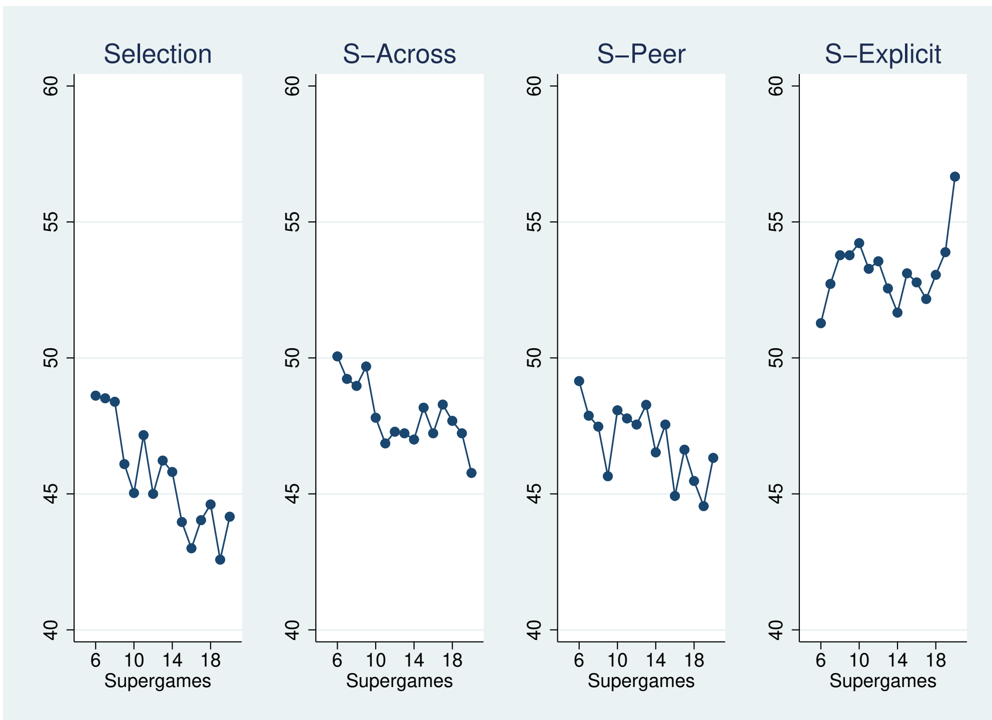


Figure B.3: Average Cutoff for Stationary Subjects in Round 1

Note: Subjects are classified as *stationary* if their cycle-21 cutoffs satisfy $|\mu_1 - \mu_2| \leq 2.5$ and $|\mu_1 - \mu_3| \leq 2.5$, where μ_t is the cutoff value chosen in round t.

One possible reason for the observed difference is the type of learning available for subjects in each case. In the S-Explicit treatment, the adverse selection was implemented through a time-dependent rule, with the rematching pool comprised of all three balls in round 1, the two lowest-value balls in round 2, and the lowest-value ball in round 3. For this reason, there were no mover types in treatment S-Explicit, and, since half of the time subjects make a decision in round 1, there is an overall lower chance of observing a bad outcome after switching in S-Explicit compared to other treatments with first, second, and third movers.

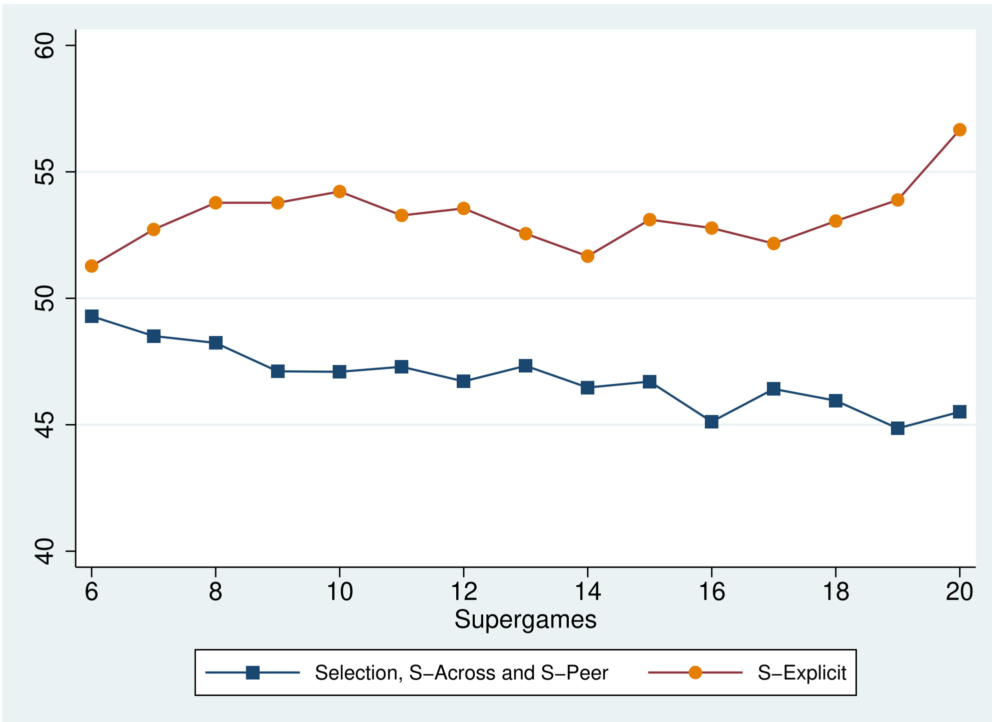


Figure B.4: Average Cutoff for Stationary Subjects in Round 1

In what follows, we estimate a simple one-parameter reinforcement model that best fit the data for treatments *Selection, S-Across,* and *S-Peer* using data only for subjects classified as *stationary*. Later, we use the optimal parameter estimate to *predict* round-1 cutoff values for the *S-Explicit* treatment, again restricting the analysis to the *stationary* subjects.

B.5.1 The Model

In supergame s = 6, each player n has an initial propensity to choose the k-th cutoff value, where $k \in [1, 100]$, given by a non-negative number $q_{nk}(1)$. We assume that each player has an equal propensity for each of the possible cutoffs, such that:

$$q_{nk}(1) = q_{nj}(1) \quad \forall n \text{ and } \forall k \tag{1}$$

The reinforcement of receiving payoff x is given by the identity function:

$$R(x) = x \tag{2}$$

Suppose that player n chooses cutoff k in supergame s. For supergame s+1, he adjusts the propensity of his j^{th}-cutoff according to:

$$q_{nj}(s+1) = \begin{cases} q_{nj}(s) + R(x), & \text{if } j = k \\ q_{nj}(s), & \text{otherwise} \end{cases} \tag{3}$$

Finally, the probability that player n chooses the k^{th} cutoff in supergame s is given by:

$$p_{nk}(s) = \frac{q_{nk}(s)}{\sum_{j=1}^{100} q_{nj}(s)} \tag{4}$$

Equation (4) is Luce's linear probability response rule. Note that, even though we assume that every cutoff has the same propensity at s = 6, we made no assumptions about the sum of propensities, which appear in the denominator of equation (4). This is the only parameter of the model, and, together with the size of the payoffs, it determines the speed of learning.

Let X be the average payoff of all players in all four treatments. The parameter s(1), which is assumed to be the same for all players, is given by:

$$s(1) = \frac{\sum_{j=1}^{100} q_{nj}(1)}{X} \tag{5}$$

which, together with (4), implies that the initial propensities are:

$$q_{nj}(1) = p_{nj}(1)s(1)X \tag{6}$$

Both $p_{nj}(1)$ and X are known: the first by assumption, and the second from the data. All we need is an estimate for s(1). The value that minimizes the Mean Squared Deviation (MSD) using data from Selection, S-Across, and S- Peer is s(1)=0.162, which indicates an extremely fast learning speed.

Figure F.3 presents the same graph as in Figure F.2, but also plots the *fitted* values for *Selection, S-Across,* and*S-Peer* and the *predicted* values for *S-Explicit.*

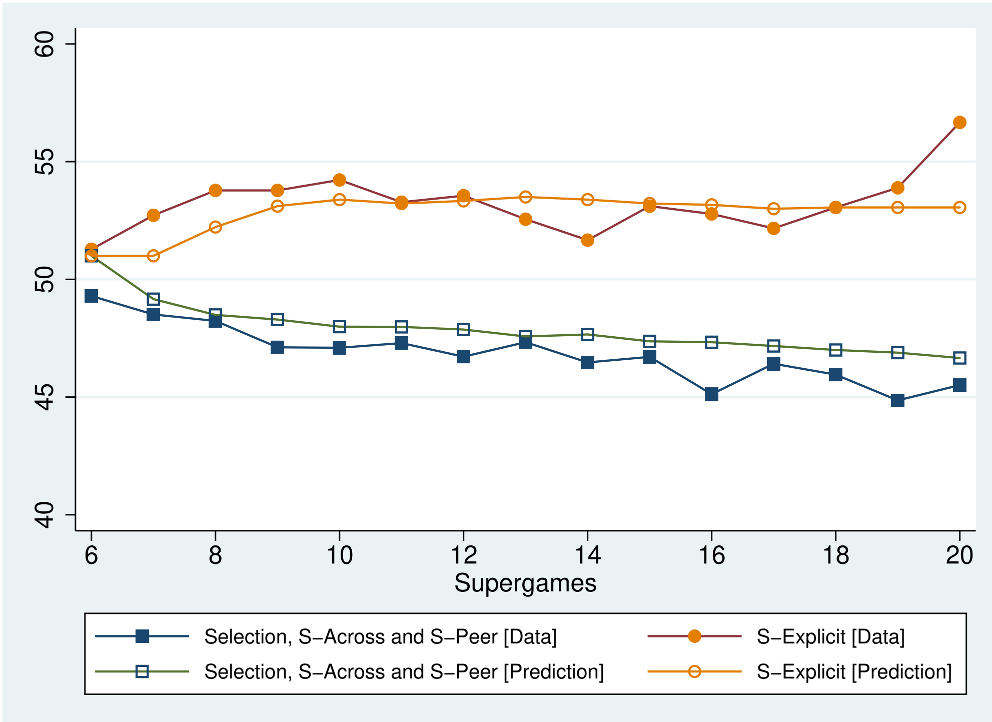


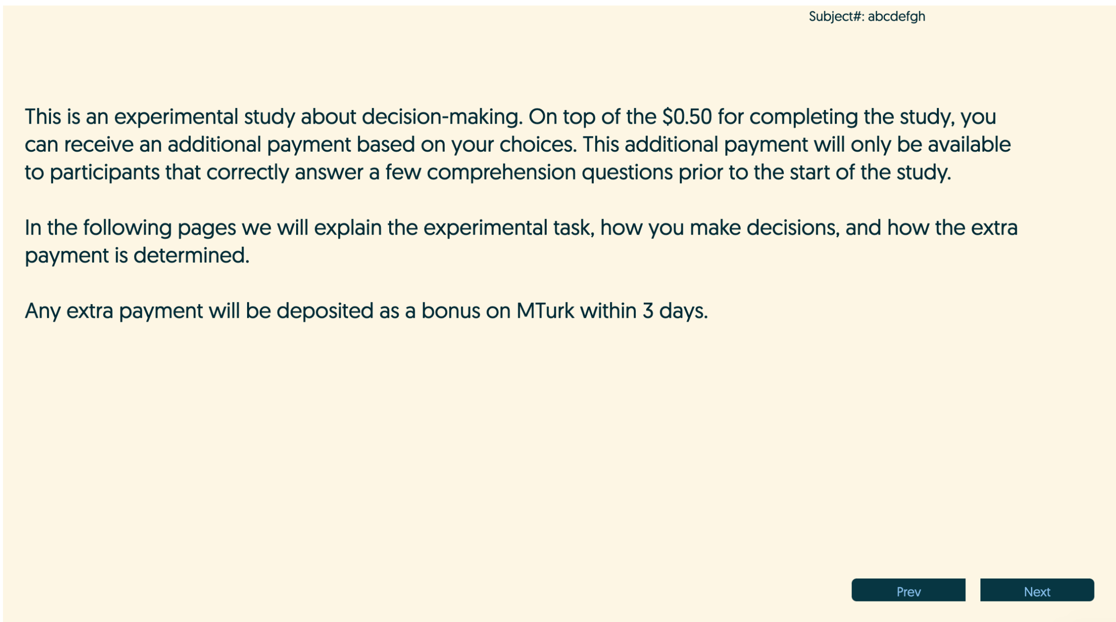
Figure B.5: Average Cutoff for Stationary Subjects in Round 1, Data and Prediction $(s(1) = 0.162)$

Note that both in S-Explicit and in the other treatments, subjects start choosing very similar cutoffs. The subsequent experience with realized payoffs, however, is different, which accounts for the increasing profile of S-Explicit cutoff choices and the decreasing profile of cutoff choices in the other treatments.

Appendix C Appendix for Chapter 4

C.1 First Stage Experiments

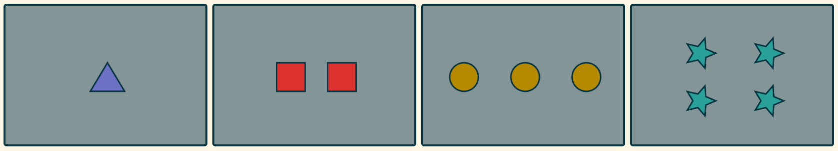
Screenshots for first-stage experiments. Different states of the same stage are indicated by sub-captions.



Subject#: abcdefgh

The study will use cards containing a certain number (1, 2, 3, or 4) of a particular symbol (triangle, square, circle, or star) with a specific color (purple, orange, red, or blue). Considering all possible values for [number, symbol, color] we end up with a deck of 4x4x4=64 unique cards.

As an example, four of those cards are shown below:



Prev Next

Subject#: abcdefgh

Even though the entire deck has 64 cards, during the experiment you will work with 12 cards at a time. Your goal is to use those 12 cards to form groups of 3 cards satisfying a simple property. Any combination of 3 cards satisfying that property is called a SET.

The property that defines a SET is this: A SET is any combination of 3 cards such that each attribute (number, symbol, and color) is either:

- the same for **all** three cards, or,
- is different for **each** of the three cards.

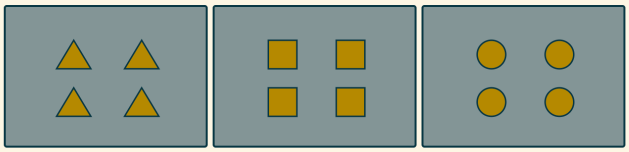
The next few pages show examples of 3-card combinations that **do** and **do not** form a SET.

Prev Next

Subject#: abcdefgh

Example 1

The following 3 cards form a SET:



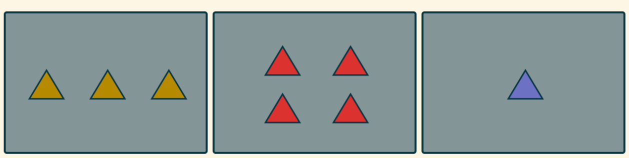
That's because all cards have the **same** color (orange), the **same** number of symbols (4), and **different** symbols (from left to right: triangle, square, and circle).

Prev Next

Subject#: abcdefgh

Example 2

The following 3 cards also form a SET:



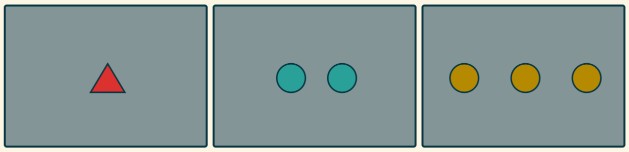
And that's because all cards have the **same** symbol (triangle), **different** colors (from left to right: orange, red, and blue), and **different** number of symbols (from left to right: 3, 4, and 1).

Prev Next

Subject#: abcdefgh

Example 3

The following 3 cards, on the other hand, do **NOT** form a SET:



Even though the middle and right-most cards have circles, the left-most card has a triangle. Hence, for the attribute symbol, the cards are neither all the same nor all different. Note that the cards are all different with respect to color and number.

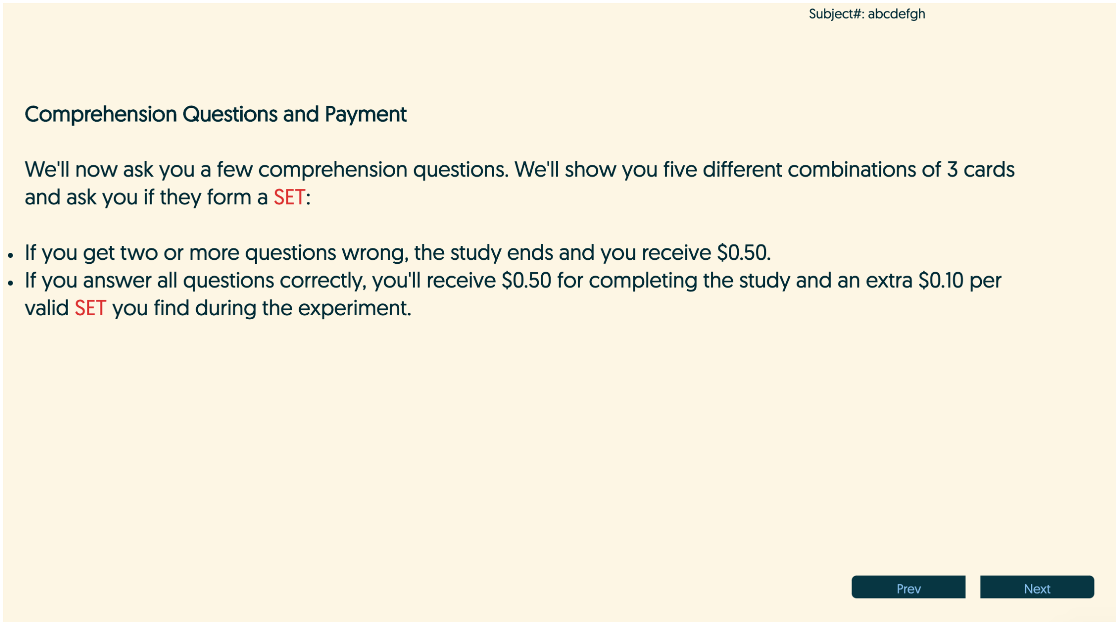
Prev Next

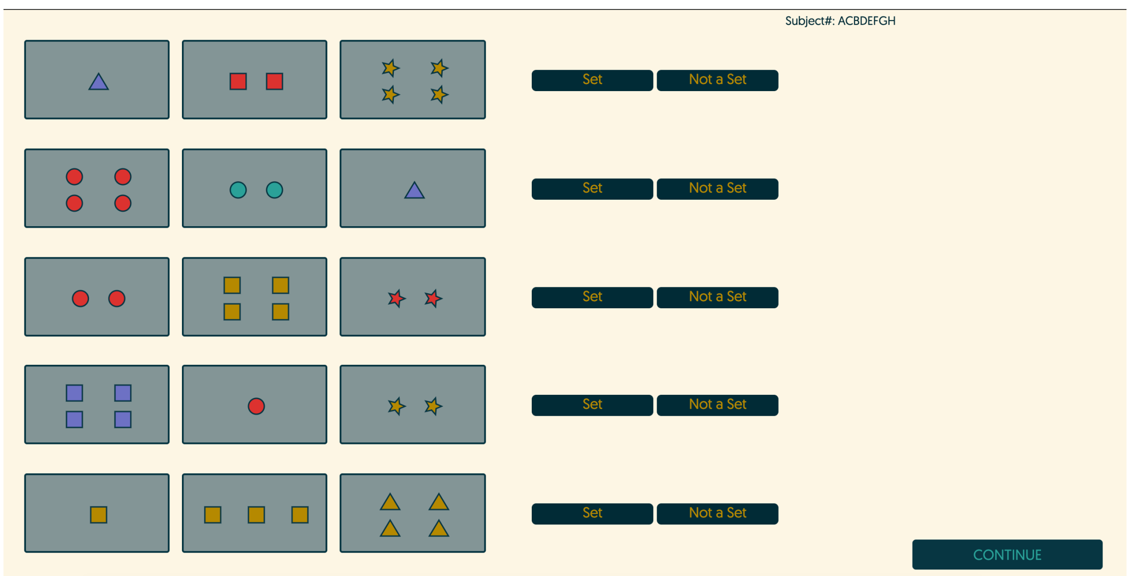
Subject#: abcdefgh

There will be **two** rounds in the experiment, and in each round you will see a 12-card deck that was randomly selected from a larger list of 12-card decks. You will have 120 seconds to form SETs in each round. You can form a SET by clicking on the cards to select, and you can click a selected card to un-select it.

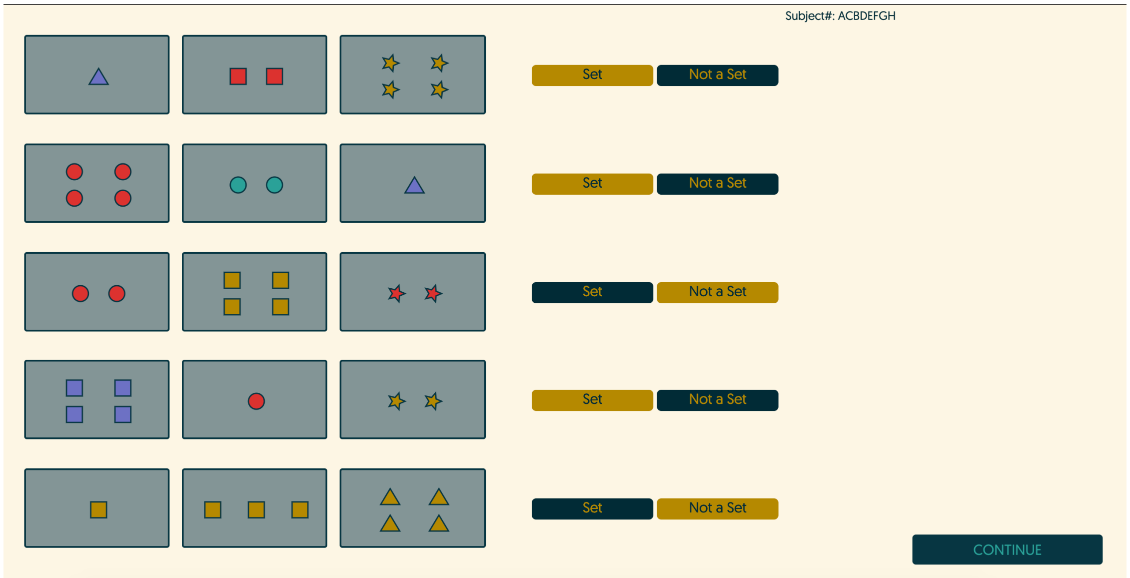
If you select 3 cards that form a SET, it will be registered on the right side of your screen. If you select 3 cards that do not form a SET, you will incur a 5-second penalty. And if you select the same SET more than once, you will also incur a 5-second penalty. Therefore, you should not randomly click on cards in the hopes of finding a SET.

Prev Next

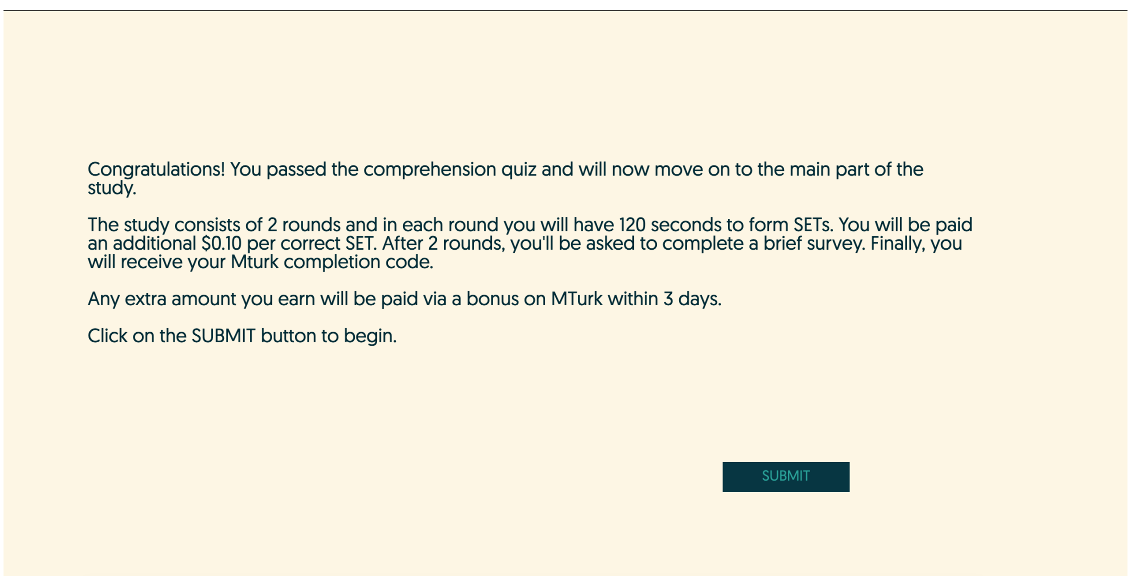


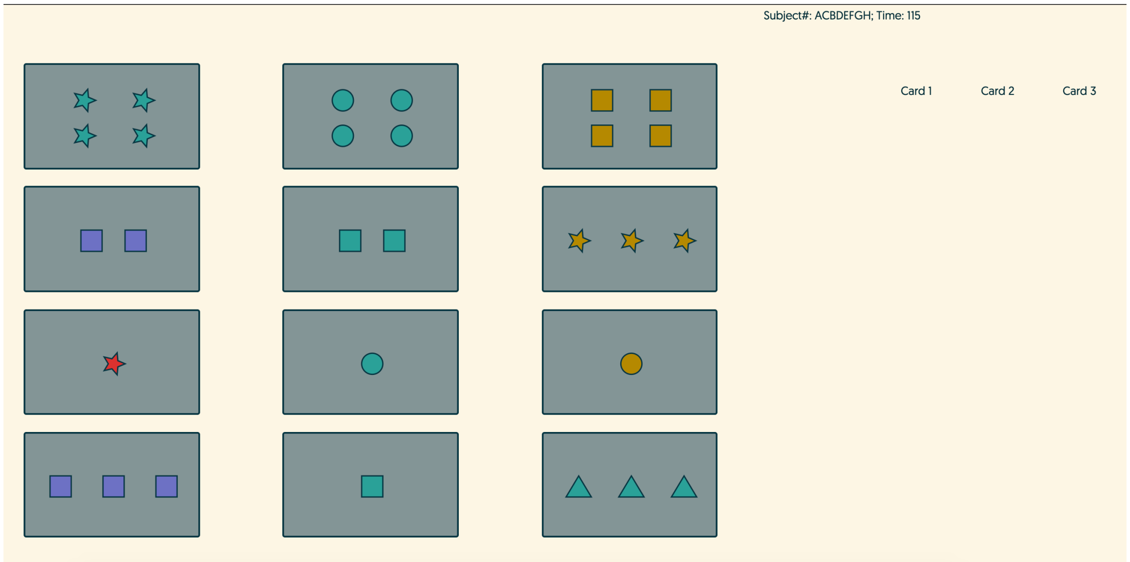


(A-1) Quiz stage with no answers

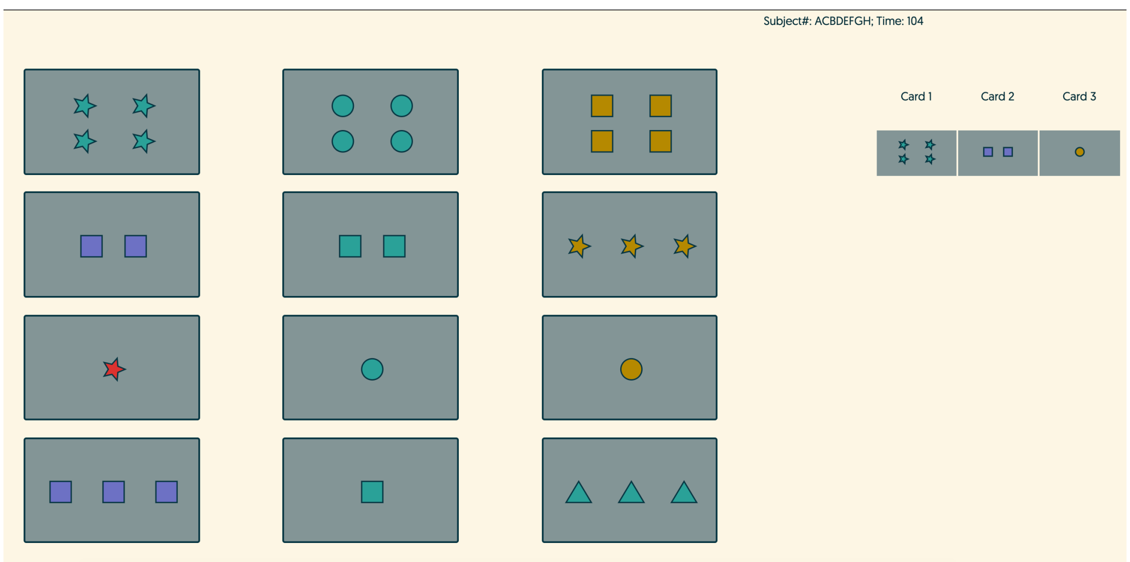


(A-2) Quiz stage with 4 correct answers

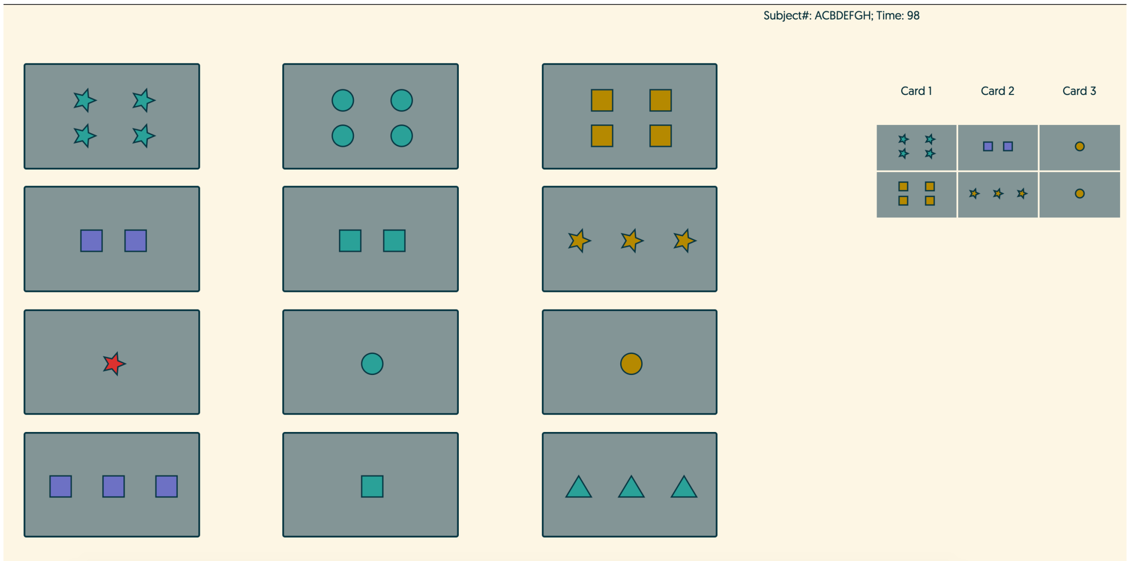




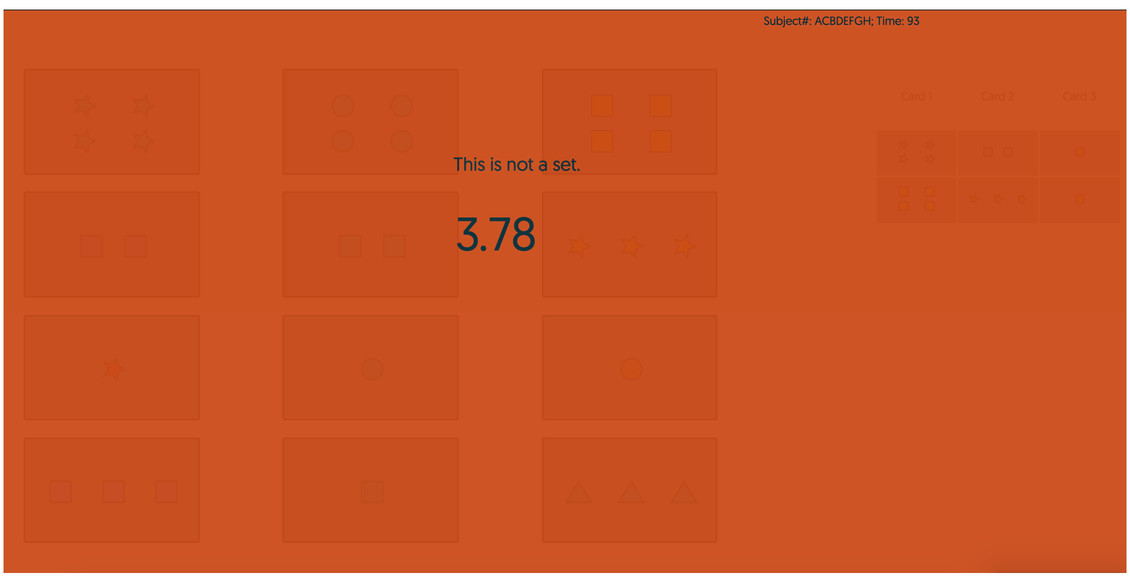
(B-1) Initial screen for round 1



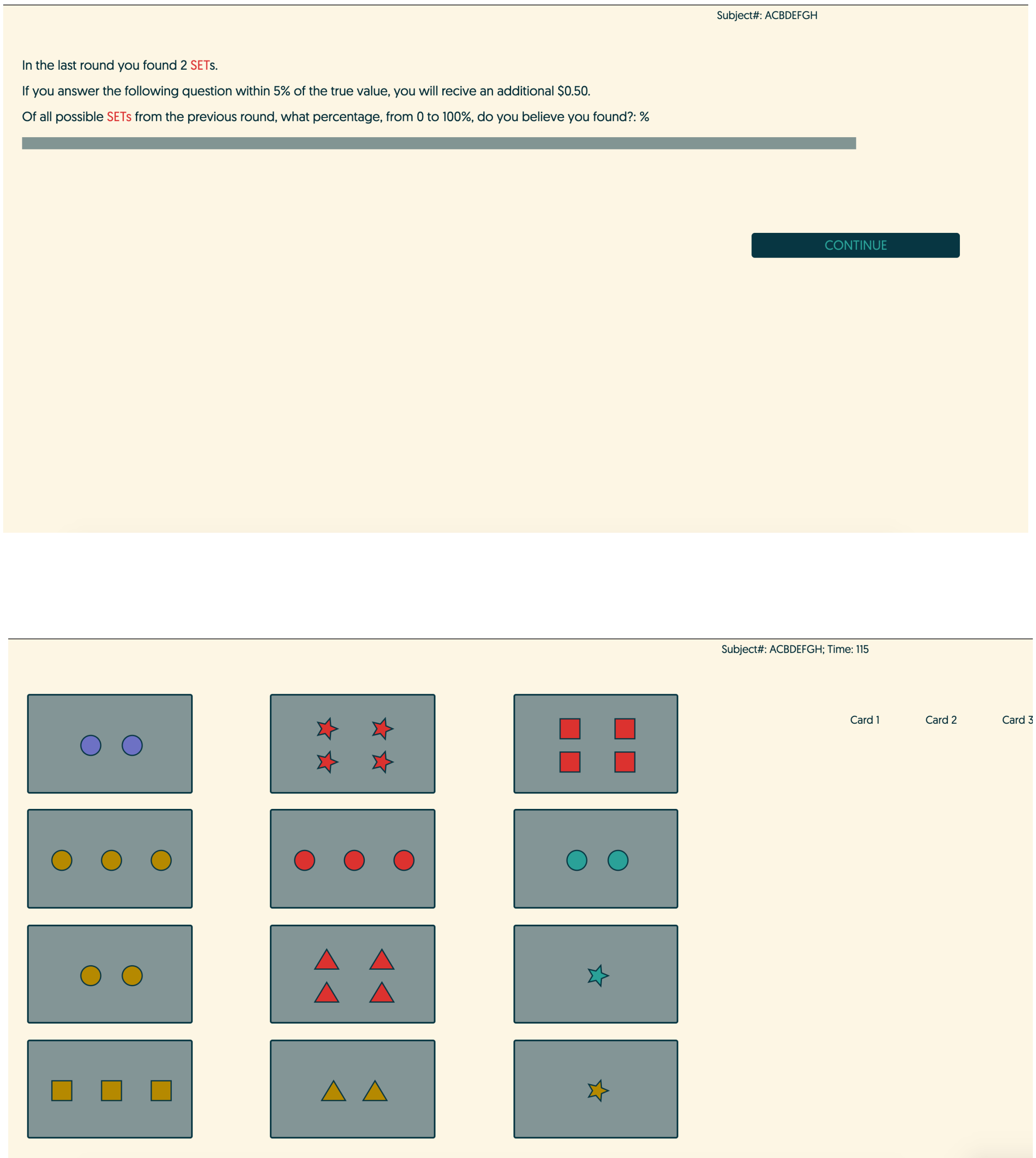
(B-2) Screen indicating 1 set found



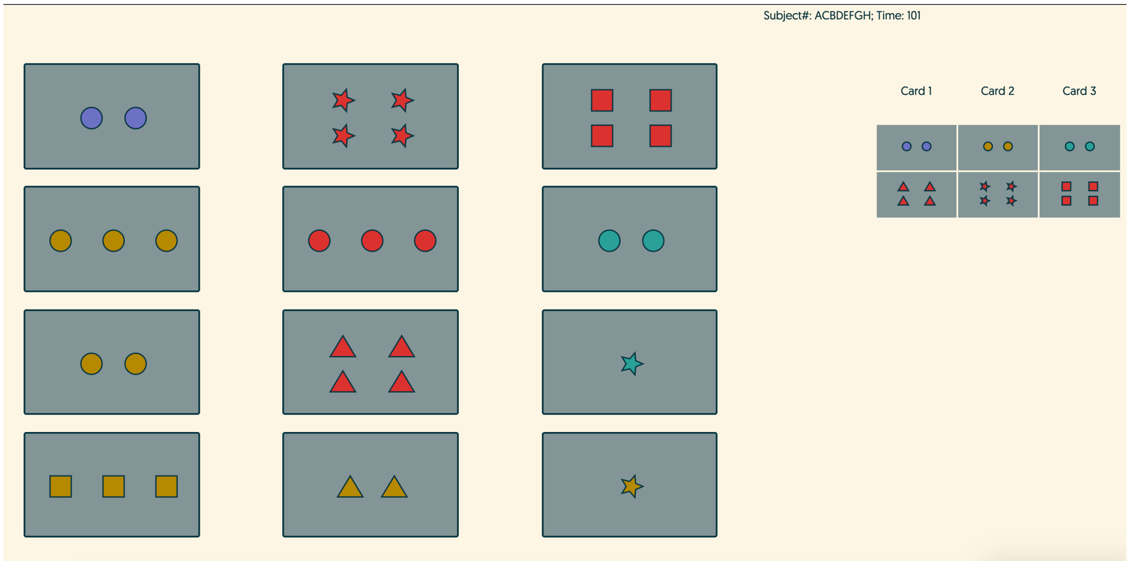
(B-3) Screen indicating 2 sets found



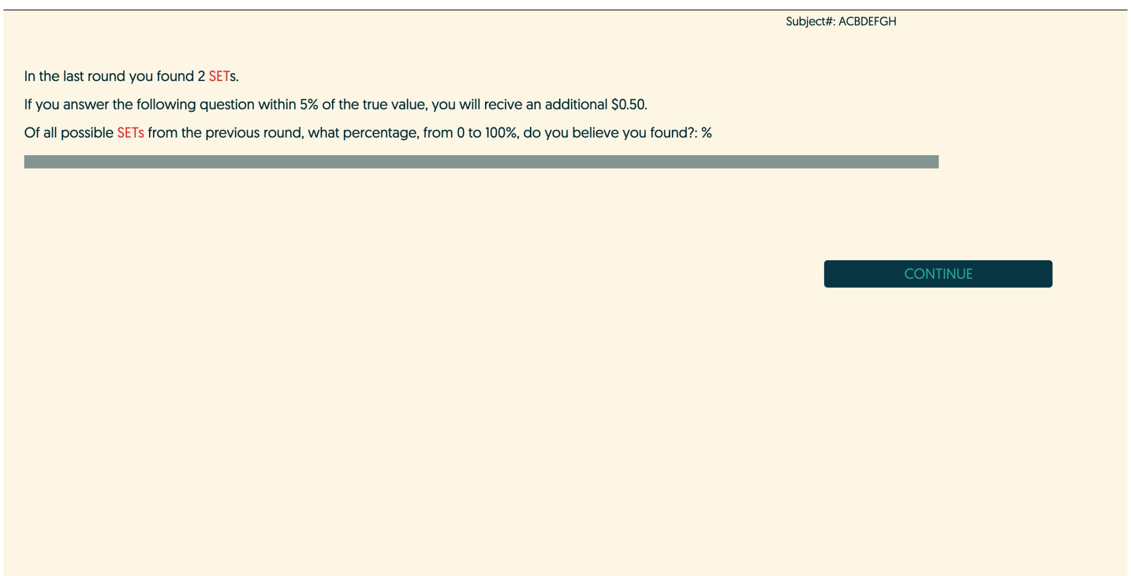
(B-4) Example of 5-second time penalty

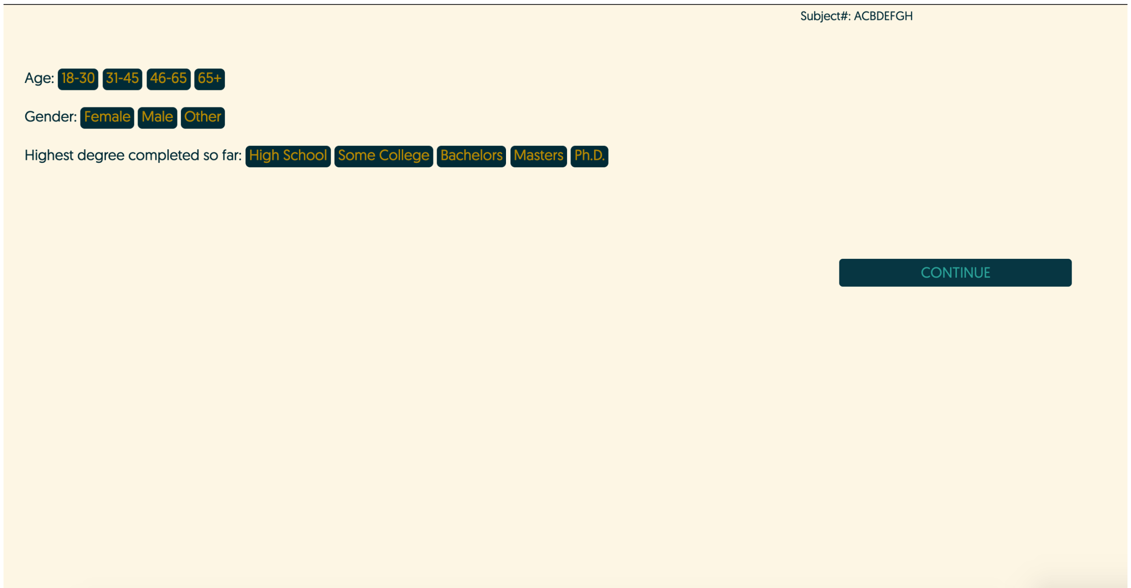


(C-1) Initial screen for round 2



(C-2) Screen indicating 2 sets found




Subject#: ACBDEFGH

Age: 18-30 31-45 46-65 65+

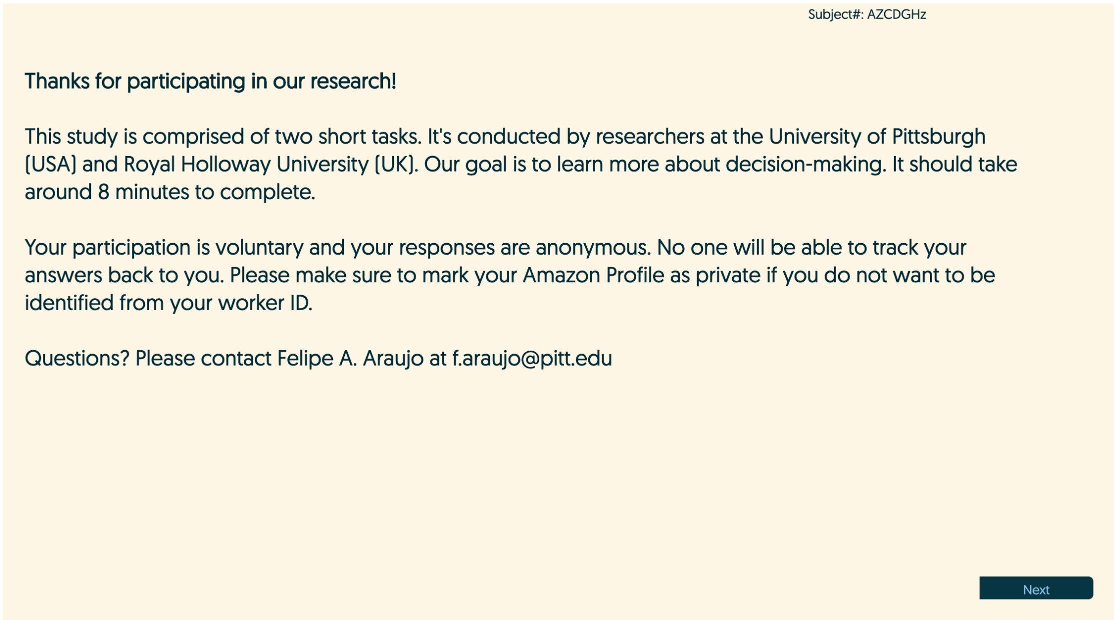
Gender: Female Male Other

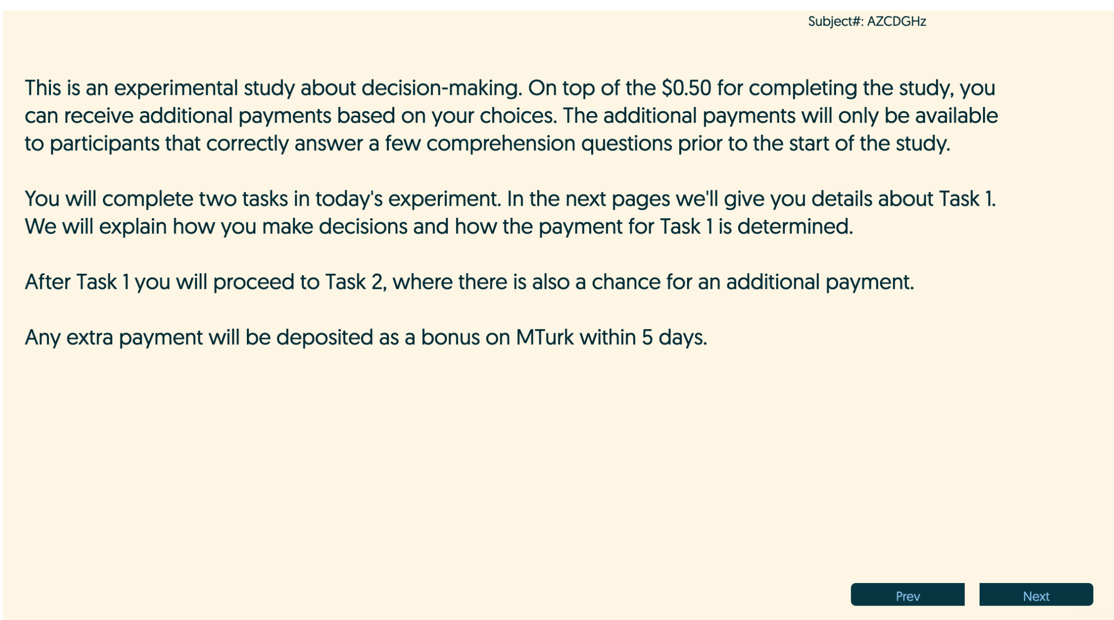
Highest degree completed so far: High School Some College Bachelors Masters Ph.D.

CONTINUE

C.2 Main Experiments

Screenshots for main experiments. Different states of the same stage — and different treatments — are indicated by sub-captions.

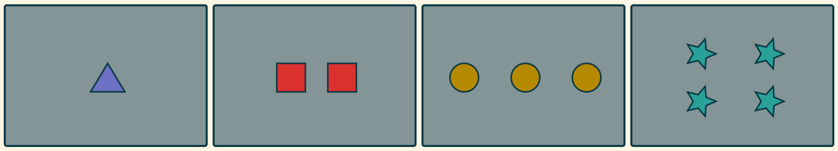




Subject#: AZCDGHz

Task 1 will use cards containing a certain number (1, 2, 3, or 4) of a particular symbol (triangle, square, circle, or star) with a specific color (purple, orange, red, or blue). Considering all possible values for [number, symbol, color] we end up with a deck of 4x4x4=64 unique cards.

As an example, four of those cards are shown below:



Prev Next

Subject#: AZCDGHz

Even though the entire deck has 64 cards, during Task 1 you will work with 12 cards. Your goal is to use those 12 cards to form groups of 3 cards satisfying a simple property. Any combination of 3 cards satisfying that property is called a SET.

The property that defines a SET is this: A SET is any combination of 3 cards such that each one of the attributes (number, symbol, and color) is either:

- the same for **all** three cards, or,
- is different for **each** of the three cards.

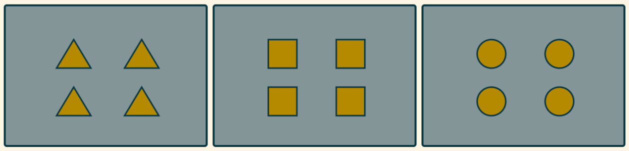
The next few pages show examples of 3-card combinations that **do** and **do not** form a SET.

Prev Next

Subject#: AZCDGHz

Example 1

The following 3 cards form a SET:



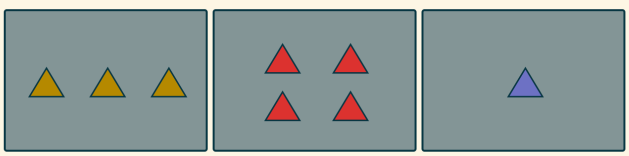
That's because all cards have the **same** color (orange), the **same** number of symbols (4), and **different** symbols (from left to right: triangle, square, and circle).

Prev Next

Subject#: AZCDGHz

Example 2

The following 3 cards also form a SET:



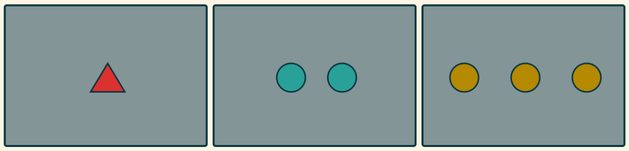
And that's because all cards have the **same** symbol (triangle), **different** colors (from left to right: orange, red, and blue), and **different** number of symbols (from left to right: 3, 4, and 1).

Prev Next

Subject#: AZCDGHz

Example 3

The following 3 cards, on the other hand, do **NOT** form a SET:



Even though the middle and right-most cards have circles, the left-most card has a triangle. Hence, for the attribute symbol, the cards are neither all the same nor all different. Note that the cards are all different with respect to color and number.

Prev Next

Subject#: AZCDGHz

On Task 1, you will see a 12-card deck that was randomly selected from a larger list of 12-card decks, and you will have 120 seconds to form SETs. You can form a SET by clicking on the cards to select, and you can click a selected card to un-select it.

If you select 3 cards that form a SET, it will be registered on the right side of your screen. If you select 3 cards that do not form a SET, you will incur a 5-second penalty. And if you select the same SET more than once, you will also incur a 5-second penalty. Therefore, you should not randomly click on cards in the hopes of finding a SET.

Prev Next

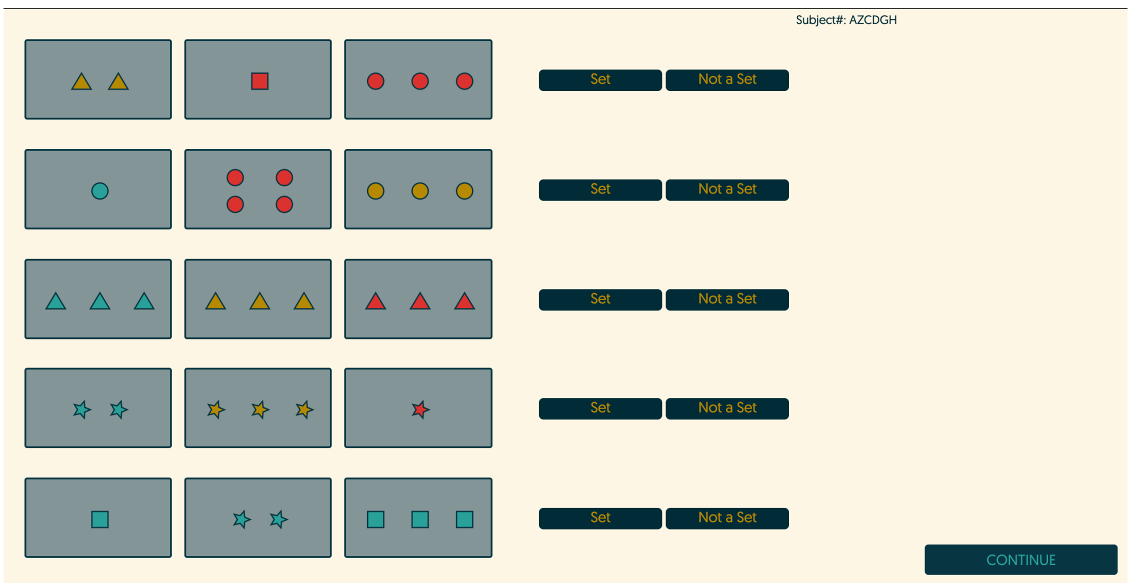
Subject#: AZCDGHz

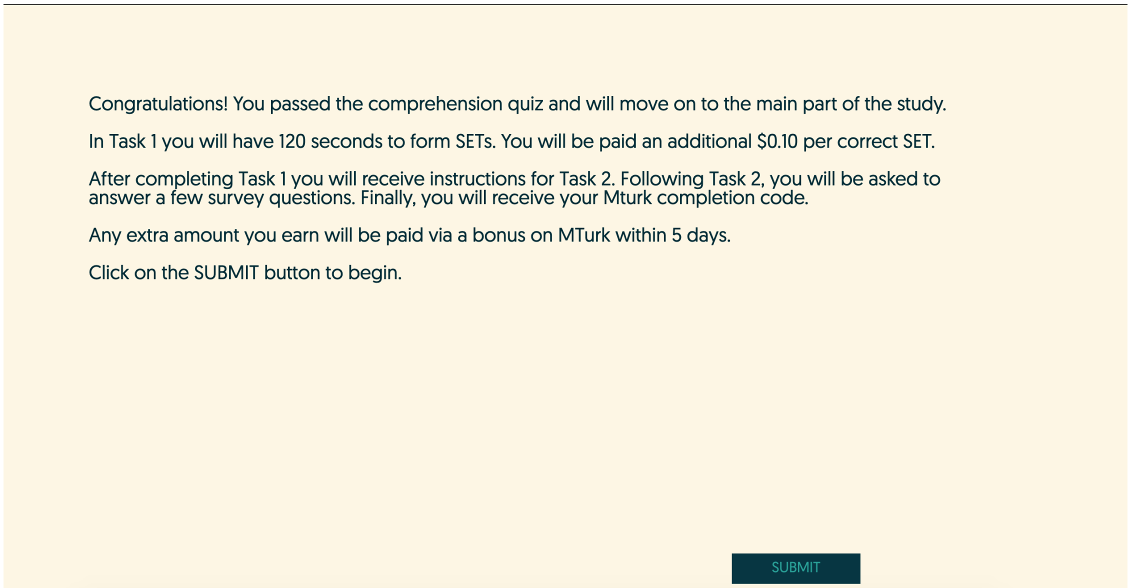
Comprehension Questions and Payment

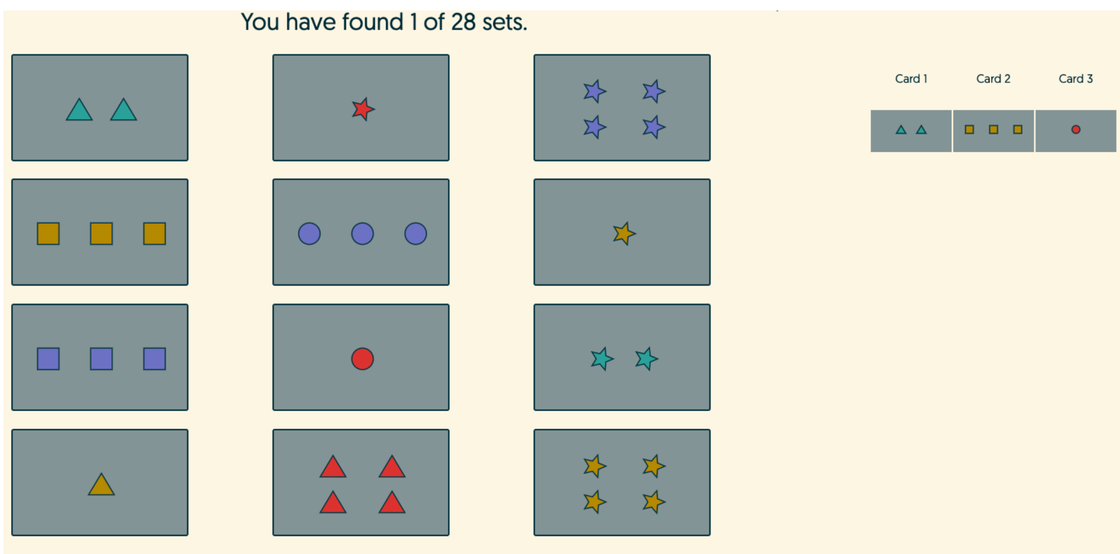
We'll now ask you a few comprehension questions. We'll show you five different combinations of 3 cards and ask you if they form a SET:

- If you get two or more questions wrong, the study ends and you receive $0.50.
- If you answer at least four of the five questions correctly, you'll receive $0.50 for completing the study and an extra $0.10 per valid SET you find during Task 1. Moreover, you will also have a chance for an additional payment on Task 2.

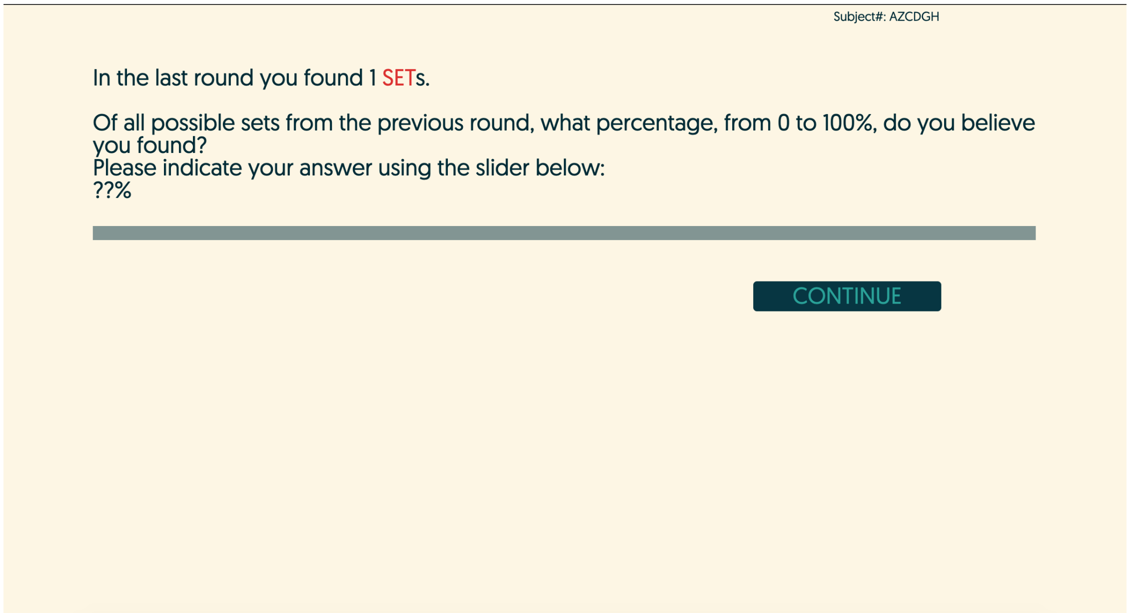
Prev Next



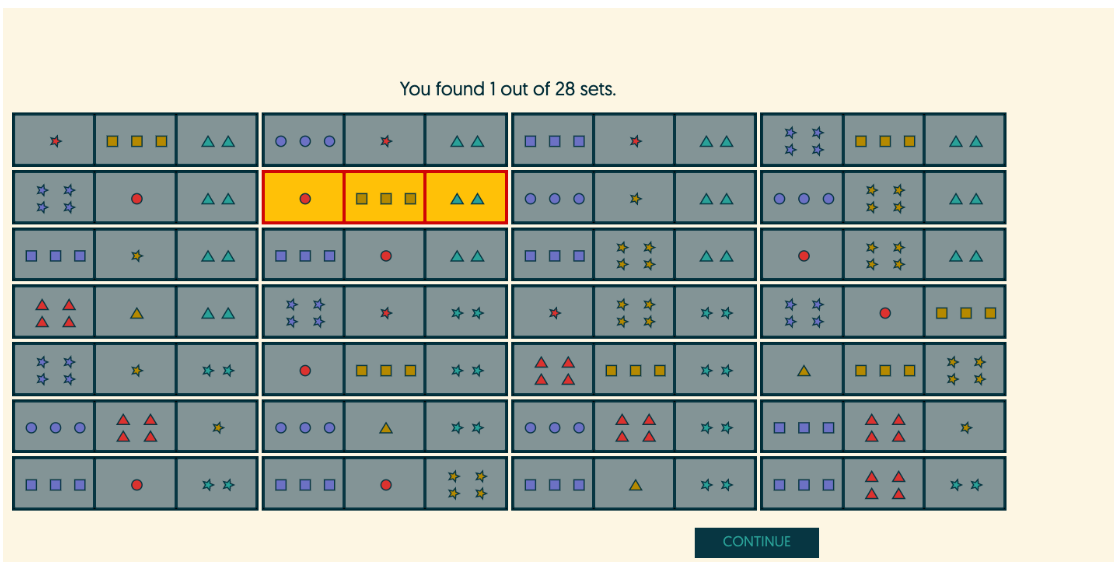




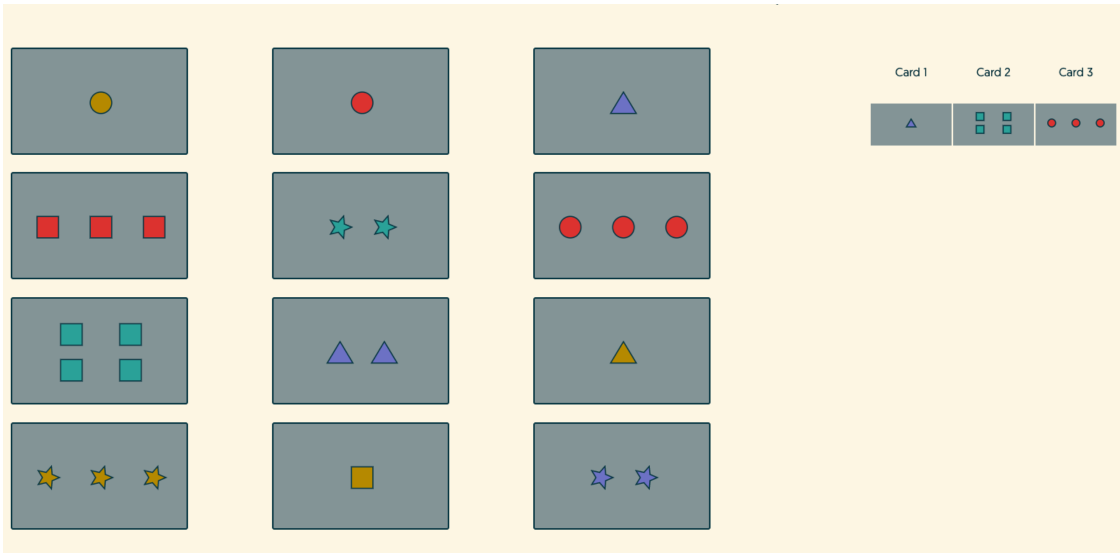
(A-1) Treatment *full information*; 1 set found



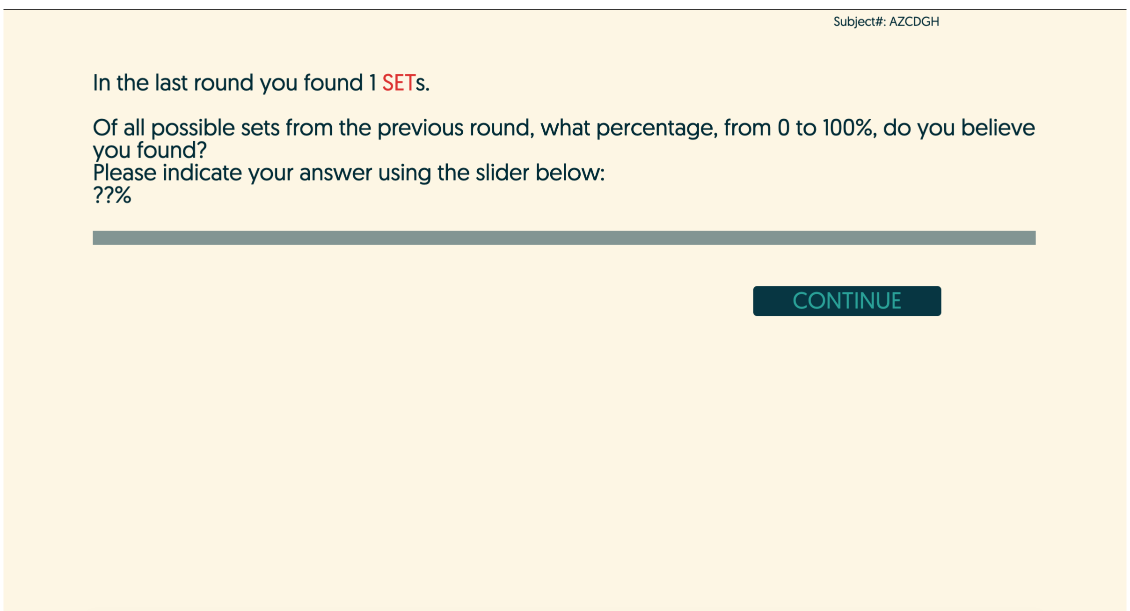
(A-2) Belief elicitation



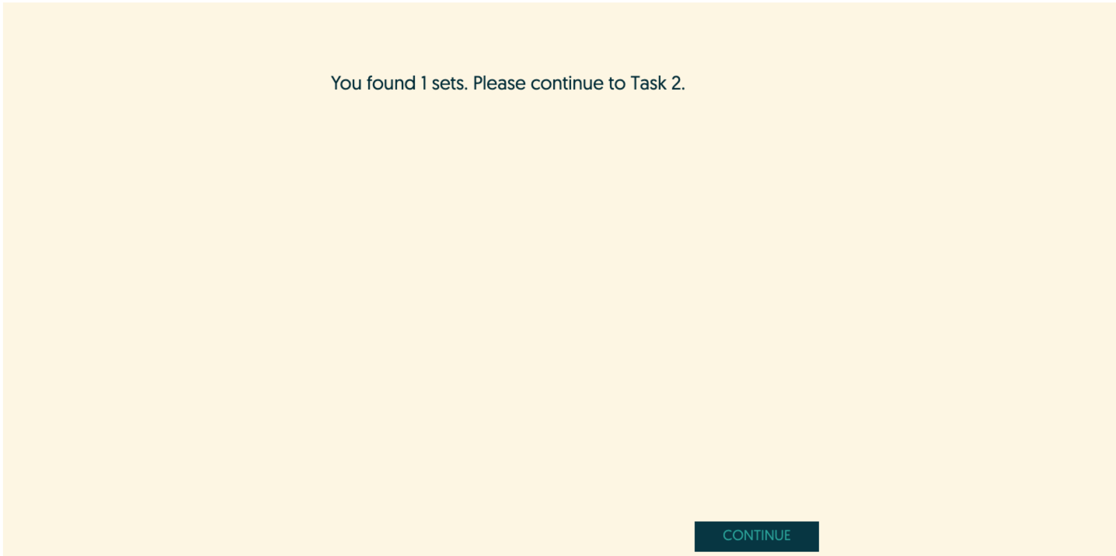
(A-3) Treatment *full information*; feedback screen



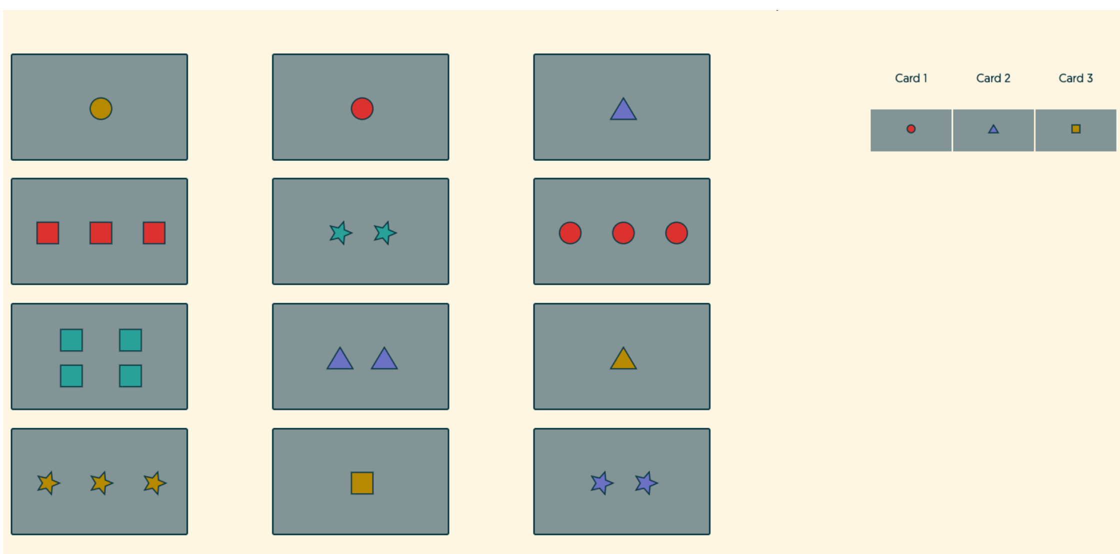
(B-1) Treatment *unawareness*; 1 set found



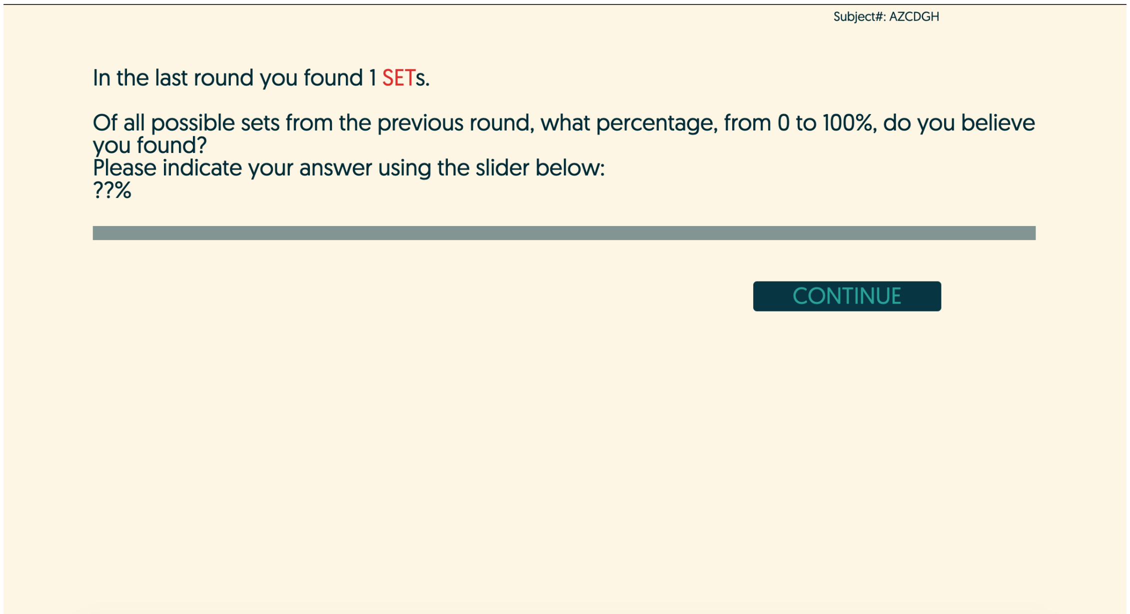
(B-2) Belief elicitation



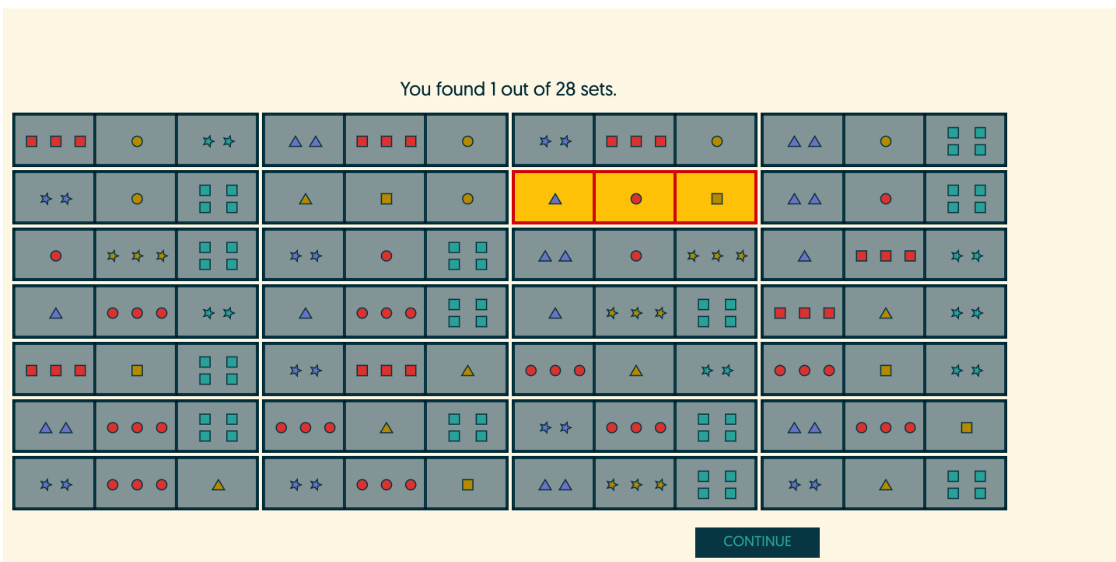
(B-3) Treatment *unawareness*; feedback screen



(C-1) Treatment *unawareness-info*; 1 set found



(C-2) Belief elicitation



(C-3) Treatment *unawareness-info*; feedback screen

Subject#: AZCDGH

Task 2

How does it work?

1. You just received an additional $1.00 for passing the comprehension quiz
2. You must decide how much of this additional $1.00 to **keep safe** and how much to **bet** in a lottery
3. You can bet any amount from $0.00 to $1.00

How do you earn money?

The amount of money you make will depend on your choices and on chance. You never lose the amount you **keep safe**, but there are two possible outcomes for the amount you **bet**:

- GOOD outcome: you earn **three times** the amount you bet
- BAD outcome: you **lose** the amount you bet

See examples on the next page.

Next

Subject#: AZCDGH

Example 1

You **bet** $0.10 and **keep safe** the remaining $0.90:

- In the GOOD outcome you end up with $0.10 * 3 + $0.90 = $1.20
- In the BAD outcome you end up with $0.90

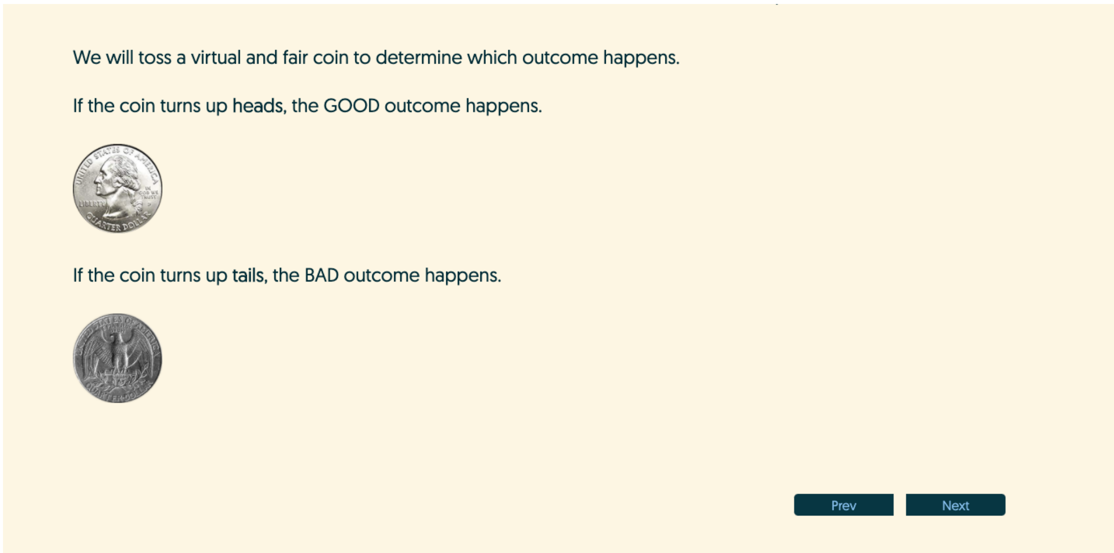
Example 2

You **bet** $0.90 and **keep safe** the remaining $0.10:

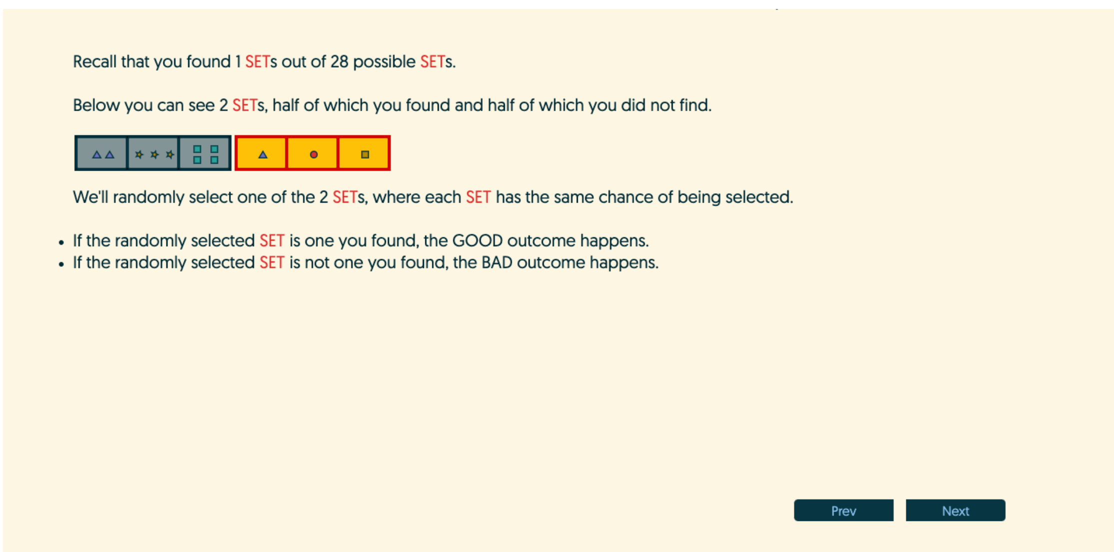
- In the GOOD outcome you end up with $0.90 * 3 + $0.10 = $2.80
- In the BAD outcome you end up with $0.10

The next page explains the chance component of the lottery.

Prev Next



(D-1) Treatment with *neutral* risk task



(D-2) Treatment with *context* risk task

Subject#: AZCDGH

The actual randomization will be done using public and verifiable means. Specifically, it will use future results from the Pennsylvania Lottery.

For more details, please e-mail Felipe A. Araujo at f.araujo@pitt.edu or check this study registration at the American Economic Association website.

Prev Next

Subject#: AZCDGH

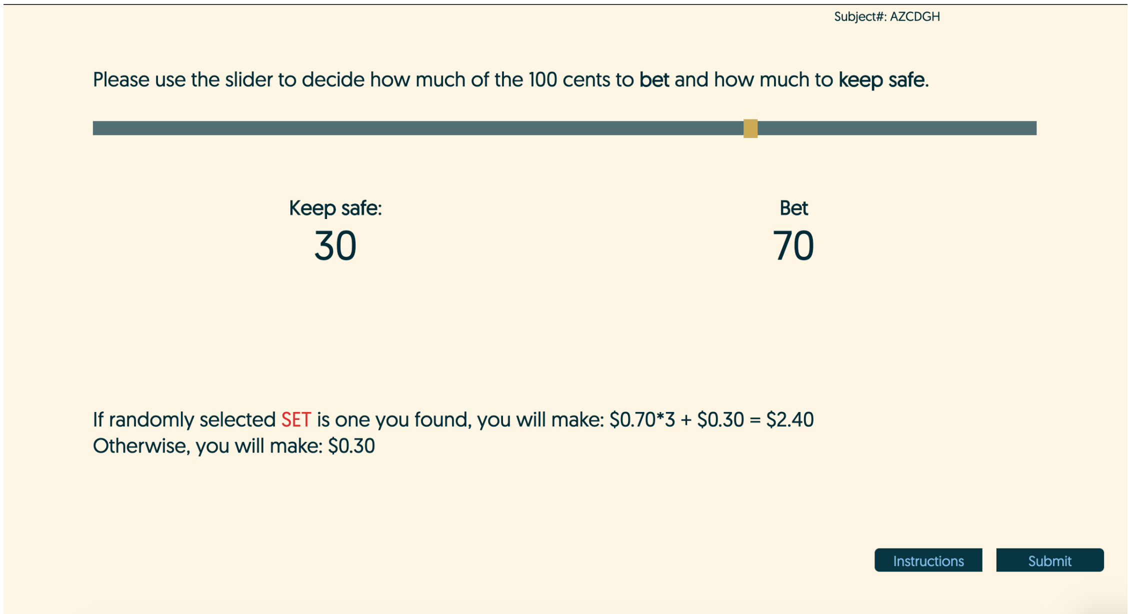
Please use the slider to decide how much of the 100 cents to **bet** and how much to **keep safe**.

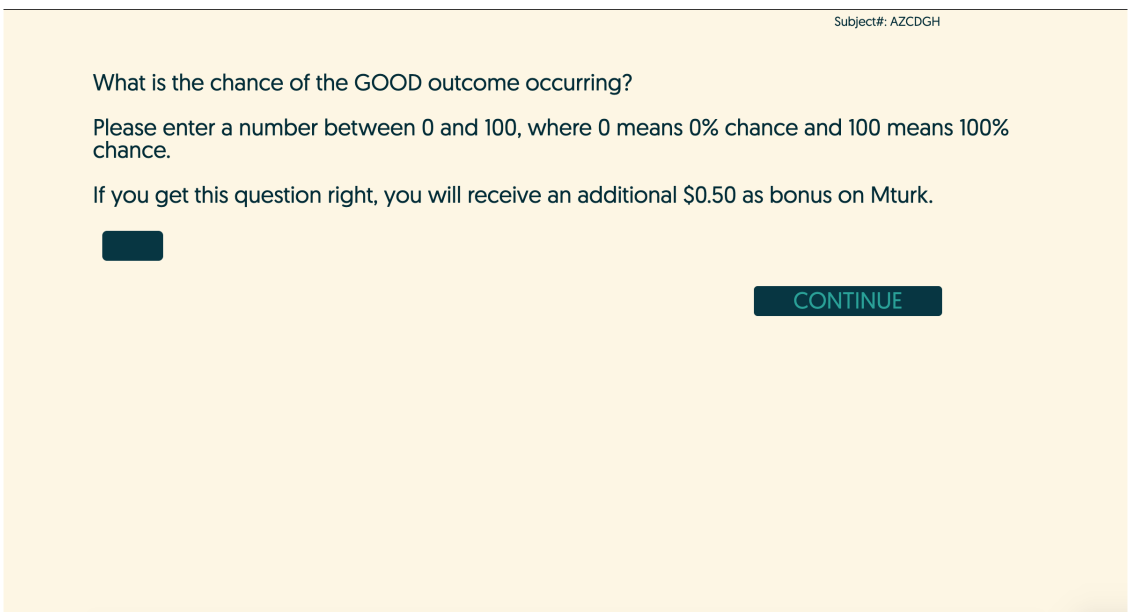
Keep safe:
??

Bet
??

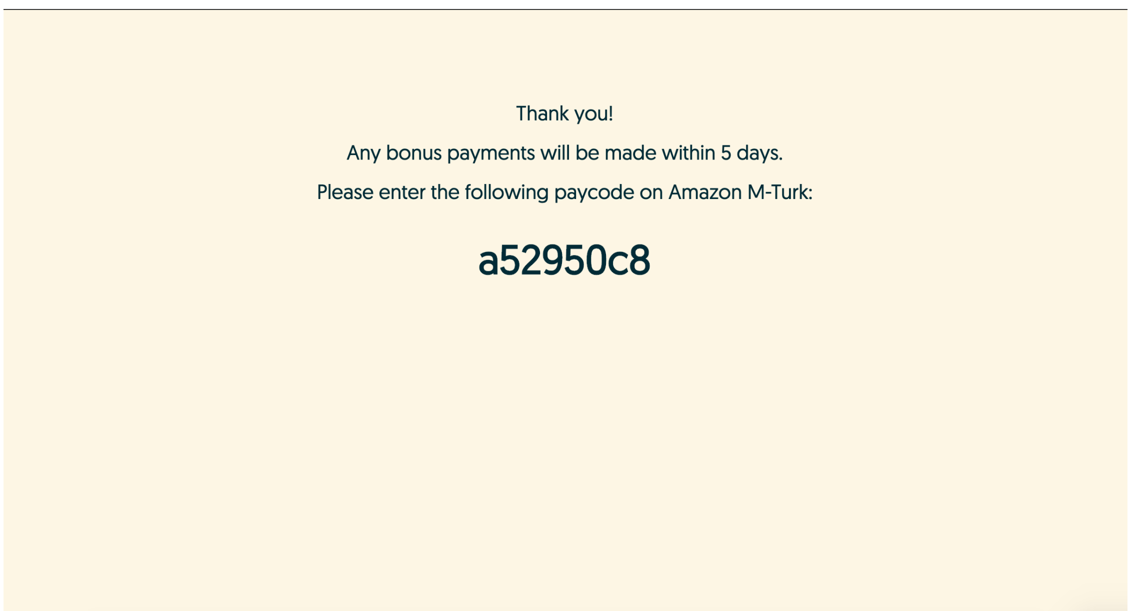
If randomly selected SET is one you found, you will make: ??
Otherwise, you will make: ??

Instructions Submit


Subject#: AZCDGH
Please use the slider to decide how much of the 100 cents to bet and how much to keep safe.
Keep safe:
30
Bet
70
If randomly selected SET is one you found, you will make: $0.70*3 + $0.30 = $2.40
Otherwise, you will make: $0.30
Instructions
Submit


Subject#: AZCDGH
What is the chance of the GOOD outcome occurring?
Please enter a number between 0 and 100, where 0 means 0% chance and 100 means 100% chance.
If you get this question right, you will receive an additional $0.50 as bonus on Mturk.
CONTINUE


Subject#: AZCDGH
Please answer a few questions about yourself
Age:
18-30 31-45 46-65 65+
Gender:
Female Male Other
Highest degree completed so far:
High School Some College Bachelors Masters Ph.D.
CONTINUE


Thank you!
Any bonus payments will be made within 5 days.
Please enter the following paycode on Amazon M-Turk:
a52950c8

Bibliography

Atila Abdulkadiroğlu, Joshua Angrist, and Parag Pathak. The elite illusion: Achievement effects at boston and new york exam schools. *Econometrica*, 82(1):137–196, 2014.

James Albrecht, Monica Robayo-Abril, and Susan Vroman. Public-sector employment in an equilibrium search and matching model. *The Economic Journal*, 129(617):35–61, 2018.

Alberto Alesina, Reza Baqir, and William Easterly. Redistributive public employment. *Journal of Urban Economics*, 48(2):219–241, 2000.

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

Peter Arcidiacono, V Joseph Hotz, and Songman Kang. Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, 166 (1):3–16, 2012.

Olivier Bargain and Blaise Melly. Public sector pay gap in france: New evidence using panel data. *IZA Disucssion Paper*, (3427):1–17, 2008.

Robert J Barro. The control of politicians: an economic model. *Public choice*, 14(1):19–42, 1973.

Charles Bellemare, Marion Goussé, Guy Lacroix, and Steeve Marchand. Physical disability and labor market discrimination: Evidence from a field experiment. Université Laval Working Paper, April 2018.

Keith A Bender. The central government-private sector wage differential. *Journal of Economic Surveys*, 12(2):177–220, 1998.

Timothy Besley and Maitreesh Ghatak. Competition and incentives with motivated agents. *American economic review*, 95(3):616–636, 2005.

Philipp Bewerunge, Harvey S Rosen, et al. Wages, pensions, and public-private sector compensation differentials. *Griswold Center for Economic Policy Studies Working Paper*, 227, 2012.

Holger Bonin, Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, 14(6):926–937, 2007.

Jake Bradley, Fabien Postel-Vinay, and Hélène Turon. Public sector wage policy and labor market equilibrium: a structural model. *Journal of the European Economic Association*, 15(6):1214–1257, 2017.

Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

Maria M Campos, Domenico Depalo, Evangelia Papapetrou, Javier J Pérez, and Roberto Ramos. Understanding the public sector pay gap. *IZA Journal of Labor Policy*, 6(1):7, 2017.

Juan D Carrillo and Thomas R Palfrey. The compromise game: two-sided adverse selection in the laboratory. *American Economic Journal: Microeconomics*, 1(1):151–81, 2009.

Briana Chang. Adverse selection and liquidity distortion. 2014.

Gary Charness and Dan Levin. The origin of the winner's curse: A laboratory study. *American Economic Journal: Microeconomics*, 1(1):207–36, 2009.

Ernesto Dal Bó, Frederico Finan, and Martín A Rossi. Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, 128(3):1169–1218, 2013.

Brendan Daley and Brett Green. Waiting for news in the market for lemons. *Econometrica*, 80(4):1433–1504, 2012.

Clément de Chaisemartin and Luc Behaghel. Estimating the effect of treatments allocated by randomized waiting lists. Technical report, National Bureau of Economic Research, 2019.

Jonathan de Quidt, Lise Vesterlund, and Alistair J Wilson. Experimenter demand effects. In *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing, 2019.

Eddie Dekel, Barton L Lipman, and Aldo Rustichini. Standard state-space models preclude unawareness. *Econometrica*, 66(1):159–173, 1998.

Ronald G Ehrenberg and Joshua L Schwarz. Public-sector labor markets. *Handbook of labor economics*, 2:1219–1260, 1986.

Benjamin Enke. What you see is all there is. July 2017.

Ignacio Esponda. Behavioral equilibrium in economies with adverse selection. *American Economic Review*, 98(4):1269–1291, 2008.

Ignacio Esponda and Demian Pouzo. Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130, 2016.

Ignacio Esponda and Emanuel Vespa. Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics*, 6(4):180–202, 2014.

Ignacio Esponda and Emanuel Vespa. Endogenous sample selection and partial naiveté in common value environments: A laboratory study. 2015.

Erik Eyster and Matthew Rabin. Cursed equilibrium. *Econometrica*, 73(5):1623–1672, 2005.

Ronald Fagin and Joseph Y Halpern. Belief, awareness, and limited reasoning. *Artificial intelligence*, 34(1):39–76, 1987.

Frederico Finan, Benjamin A. Olken, and Rohini Pande. *Handbook of Field Experiments*, chapter The Personnel Economics of the Developing State, pages 467–511. North-Holland, 2017.

Miguel N Foguel, Indermit Gill, Rosane Mendonça, and Ricardo Paes de Barros. The public-private wage gap in brazil. *Revista brasileira de economia*, 54(4):433–472, 2000.

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

William Fuchs and Andrzej Skrzypacz. Government interventions in a dynamic market with adverse selection. *Journal of Economic Theory*, 158:371–406, 2015.

Drew Fudenberg and Emanuel Vespa. Heterogeneous play and information spillovers in the lab. UCSB Working Paper, March 2018.

Alex Gershkov and Motty Perry. Dynamic contracts with moral hazard and adverse selection. *Review of Economic Studies*, 79:268–306, 2012.

Vladimir Gimpelson, Anna Lukiyanova, and Anna Sharunina. Estimating the public-private wage gap in russia: What does quantile regression tell us? *Higher School of Economics Research Paper No. WP BRP*, 104, 2015.

TH Gindling, Zahid Hasnain, David Locke Newhouse, and Rong Shi. Are public sector workers in developing countries overpaid? evidence from a new global data set. *Evidence from a New Global Data Set (February 25, 2019). World Bank Policy Research Working Paper*, (8754), 2019.

Uri Gneezy and Jan Potters. An experiment on risk taking and evaluation periods. *The quarterly journal of economics*, 112(2):631–645, 1997.

Gustavo Gonzaga and Sergio Firpo. Going private: Public sector rents and privatization in brazil. In *32º Meeting of the Brazilian Econometric Society*, 2010.

Robert G Gregory and Jeff Borland. Recent developments in public sector labor markets. *Handbook of labor economics*, 3:3573–3630, 1999.

Veronica Guerrieri and Robert Shimer. Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality. *American Economic Review*, 104(7):1875–1908, 2014.

Luigi Guiso and Monica Paiella. Risk Aversion, Wealth, and Background Risk. *Journal of the European Economic Association*, 6(6):1109–1150, 12 2008.

Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.

Joseph Y Halpern and Evan Piermont. Dynamic unawareness. 2020.

Joseph Y Halpern and Leandro C Rêgo. Reasoning about knowledge of unawareness revisited. *Mathematical Social Sciences*, 65(2):73–84, 2013.

Rema Hanna, Sendhil Mullainathan, and Joshua Schwartzstein. Learning through noticing: Theory and evidence from a field experiment. *Quarterly Journal of Economics*, 129(3): 1311–1353, 2014.

Igal Hendel, Alessandro Lizzeri, and Marciano Siniscalchi. Efficient sorting in a dynamic adverse-selection model. *Review of Economic Studies*, 72(2):467–497, 2005.

Guido Imbens and Karthik Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3):933–959, 2012.

Guido W Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635, 2008.

Asen Ivanov, Dan Levin, and Muriel Niederle. Can relaxation of beliefs rationalize the winner's curse?: An experimental study. *Econometrica*, 78(4):1435–1452, 2010.

Philippe Jehiel. Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123 (2):81–104, 2005.

Philippe Jehiel and Frédéric Koessler. Revisiting games of incomplete information with analogy-based expectations. *Games & Economic Behavior*, 62(2):533–557, 2008.

Philippe Jehiel and Dov Samet. Valuation equilibrium. *Theoretical Economics*, 2(2):163–185, 2007.

Ginger Zhe Jin, Michael Luca, and Daniel Martin. Is no news (perceived as) bad news? an experimental investigation of information disclosure. Working Paper 21099, National Bureau of Economic Research, April 2015.

John H Kagel. Cross-game learning: Experimental evidence from first-price and english common value auctions. *Economics Letters*, 49(2):163–170, 1995.

John H Kagel and Dan Levin. *Common Value Auctions and the Winner's Curse*. Princeton University Press, Princeton, NJ, 2002.

Edi Karni and Marie-Louise Vierø. Awareness of unawareness: a theory of decision making in the face of ignorance. *Journal of Economic Theory*, 168:301–328, 2017.

Michal Kolesár and Christoph Rothe. Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8):2277–2304, 2018.

Alan B Krueger. Are public sector workers paid more than their alternative wage? evidence from longitudinal data and job queues. In *When public sector workers unionize*, pages 217–242. University of Chicago Press, 1988.

Jelena Lausev. *Journal of Economic Surveys*, 28(3):516–550, 2014.

Dan Levin, John H Kagel, and Jean-Francois Richard. Revenue effects and information processing in english common value auctions. *American Economic Review*, pages 442–460, 1996.

Wenjun Ma and Burkhard C Schipper. Does exposure to unawareness affect risk preferences? a preliminary result. *Theory and Decision*, 83(2):245–257, 2017.

Gangadharrao S Maddala. Disequilibrium, self-selection, and switching models. *Handbook of econometrics*, 3:1633–1688, 1986.

Charles F Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.

Alejandro Martínez-Marquina, Muriel Niederle, and Emanuel Vespa. Probabilistic states versus multiple certainties: The obstacle of uncertainty in contingent reasoning. Working Paper 24030, National Bureau of Economic Research, November 2017. URL `http://www.nber.org/papers/w24030`.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714, 2008.

Blaise Melly. Public-private sector wage differentials in germany: Evidence from quantile regression. *Empirical Economics*, 30(2):505–520, 2005.

Friederike Mengel, Elias Tsakas, and Alexander Vostroknutov. Past experience of uncertainty affects risk aversion. *Experimental Economics*, 19(1):151–176, 2016.

Alejandra Mizala, Pilar Romaguera, and Sebastián Gallegos. Public–private wage gap in latin america (1992–2007): A matching approach. *Labour Economics*, 18:S115–S131, 2011.

Salvatore Modica and Aldo Rustichini. Awareness and partitional information structures. *Theory and decision*, 37(1):107–124, 1994.

John S Moran, Harlan R Janes, Thomas A Peterman, and Katherine M Stone. Increase in condom sales following aids education and publicity, united states. *American Journal of Public Health*, 80(5):607–608, 1990.

Brent R Moulton. A reexamination of the federal-private wage differential in the united states. *Journal of Labor Economics*, 8(2):270–293, 1990.

Richard E Mueller. Public–private sector wage differentials in canada: evidence from quantile regressions. *Economics Letters*, 60(2):229–235, 1998.

Kathleen Ngangoué and Georg Weizsäcker. Learning from unrealized versus realized prices. *Working Paper*, 2017.

Ben Ost, Weixiang Pan, and Douglas Webber. The returns to college persistence for marginal students: Regression discontinuity evidence from university dismissal policies. *Journal of Labor Economics*, 36(3):779–805, 2018.

Evan Piermont. Algebraic semantics for propositional awareness logics. *arXiv preprint arXiv:1910.07275*, 2019a.

Evan Piermont. Unforeseen evidence. *arXiv preprint arXiv:1907.07019*, 2019b.

Ernesto Reuben, Matthew Wiswall, and Basit Zafar. Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186, 2017.

Michael Rosholm and Nina Smith. The danish gender wage gap in the 1980s: A panel data study. *Oxford Economic Papers*, 48(2):254–279, 1996.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Donald B Rubin. Assignment to treatment group on the basis of a covariate. *Journal of educational Statistics*, 2(1):1–26, 1977.

Max Schanzenbach. Explaining the public-sector pay gap: The role of skill and college major. *Journal of Human Capital*, 9(1):1–44, 2015.

Sharon P Smith. Pay differentials between federal government and private sector workers. *ILR Review*, 29(2):179–197, 1976.

Ran Spiegler. Bayesian networks and boundedly rational expectations. *Quarterly Journal of Economics*, 131(3):1243–1290, 2016.

Jacques Van der Gaag and Wim Vijverberg. A switching regression model for wage determinants in the public and private sectors of a developing country. *The review of economics and statistics*, pages 244–252, 1988.

Hans Van Ophem. A modified switching regression model for earnings differentials between the public and private sectors in the netherlands. *The Review of Economics and Statistics*, pages 215–224, 1993.

Liviu Voinea and Flaviu Mihaescu. A contribution to the public–private wage inequality debate: The iconic case of romania 1. *Economics of Transition*, 20(2):315–337, 2012.

Stephanie W Wang, Michelle Filiba, and Colin F Camerer. Dynamically optimized sequential experimentation (dose) for estimating economic preference parameters. September 2010.