


5-2022

## Economic Experiments on Group Identity and Bias

Nathaniel Christopher Burke  
*University of Arkansas, Fayetteville*

Follow this and additional works at: <https://scholarworks.uark.edu/etd>

 Part of the [Behavioral Economics Commons](#), [Economic Policy Commons](#), [Economic Theory Commons](#), [Growth and Development Commons](#), [Political Economy Commons](#), and the [Public Economics Commons](#)

---

### Citation

Burke, N. C. (2022). Economic Experiments on Group Identity and Bias. *Graduate Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/4500>

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact [uarepos@uark.edu](mailto:uarepos@uark.edu).

**Economic Experiments on Group Identity and Bias**

**A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy in Economics**

**by**

**Nathaniel Christopher Burke  
Manhattan College  
Bachelor of Arts in Economics, 2011  
University of Alaska Fairbanks  
Master of Science in Resource and Applied Economics, 2017**

**May 2022  
University of Arkansas**

**This dissertation is approved for recommendation to the Graduate Council.**

---

**Sherry Li, Ph.D.  
Dissertation Co-Chair**

---

**Andy Brownback, Ph.D.  
Dissertation Co-Chair**

---

**Arya Gaduh, Ph.D.  
Committee Member**

## **Abstract**

Experiments in economics have been a valuable tool to understand the behavioral implications of incentives on the decision-making process. Particularly, aspects of decision making that cannot be observed in empirical data can be better isolated in an experimental setting such as bias and identity impacts. This dissertation uses three distinct experiments to further the understanding of individual biases, perceptions, and identity and how they impact the way people defer to these internal traits under incentives. This dissertation looks at how well individuals can make inferences about polling data that was collected from individuals susceptible to socially desirable responding. It also explores the identity importance of gender in both a public goods game setting as well as a setting where individuals must make predictions about the risk and time preferences of others.

## **Dedication**

This work is dedicated to my mother, Iris Osorio-Sevilla, who gave everything for my education and to all the kids that look like me, coming from a community too often forgotten, marginalized, and undervalued.

“Bringing the gifts that my ancestors gave, I am the dream and the hope of the slave.  
I rise. I rise. I rise.” Maya Angelou

## **Acknowledgements**

For all of their guidance and support through this process, I want to acknowledge my advisors, Sherry Li and Andy Brownback, who have made me into the researcher I am today.

I also want to acknowledge the guidance from my previous advisors from my Master's program, Joe Little, and my undergraduate advisor, Gwendolyn Tedeschi, who were both instrumental in making me fall in love with economic research and using it as a means to bring positive change; Dr. Sarah Jacobson for welcoming me into the profession; my cohort, Logan Miller, Ahmad Shah Mobariz, and James Willbanks, for their constant feedback, encouragement, and never-ending support; Drs. Barbara Lofton and Synetra Hughes from the Office of Diversity and Inclusion; 1SG Durvell Smith, USA Ret., 1SG Paul Peters, USA (Ret., and MSG Edson Rodriguez, USA for teaching me what it means to be a leader and the essence of military bearing.

I also want to acknowledge the never ending support, love, and motivation from my wife and son, who were patient through the missed bike rides and family dinners as they supported me during the dissertation process.

Finally, I want to acknowledge all of the ADD/ADHD kids, the bilingual kids, the Latino kids, and the Black kids who keep showing up everyday, breaking barriers, and keeping me inspired to do the work I'm doing. Thank you for teaching me what we do this for.

## Contents

<b>1 Introduction</b>	<b>1</b>
<b>2 Chapter 1: Inferences from Biased Polls</b>	<b>3</b>
Introduction . . . . .	4
Experimental Design . . . . .	12
Primary Hypotheses . . . . .	18
Data Description and Manipulation Checks . . . . .	24
Main Results . . . . .	29
Discussion and Conclusion . . . . .	39
References . . . . .	42
Appendices . . . . .	47
<b>3 Chapter 2: Group Identity Switching</b>	<b>75</b>
Introduction . . . . .	76
Experimental Design . . . . .	80
Results . . . . .	87
Discussion and Conclusion . . . . .	95
References . . . . .	98
Appendices . . . . .	101
<b>4 Chapter 3: Identity Based Risk and Time Preference Predictions</b>	<b>109</b>
Introduction . . . . .	110
Experimental Design . . . . .	114
Results . . . . .	117
Conclusions . . . . .	131
References . . . . .	135
Appendices . . . . .	137



## Introduction

Economics has made major advances in the last 20 years in the understandings of how identity impacts decision making, informs biases, and motivates certain types of behaviors. Additionally, as a society, we have become more aware of the social and natural identity groups that we identify with in order to find places of belonging and reaffirm our beliefs and perspectives about the world. The three papers in this dissertation come together to look at how behavioral preferences, identity, and biases impact the decision making processes of individuals under incentivized conditions. More specifically, I look into the real impacts of behavioral perceptions on earnings in a laboratory experimental setting to gain a better economic understanding of how people perceive and interact with each other.

This line of research was inspired by trying to understand the story of why people make decisions that align with group ideals over individual goals. Two of the most easily relatable examples of this are how people make sacrifices in their educational or career pursuits in the interest of staying in a particular community or staying near family. This basic concept led to the consideration of the second paper in this work, where I use an experimental design that puts individuals in a situation where they have to choose between their own personal financial gain and the overall financial gain of the group. This topic may seem to be quite different from the first and third works that look at the ability of individuals to predict preferences and biases but the behavioral mechanism is very similar. It comes down to the idea that people have a constant incentive to belong to a certain group and that pursuit for a sense of belonging also creates environments where individuals create stereotypes and have biases about those that not only do not belong to their group but also are a part of their own group.

This first paper in this work looks at how individuals lie to themselves and to researchers/surveyors in order to make them seem more socially desirable, referred to as socially desirable responding, and then how well other people can detect this socially de-



sirable responding and properly weight those responses to predict the truth in the adjusted statements. Once again, the motivating factor is this sense of belonging. Though that experiment does not incite a particular identity dimension, it does use the desire to be seen as generally sociable in general. The third paper looks at how well people can make predictions about the time and risk preferences of people that belong to their group versus the general public to see if there is any advantage to being in the same group but not actually having any other information when it comes to knowing each other.

## Chapter 1: Inference from Biased Polls<sup>1</sup>

Andy Brownback<sup>2</sup> Nathaniel Burke<sup>3</sup> Tristan Gagnon-Bartsch<sup>4</sup>

### Abstract

Poll respondents often attempt to present a positive image by overstating virtuous behaviors. We examine whether people account for this “socially desirable responding” (SDR) when drawing inferences from poll data. In an experiment, we incentivize “predictors” to guess others’ choice behaviors across eight actions with varying social desirability. Predictors observe random subsamples of either (i) incentivized choices or (ii) hypothetical claims from polls. The hypothetical claims exhibit predictable SDR and predictors are reasonably skeptical of them. However, their skepticism is not tailored to the direction or magnitude of SDR. This under-correction occurs even though subjects’ explicit responses can predict SDR.

**JEL classification:** D91, D84, D72.

**Keywords:** Polling, Social Desirability, Inference, Signaling, Selection Bias.

---

<sup>1</sup>For helpful comments, we thank Chiara Aina, Benjamin Bushong, Jonathan de Quidt, Uri Gneezy, David Huffman, Alex Imas, Michael Kuhn, Sherry Li, Peter McGee, Matthew Rabin, Joshua Schwartzstein, Marta Serra-Garcia, and seminar audiences at the SEA Annual Meeting, AEA Mentoring Pipeline Conference, and West Virginia University. AEA RCT registry number: AEARCTR-0005186 (available at <https://doi.org/10.1257/rct.5186-1.0>). We gratefully acknowledge support from the AEA Mentoring Program (NSF Award #1730651).

<sup>2</sup>University of Arkansas: [ABrownback@walton.uark.edu](mailto:ABrownback@walton.uark.edu)

<sup>3</sup>West Virginia University: [nathaniel.burke@mail.wvu.edu](mailto:nathaniel.burke@mail.wvu.edu)

<sup>4</sup>Harvard University: [gagnonbartsch@fas.harvard.edu](mailto:gagnonbartsch@fas.harvard.edu)

## 1 Introduction

Presenting a positive image is a widespread human desire, and many are willing to incur significant costs to do so (Bagwell and Bernheim, 1996; Bursztyn, Ferman, Fiorin, Kanz and Rao, 2018; Veblen, 1899). Thus, we expect people to take advantage of opportunities to costlessly inflate their own image. Indeed, people often do so by misrepresenting their views, traits, or behaviors in response to unincentivized elicitations such as opinion surveys, self-reports, or political polls. This is known as *socially desirable responding (SDR)* (Edwards, 1957; Maccoby and Maccoby, 1954; Paulhus, 1984).

Unincentivized elicitations are often the best available source of information even when they are plagued by SDR. For instance, doctors rely on self-reports to design treatments around alcohol use even though these reports exhibit well-documented biases (Del Boca and Noll, 2000; Latkin, Edwards, Davey-Rothwell and Tobin, 2017). And businesses use political polls to predict and prepare for changes in government policies despite the potential bias in such polls (Finkel, Guterbock and Borg, 1991). While unincentivized elicitations come in many forms, we refer to them simply as *polls* and the elicited responses as *poll data*.

We experimentally study whether people account for SDR in poll data when using it to draw inference about subsequent choice behavior. SDR suggests a general tendency to misrepresent preferences, but many people still respond to polls truthfully.<sup>5</sup> Thus, even though poll data may be biased by SDR, a careful observer of this data may be able to extract an unbiased signal from it. That is, even biased poll data can still bear useful information. However, extracting accurate signals from this data requires an appreciation that responses cannot always be taken at face value and an understanding of how they might be distorted. One must anticipate SDR and discount claims of virtuous behavior while also recognizing that people are unlikely to be lying when they admit to stigmatized behaviors.

---

<sup>5</sup>People may respond truthfully both because of a preference for being honest and a preference to appear honest, as documented by a large experimental literature on an aversion to lying; see, e.g., Abeler, Nosenzo and Raymond (2019) for a meta-study of 90 studies using designs similar to Fischbacher and Föllmi-Heusi (2013).

We refer to this ability to interpret poll data in a way that corrects for misreporting due to SDR as *social sophistication*.

How to draw accurate inference from biased signals is a longstanding question in economics. For instance, Crawford and Sobel (1982) explore this concept through “cheap-talk” equilibria in which receivers extract information from signals sent by senders with misaligned incentives. Similarly, Kartik (2009) demonstrates the informativeness of communication in such settings when senders bear some cost to misreporting their private information. Experimental studies confirm the benefits of communication even when incentives are not aligned (see Farrell and Rabin, 1996 and Crawford, 1998 for reviews). The real-world value of using social sophistication to identify and correct for misreported information is clear. For example, doctors can provide better care when they anticipate that patients are reluctant to reveal mental health issues (Bharadwaj, Pai and Suziedelyte, 2017) and accordingly discount claims of perfect mental health. And job-seekers can better target their search efforts when they appreciate that other workers often relay overly-optimistic information about job prospects (Arnold, Feldman and Purbhoo, 1985).

In our study, we elicit both poll responses and actual choice behaviors, allowing us to clearly measure the SDR in our poll data. We then examine the degree of social sophistication present when people are given this poll data and asked to predict others’ actual choice behaviors. Since we directly observe the effect of SDR on the poll data, we can similarly observe how people correct for it in their predictions. We find ample evidence of social sophistication along fundamental dimensions: people do anticipate the potential for biased poll data and discount the hypothetical claims of others. However, we find no evidence of more complex dimensions of sophistication: people do not properly tailor their discounting to the direction or magnitude of SDR.

Moreover, our experiment develops a novel methodology of information-provision to identify how beliefs respond to poll data.<sup>6</sup> We reveal random subsamples from an assigned infor-

---

<sup>6</sup>A large literature demonstrates that information provision influences beliefs and attitudes across numerous policy-relevant domains; see Haaland, Roth and Wohlfart (2020) for a review.

mation source (either poll data or actual choice behaviors), inducing mechanically random sampling variation in signals. Controlling for differences in the distribution of signals isolates signal variation from sampling noise—revealing experimentally-random changes in signals and their *causal* impact on inferences. Our random assignment of information sources also provides causal evidence on the heterogeneous interpretation of information from different sources.

Our first step in studying how people account for SDR in poll data is to construct a setting where we can observe how SDR affects responses to eight separate actions. The initial stage of our experiment measures actual choices and hypothetical claims using parallel elicitations with two distinct groups of subjects. Participants in each group answer whether they would take each of the actions, which vary in their social desirability (e.g., donating to St. Jude Children’s Hospital or stealing from an experimental subject in another session).<sup>7</sup> In the *incentive-compatible (IC) group*, we use an incentivized revealed-preference elicitation to measure actual choices. In the *hypothetical (H) group*, we use an unincentivized stated-preference elicitation to measure hypothetical claims about behavior for the same eight actions. SDR prompts the H group to overstate (understate) their demand relative to the IC group for actions they believe to be virtuous (stigmatized).

Our design begins with no ex-ante perspective on which actions are virtuous or stigmatized. Instead, we recruit a separate *sentiment group* to rate the social desirability of each action. We use this independent evaluation to establish that SDR is identifiable and predictable in our controlled setting. A one standard-deviation (SD) increase in how the sentiment group scores an action’s social desirability is associated with a 3.1 percentage-point increase in the H group’s overstatement of demand for that action ( $p < 0.001$ ). For example, our sentiment group evaluated donating to St. Jude Children’s Hospital as the most

---

<sup>7</sup>In total, we consider eight different actions. Six involve deciding whether to donate \$1 to an organization: St. Jude Children’s Hospital, a local NPR affiliate, the Democratic National Committee, the Republican National Committee, Joe Biden’s campaign, and Donald Trump’s campaign. We also consider stealing \$1 from another participant in the study and taking \$1 for yourself from a planned donation to the Make-A-Wish Foundation.

virtuous action (2.24 SD above the mean). This is associated with a 13 percentage-point overstatement of claimed desire to donate: 75% *claim* they would donate, but only 62% do.

We then evaluate the degree to which people anticipate and correct for the SDR manifest in the H group’s claims. To do so, we incentivize *predictors* to guess the aggregate choice behavior of the IC group for each action.<sup>8</sup> Predictors make initial guesses about choice behavior. They are then randomly assigned to observe “signals,” which are subsamples of either (i) choices from the IC group itself or (ii) claims from the H group. Predictors then make updated guesses about the behavior of the IC group. By observing predictors’ updating behavior, we can deduce the differential weighting of information from the two sources and, thus, their social sophistication. Specifically, we assess whether predictors account for SDR by appropriately discounting the claims of the H group.

We follow our pre-registration and examine several hypotheses regarding the differential treatment of signals from the two information sources.

Our first main hypothesis examines whether predictors *anticipate* SDR and accordingly down-weight the (potentially biased) claims from the H group. We find that they do. 31% of predictors’ updating from IC-group signals is “extra updating” attributable to the added weight given to the IC-group’s choices relative to the H-group’s claims ( $p < 0.01$ ).<sup>9</sup>

We are motivated by polling contexts in which people regularly participate in both the choice and prediction groups—e.g. the average voter likely has experience with participating in political polls and has made predictions based on data from such polls. For this reason, we examine how prior experiences may influence social sophistication. To do so, we recruited a mix of predictors: some were newly recruited while others had previously participated in either the IC or H group. We directly compare responses across the two groups and find suggestive but inconclusive evidence that predictors who previously participated in the H group discount the claims of the H group more than predictors without this experience

---

<sup>8</sup>Predictors are a mix of newly recruited participants and returners from the IC and H groups. As we detail later, this allows us to examine a key question about how experience with SDR affects predictor behavior.

<sup>9</sup>All estimates presented in the introduction are derived from our within-subjects specification. See Section 5 for details on our analysis.

( $p = 0.125$ ). This is consistent with the idea that these subjects are more skeptical of hypothetical claims because they are familiar with the impulse to lie when making such claims.

While discounting the average signal from the H group is a fundamental part of sophisticated inference, discounting all signals equally would not reflect full social sophistication. Full sophistication calls for people to adjust their discounting depending on the direction and magnitude of the bias. Predictors' guesses offer little evidence of these more challenging dimensions of social sophistication.

Our second main hypothesis explores more complex social sophistication by asking whether predictors recognize the *direction* of SDR; that is, whether it is socially desirable to overstate or understate demand for an action. Social sophistication rests on such knowledge as it enables a predictor to determine whether the signal they receive reflects “perception-inflating” boasting—which should be discounted—or reflects “perception-deflating” confessing—which should be given additional weight. We define a signal to be perception inflating if it implies greater social desirability *than the predictor initially guessed*. A signal is perception deflating if it implies lesser social desirability *than the predictor initially guessed*.<sup>10</sup> A perception-deflating signal is particularly informative because it suggests that more respondents than expected admit to socially undesirable behavior despite the opportunity to freely claim virtuous behavior. We find that predictors fail to recognize this. While they correctly discount perception-inflating signals from the H group by 18% relative to the IC group ( $p < 0.001$ ), they treat perception-deflating signals from the H group almost identically to those from the IC group.

Our third main hypothesis asks if predictors recognize the *relative magnitude* of SDR across the eight actions. To answer this question, we examine whether predictors discount

---

<sup>10</sup>Note that we define a perception-inflating (-deflating) signal entirely with respect to how it would change a predictor's perception of the IC group's behavior assuming that the signal reflected truthful responses. Thus, even a perception-deflating signal from the H group is consistent with SDR. SDR suggests that more people hypothetically claim socially desirable behavior than actually choose it, while a perception-deflating signal shows fewer people claiming socially desirable behavior relative to *the predictor's initial guess* of how many would actually choose it.

signals for each action based on the degree of SDR for that specific action. When considering actions that are notably biased, predictors should treat claims from the H group with increased skepticism. However, we find no evidence of this dimension of social sophistication. If anything, our point estimates suggest that predictors’ guesses place *more* weight on claims from the H group as SDR becomes more extreme ( $p > 0.10$ ).

The lack of social sophistication demonstrated by predictors’ guesses stands in striking contrast with the responses of the sentiment group. When explicitly asked to evaluate the social desirability of each action, the sentiment group produced reasonably accurate assessments. Hence, the knowledge of which actions tend to incite greater social-image concerns does exist in our population. However, it appears that predictors neglect this knowledge when deciding how much to discount claims from the H group.<sup>11</sup>

We also find irregularities in the confidence predictors place in their guesses. After each guess we elicited from predictors, we elicited their confidence in that guess. For initial guesses, we find a negative correlation between the accuracy of a predictor’s guess and their confidence ( $p < 0.01$ ). This “Dunning-Kruger” effect (Kruger and Dunning, 1999) also emerges between updated guesses and confidence among predictors who receive information from the H group ( $p < 0.05$ ). However, it is diminished and no longer significant for predictors receiving the higher-quality information from the IC group. These results highlight the downstream consequences of biased polling information: not only do predictors under-correct for biases in the poll data, but this low-quality data may also allow false confidence to persist.

SDR is typically interpreted as a form of social-signaling in which people project a positive image of themselves to others. Interestingly, this behavior persists in many online and anonymous contexts such as ours. This could be signaling to the experimenter or pollster, but it is not distinguishable from the form of self-signaling outlined in Bénabou and Tirole (2002).

---

<sup>11</sup>A similar pattern emerges in experiments on “cursed thinking” (Eyster and Rabin, 2005) in trading environments with asymmetric information. Subjects often accept financial trades with better-informed parties to their own detriment (e.g., Samuelson and Bazerman, 1985). However, when explicitly asked, a typical subject in such settings correctly predicts that her better-informed partners will only agree to trades that are detrimental for her to accept (Hales, 2009).



In most contexts, SDR is likely driven by a combination of both social- and self-signaling and these two mechanisms are often observationally equivalent. While our anonymous, online context likely mutes the impact of SDR, it has a few advantages. First, it provides a test of our key hypotheses in a relevant context—real-world polls intentionally minimize confounds like SDR and experimenter demand effects. Second, despite the muted SDR, the evaluations of our sentiment group still predict the response biases. With this test satisfied by our sentiment group, it is natural to extend a similar test of social sophistication to predictors.

In a recent study of “political correctness,” Braghieri (2021) finds that SDR has a meaningful impact on the information content of public statements relative to those made in private. Our study offers a complement to this conceptually related paper. While both papers explore biases that arise in statements made about sensitive topics, they examine two different regimes that may limit such biases. Braghieri (2021) considers greater privacy—focusing on the “wedge” between private and public statements—while our paper considers greater incentives for truth-telling—focusing on the wedge between statements and consequential choices. Both papers take the analysis one step further and explore the degree to which others anticipate how responses will change as a result of the respective regime (privacy or incentives). Subjects in Braghieri (2021) are able to predict average differences between public and private statements, but similar to our findings, subjects exhibit limited sophistication when predicting heterogeneity in the bias. Other experimental differences may be informative about the mechanisms at play. Braghieri (2021) explicitly asks subjects to predict differences between private and public statements, while we use a difference-in-differences design to examine how predictors respond to information from the different sources. Explicitly eliciting beliefs about the wedge may prompt subjects to consider the possibility of misreported statements and hence may explain why subjects in Braghieri (2021) exhibit greater sophistication than subjects in our study.<sup>12</sup>

---

<sup>12</sup>Our analysis of sophistication is also related to Charness, Oprea and Yuksel (2021), as they similarly examine how subjects gather information from biased sources. However, instead of asking whether people can identify and correct biased data, as we do, they ask whether subjects can optimally select between data sources that have known biases. Their subjects fail to maximize the information content they can extract

Our exploration of social sophistication advances the literature on social norms in general and on SDR specifically. Krupka and Weber (2013) demonstrate that social norms are well-anticipated by experimental subjects. We extend the exploration of the predictability of norms by asking whether observers of poll data account for SDR when drawing inference.

Given the limited social sophistication that we find, the benefits of collecting and disseminating more accurate data are clear. Researchers have developed several tools, such as the randomized-response technique (Warner, 1965) and list experiments (Karlan and Zinman, 2012; Raghavaram and Federer, 1979), to identify underlying preferences when SDR is prevalent and incentivized elicitation is not possible. These tools have identified SDR in a broad set of stigmatized and virtuous domains.<sup>13</sup> Moreover, Rosenfeld, Imai and Shapiro (2016) find that these techniques can correct biased estimates and improve inference from polls. In light of our results, we believe there is strong evidence in favor of using these tools in regular audits to identify SDR and recalibrate polls. Independent sentiment surveys could also be used to predict susceptibility to SDR.

While we focus primarily on polling, biases from SDR affect incentivized economic experiments as well. “Experimenter demand”—where subjects respond in a manner they perceive to be consistent with the experimenter’s intention—is one expression of SDR. de Quidt, Haushofer and Roth (2018) find that the impact of this bias has clear bounds. Our results suggest that, though the impact of experimenter demand may be limited, observers of biased experimental data are unlikely to accurately predict the direction or degree of the bias.

Our paper begins with an explanation of our pre-registered experimental design in Section 2. Section 3 follows with our hypotheses and a simple model that develops the intuition

---

from the data because they tend to over-select sources biased towards giving confirmatory evidence.

<sup>13</sup>Tourangeau, Rips and Rasinski (2000) and Tourangeau and Yan (2007) provide reviews. SDR has been identified in political polls—often called a “Bradley Effect” or “Shy Tory Factor” (Brownback and Novotny, 2018; Hopkins, 2009; Reeves et al., 1997); polls for female and minority candidates (Brown-Iannuzzi, Najle and Gervais, 2019; Heerwig and McCabe, 2009; Kane, Craig and Wald, 2004; Stephens-Davidowitz, 2014; Streb, Burrell, Frederick and Genovese, 2008); sentiment surrounding race (Krysan, 1998), immigration (Janus, 2010), and same-sex marriage (Coffman, Coffman and Ericson, 2017; Powell, 2013); revelation of vote-buying behavior (Gonzalez-Ocantos, De Jonge, Meléndez, Osorio and Nickerson, 2012); voter turnout (Holbrook and Krosnick, 2010); and religious attendance (Jones and Elliot, 2016).

behind them. We then evaluate these hypotheses in Sections 4 and 5. Section 6 concludes.

## 2 Experimental Design

Our study consisted of three stages: the Sentiment Stage, the Choice Stage, and the Prediction Stage. Each stage took place online with subjects recruited from the University of Arkansas. Each stage featured the same eight actions framed as binary choices.

We recruited 39 subjects for the Sentiment Stage. For the Choice Stage, we recruited 187 subjects and split them into two groups. In the Prediction Stage, we recruited 95 new subjects to combine with returners from the Choice Stage.

### 2.1 Actions

Subjects considered eight binary choices to take an action or not. The eight actions were:

**St Jude Donation:** Donate \$1 to the St. Jude Children’s Hospital.

**NPR Donation:** Donate \$1 to KUAF radio station, the local NPR affiliate.

**Steal:** Steal \$1 from a participant in another stage of the study.

**Take Donation:** Take \$1 for yourself from a planned \$50 donation to the Make-A-Wish Foundation.

**Trump Donation:** Contribute \$1 to Donald Trump’s presidential campaign.

**Biden Donation:** Contribute \$1 to Joe Biden’s presidential campaign.

**RNC Donation:** Contribute \$1 to the Republican National Committee.

**DNC Donation:** Contribute \$1 to the Democratic National Committee.

We made no attempt to label each action as “virtuous” and “stigmatized” based on our a priori perceptions. We designed our experiment and all hypotheses to be agnostic about the sentiment surrounding actions; instead, we classify actions based solely on the evaluations of the sentiment group, who are drawn from the same population. In this way, all of our tests could be based on perceptions that are observably present in the population.

Though the emotional valence of a specific action was unimportant to our design, it was important that the actions we selected possessed a wide range of emotional valence so that we could test for sensitivity to *differences* in social desirability. It was also important that the actions did not exist solely at the extremes of virtue and stigma so that we could identify effects both *between* and *within* stigmatized and virtuous domains. Many actions were chosen in pairs so that the sentiment surrounding them would be likely to covary negatively. These steps were taken to increase the variance in choice behaviors and predictions so that we would not spuriously attribute general behaviors to systematic differences in behavior towards stigmatized and virtuous actions.<sup>14</sup>

The binary nature of decisions (either take an action or not) simplified the experiment and allowed us to send easily-understood signals of behavior to our predictors. All choices were made privately through online surveys. Subjects were assured that no individual responses would ever be viewed by anyone except the researchers. This is a conservative approach that likely mutes the impact of social desirability, since SDR is often dependent on the anticipated reactions of observers. As previously discussed, this provides a more natural test of social sophistication about SDR without experimenter demand effects. Actions were described identically and in detail to all subjects in all stages of the study, including information about the anonymity under which choices and statements were made. See Appendix Section C for the full description given to subjects.

## 2.2 Sentiment Stage

We recruited 39 subjects to evaluate the sentiment associated with each of the eight actions listed above. Subjects who participated in the Sentiment Stage did not participate in any other portion of the experiment; they were paid a flat fee of \$5.

For each of our eight actions, subjects answered the three questions below on a scale of 0-10, where 0 represented “Very Negative” and 10 represented “Very Positive” sentiment.

---

<sup>14</sup>Moreover, we selected several actions related to political views since this is a familiar domain in which people observe poll data.

1. How would you feel about taking this action yourself?
2. How would you feel about other people who take this action?
3. How do you think most other people would feel about people who take this action?

For each action  $A$ , let  $Q_{i,j,A}$  denote subject  $i$ 's response to question  $j \in \{1, 2, 3\}$  above. We then construct subject  $i$ 's "perceived virtue" of action  $A$ , denoted  $V_{i,A}$ , by taking the within-subject mean of these responses:  $V_{i,A} \equiv \frac{\sum_{j=1}^3 Q_{i,j,A}}{3}$ . Letting  $N_S$  denote the number of subjects in the Sentiment Stage, we will use the following indices to measure the perceived virtue of action  $A$ :

$$V_A \equiv \frac{\sum_{i=1}^{N_S} V_{i,A}}{N_S}, \quad (1)$$

$$\widehat{V}_{i,A} \equiv \frac{V_{i,A} - \bar{V}_i}{\sigma_i}, \quad (2)$$

where  $\bar{V}_i$  and  $\sigma_i$  are subject  $i$ 's mean and standard deviation of  $V_{i,A}$  across all eight actions.

Our pre-registered measure of social desirability,  $V_A$ , captures the perceived virtue of action  $A$  averaged across individuals. This measure suffers from a lack of statistical power since each action has only one observation. To leverage our full sample of sentiment data and increase statistical power, we replicate our analyses using  $\widehat{V}_{i,A}$ , which normalizes responses within each individual.

### 2.3 Choice Stage

In the Choice Stage, subjects evaluated all eight actions after being assigned to one of two groups. The first group, the "IC" group, revealed their preferences through choices in an incentive-compatible elicitation. The second group, the "H" group stated their preferences through claims in a hypothetical elicitation. The IC group had 91 subjects and the H group had 96 subjects.<sup>15</sup>

---

<sup>15</sup>We restricted subjects to participate in one survey or the other, and to participate in that survey only once. We dropped 15 submissions from the IC group and 7 from the H group for violating this restriction.

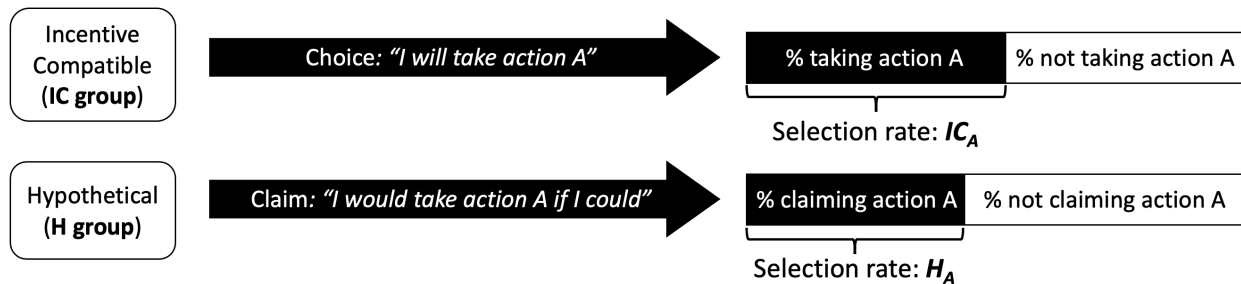
The only difference between the IC and H groups was the incentive-compatibility of the IC group’s choice elicitation. For instance, if a subject in the IC group chose to donate \$1 to St. Jude, then that subject actually sacrificed \$1 of their payment and St. Jude actually received a \$1 donation. If a subject in the H group made such a claim, they sacrificed nothing and St. Jude received nothing. Unlike subjects in the IC group, those in the H group faced no explicit incentives to make claims consistent with their true preferences.

For each action  $A$ , let  $IC_A \in [0\%, 100\%]$  and  $H_A \in [0\%, 100\%]$  denote the “selection rate” for action  $A$  among the IC and H group, respectively. We then define socially desirable responding (SDR) as the overstatement of demand for an action when subjects did not have to pay the cost of taking the action:

$$SDR_A \equiv H_A - IC_A. \tag{3}$$

We consider action  $A$  to be socially desirable if  $SDR_A > 0$ ; that is, the H group inflated their claimed desire to take that action relative to the choices of the IC group. Importantly,  $SDR_A$  can take on negative values, indicating a socially undesirable action. Figure 1 depicts the flow of the Choice Stage for an example where the H group understates demand for an action (i.e.  $SDR_A < 0$ ).

**Figure 1.** Experimental Design: Choice Stage



All subjects received a \$5 participation payment in the Choice Stage. This amount was subject to change for the IC group because one of their decisions was randomly selected to be binding (e.g., if they chose to donate to St. Jude, and this decision was randomly selected

to bind, their payment would decrease by \$1 and St. Jude would gain \$1). All subjects in the Choice Stage were told that they must participate in an additional stage (the Prediction Stage, described below) during which they could earn more money; subjects were not given any description of this additional stage until the Prediction Stage began.

## 2.4 Prediction Stage

All of the subjects who participated in the Choice Stage were required to participate as “predictors” in the Prediction Stage in order to receive their full payment. In addition, we recruited 95 new predictors who had not participated in any previous stage. In total, the Prediction Stage featured 271 subjects: 84 returners from the IC group, 92 returners from the H group, and 95 new predictors. All subjects received a \$5 participation payment for completing this stage along with any earnings gained from accurate predictions.

In the Prediction Stage, predictors observed the exact same descriptions of the actions as subjects in the Choice Stage and were tasked with guessing the choice behavior of the IC group for each of the eight actions.<sup>16</sup> To simplify the procedure, we asked subjects to guess the share of subjects (between 0 and 100, inclusive) from the IC group choosing to take each action. Predictions were incentivized using a Becker-DeGroot-Marschak mechanism (Becker, DeGroot and Marschak, 1964).<sup>17</sup>

For each of the eight actions, predictors made two guesses about the IC group’s selection rate,  $IC_A$ , one before receiving information and one after. Let  $GUESS_{i,1,A}$  denote Predictor  $i$ ’s initial guess. Each predictor was then given a randomly drawn “signal” revealing selections

---

<sup>16</sup>As mentioned in Subsection 2.3 (Footnote 15), some subjects violated the restriction for duplicate participation. These duplicates were discovered after the Prediction Stage completed, meaning that signals about the IC group were drawn prior to dropping these duplicates. Accordingly, predictors were incentivized based on responses from the full dataset. Differences in choice rates between the full dataset and our restricted dataset never differ by more than 1.3 percentage points per action. We limit our analysis to predictors who followed our procedures in order to honor our experimental protocols. However, we will present results on how sentiment and accuracy relate to the full dataset, because that is the dataset from which signals were drawn and guesses were incentivized.

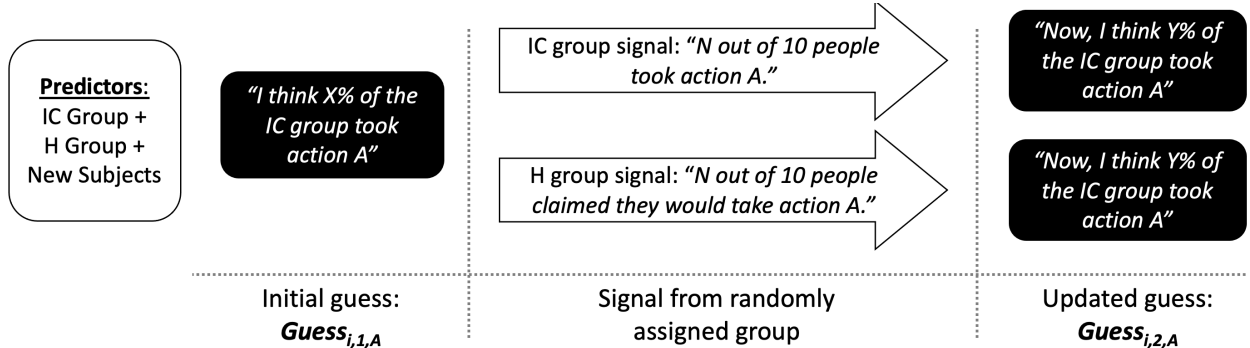
<sup>17</sup>Predictors stood to gain an extra \$5 payment based on the outcome of a lottery. The probability of winning the lottery was either (a) a random draw from a uniform distribution from 0 to 1, or (b) equal to  $IC_A$ . Predictors were paid based on option (a) unless their prediction of  $IC_A$  exceeded their random draw from option (a); in this case, they were paid based on option (b).

from either the IC or H group. Rather than observing the full selection rate, predictors observed a random sub-sampling of behavior. Specifically, predictor  $i$  received a signal,  $s_{i,A} \in \{0, 1, \dots, 10\}$ , conveying the selections on action  $A$  of 10 randomly-sampled respondents from their assigned group.<sup>18</sup> Thus, for information from the IC group,  $s_{i,A} \sim \text{Bin}(10, IC_A)$ ; for information from the H group,  $s_{i,A} \sim \text{Bin}(10, H_A)$ . Note that these signals were drawn with two independent sources of randomness that are critical to our novel identification strategy: random assignment of the information source—the IC or H group—and random sampling of the information *conditional on its source*.

Predictors were given detailed information about the choice procedures of their assigned group so that they could appropriately tailor the weight given to these signals. Predictors were required to correctly answer comprehension questions about these procedures before advancing.<sup>19</sup>

After receiving signals, each predictor submitted updated guesses about the selection rate. Let  $\text{GUESS}_{i,2,A}$  denote predictor  $i$ 's updated guess about the selection rate,  $IC_A$ . Figure 2 depicts the flow of the Prediction Stage.

**Figure 2.** Experimental Design: Prediction Stage



Immediately after revealing their guesses, predictors stated their confidence in each of

<sup>18</sup>More specifically,  $s_{i,A}$  counts the number of 10 randomly-chosen respondents who elected to take action  $A$ . A predictor received such a signal for each action, and thus received 8 signals in total. Furthermore, a predictor's signals were all drawn from the same group: a predictor either observed 8 signals from the IC group or 8 from the H group. The 10 observed respondents who comprise any given signal were randomly drawn with replacement for each  $s_{i,A}$ .

<sup>19</sup>See Appendix Section C for the exact instructions and comprehension questions.



their guesses. This confidence was elicited on a scale from 0 (very uncertain) to 10 (very confident). This elicitation was not incentivized.

## 2.5 Recruitment Summary

Table 1 breaks our sample down by assignment.<sup>20</sup>

**Table 1.** Subject participation by treatment

	Sentiment Stage	Choice Stage	Prediction Stage
Sentiment Group	39 Subjects		
IC Group		91 Subjects	84 Returners
			7 Non-Returners
H Group		96 Subjects	92 Returners
			4 Non-Returners
New Predictors			95 New Subjects
Totals	39 Subjects	187 Subjects	271 Subjects

*Notes:* “Non-Returners” failed to complete the Prediction Stage after successfully completing the Choice Stage.

## 3 Primary Hypotheses

Our research questions focus on our notion of “social sophistication.” We define social sophistication as actively anticipating SDR and appropriately weighting claims from the H group based on their susceptibility to SDR. To assess the extent to which predictors account for SDR, we measure the weight they assign to (potentially biased) signals from the H group relative to the weight assigned to signals from the IC group. Social sophistication requires that predictors both (i) anticipate the existence of SDR and (ii) adjust for the direction and magnitude of the bias.

To develop intuition for social sophistication, we present a stylized model of SDR and derive hypotheses regarding how a socially-sophisticated Bayesian would respond to information that is subject to SDR. We address the relative weight that should be given to responses

<sup>20</sup>Participation in the Prediction Stage was just shy of our target to recruit 100 predictors from each of three groups: (i) IC group participants, (ii) H group participants, and (iii) new participants.

that may be biased by SDR and how this weighting depends on perceptions of the direction and magnitude of SDR.

Recall that the H group faces no incentives based on their claims. Hence, it is costless for these subjects to claim that they would take the socially desirable action if given the opportunity. In contrast, the IC group must face the consequences of their choices. For simplicity, we refer to choices made with consequences as revealing “true” preferences.

Suppose that, due to the lack of consequences, there exists a fraction  $\theta_A \in [0, 1]$  of subjects in the H group who claim a preference toward action  $A$  in the socially desirable way regardless of their true preference.<sup>21</sup> If action  $A$  is virtuous, then such a bias leads subjects in the H group to inflate their claimed desire to take the action relative to the IC group. The expected selection rate in the H group is then  $H_A = (1 - \theta_A)IC_A + \theta_A$ : a fraction  $1 - \theta_A$  of subjects reveal their true preference, and the remaining fraction,  $\theta_A$ , claim they would take action  $A$  regardless of their true preference. Our measure of SDR for action  $A$  (Equation 3) is therefore  $SDR_A = H_A - IC_A = \theta_A(1 - IC_A)$ .

If action  $A$  is instead stigmatized, then H-group subjects will deflate their claimed desire to take the action. Their expected selection rate is thus  $H_A = (1 - \theta_A)IC_A$ : a fraction  $1 - \theta_A$  of subjects again reveal their true preference, while the remaining subjects claim they would refuse the action. Our measure of SDR in this case is  $SDR_A = -\theta_A IC_A$ .

Recall that for each action  $A$ , a predictor in our experiment observes the choices of 10 random subjects from either the IC or H group. If predictor  $i$  is assigned to receive information from the IC group, then  $s_{i,A} \sim \text{Bin}(10, IC_A)$ . If predictor  $i$  instead receives information from the H group, then our stylized model implies that, if action  $A$  is virtuous, then  $s_{i,A} \sim \text{Bin}(10, (1 - \theta_A)IC_A + \theta_A)$ , and if it is stigmatized, then  $s_{i,A} \sim \text{Bin}(10, (1 - \theta_A)IC_A)$ . Although the distribution of signals depends on  $\theta_A$ , we do not assume perfect knowledge of  $\theta_A$  in developing our hypotheses about social sophistication. Our hypotheses hold under uncertainty about the precise value for  $\theta_A$  and focus on directional predictions

---

<sup>21</sup>Equivalently, a fraction  $1 - \theta_A$  of subjects in the H group report honestly despite no explicit incentive to do so. This could be driven, for instance, by a preference for truth telling (e.g., Abeler et al., 2019).

about how knowledge of  $\theta_A$  influences the relative weight given to signals from the IC and H groups.<sup>22</sup>

We now present the hypotheses that we test in Sections 4 and 5.

### 3.1 Socially Desirable Responding

We begin our analysis with manipulation checks to demonstrate (i) SDR is present and predictable—overstatement of claimed demand from the H group correlates with evaluation of virtue from the sentiment group—and (ii) predictors place positive weight on signals—the quality of predictors guesses correlates with the quality of their signals.

By confirming these manipulation checks, we can conclude that claims from the H group do, in fact, possess information relevant for predicting the choices from the IC group. These manipulation checks also rule out the possibility that all differences in the two information sources can be wholly attributed to beliefs about noise or random choice errors. If this were the case, then social sophistication would not provide any improvement toward a predictor’s guesses.

**Manipulation Check 1 (SDR):** *Socially desirable responding will cause the H group to increasingly overstate claimed demand for an action as the perceived virtue of that action grows.*

All else equal, the more virtuous an action is perceived to be, the more beneficial it is to portray oneself as a type who takes that action. Thus, an increase in perceived virtue should increase the incentive to overstate claimed demand. In the model above, if an action  $A'$  is perceived to be more virtuous than action  $A$ , then  $\theta_{A'} > \theta_A$ . That is, more subjects are inclined to lie in the socially desirable way when incentives are removed. The measures of SDR for both virtuous and stigmatized actions, derived above, are increasing in  $\theta_A$ .<sup>23</sup>

---

<sup>22</sup>Below, we will impose an additional key assumption: although signals may influence a predictor’s estimate of  $\theta_A$ , they must also influence their beliefs about  $IC_A$ .

<sup>23</sup>This can be thought of as a local phenomenon, as it assumes that choice rates,  $IC_A$ , are held constant as perceived virtue changes. This is one limitation of such a stylized model, because a sufficiently large

Confirming the predictive validity of our measures of perceived virtue from the sentiment group establishes what we call “sentiment sophistication.” That is, the evaluations of sentiment we collect represent useful social knowledge for predicting choice behavior.

**Manipulation Check 2 (Accuracy):** *If predictors assign positive weight to their signals, then they will be relatively more accurate with information from the IC group.*

Our design cannot identify social sophistication if predictors never update their guesses in response to signals. We present a simple test to demonstrate that predictors assign positive weight to signals—we measure if updated guesses about the IC group are more accurate for predictors who receive their signals from the IC group rather than the H group. That is, do more accurate signals result in more accurate guesses?

### 3.2 Social Sophistication

Our analysis proceeds by evaluating hypotheses about social sophistication—the anticipation and correction for SDR. For these hypotheses, we use our stylized framework to describe how a sophisticated understanding of  $\theta_A$  *should* influence the way that predictors respond to signals.

In a comprehensive review, Benjamin (2019) describes how prevalent statistical biases— independent of the concepts we study here—can generate both over- and under-updating from new information. For this reason, all of our hypotheses about updating focus on differences in updating across the two information sources—how updating differs in response to IC-group and H-group signals—rather than comparisons to the Bayesian benchmark. In this way, we can evaluate social sophistication in isolation instead of evaluating the joint test of social sophistication *and* statistical sophistication.<sup>24</sup>

---

change in the virtue or stigma of an action would likely affect choice rates. However, this should have little impact on our results, as our actions are all within a reasonable range of stigma or virtue. Following the logic of our model provides useful intuition in this context.

<sup>24</sup>Our focus on differential updating across groups also mitigates concerns about anchoring. Since we elicit each subject’s prior and updated guess, they may update insufficiently if their second guess is anchored toward their first. However, our strategy of focusing on differences in updating across groups largely

**Hypothesis 1 (Anticipation of SDR):** *Predictors with social sophistication will give greater weight to incentive-compatible information.*

Social sophistication allows predictors to leverage signals from the H group to make unbiased guesses about  $IC_A$ . However, those guesses will be inherently noisier. Sophisticated predictors will recognize that signals from the H group carry less information and will discount them relative to the more informative signals from the IC group. Thus, with social sophistication, updated guesses about the behavior of the IC group will react more strongly to signals from the IC group.

Hypothesis 1 tests a fundamental aspect of social sophistication. In our stylized model, testing Hypothesis 1 simply amounts to testing whether predictors treat  $\theta_A$  as non-zero.

**Hypothesis 2 (Direction):** *Predictors with social sophistication will discount “perception-inflating” signals from the H group relative to similar signals from the IC group, but they will give greater relative weight to “perception-deflating” signals from the H group.*<sup>25</sup>

The effect of  $\theta_A$  on predictions depends on whether action  $A$  is stigmatized or virtuous. Thus, sophisticated inference requires a predictor to first assess whether an action is virtuous (where  $\theta_A$  correlates with overstatement of claimed demand by the H group) or stigmatized (where  $\theta_A$  correlates with understatement of claimed demand by the H group). Predictors can then categorize the signal they observe as a perception-inflating boast or a perception-deflating confession. We define perception-inflating and perception-deflating signals relative to a predictor’s initial guesses. A predictor’s signal is perception-inflating (-deflating) if it indicates greater (lesser) demand for socially-desirable actions *than the predictor initially guessed*. Full social sophistication requires a heterogeneous treatment of perception-inflating and perception-deflating signals. Sophisticated predictors should discount perception-inflating signals from the H group as they are likely over-optimistic about

---

sidesteps this issue so long as anchoring behavior is independent of group assignment.

<sup>25</sup>This hypothesis was not included in our pre-analysis plan. We include it here and provide results in the following section because they meaningfully add to our understanding of social sophistication among predictors. Our analysis faithfully replicates the analysis we used to evaluate every other hypothesis.

socially-desirable choices. But, in the rare event that a predictor observes a perception-deflating signal from the H group, this signal should be given *more* weight than an equivalent signal from the IC group.<sup>26</sup> This is because a sophisticated predictor realizes that they have observed a perception-deflating signal *despite* the H group’s desire to overstate their claimed demand for socially desirable behavior. Thus, choices of the IC group are probably even lower than this signal suggests.

For example, donations to St. Jude are categorized as virtuous because  $SDR_{St.Jude} > 0$ . Now, suppose a predictor initially guesses that 50% of the IC group would donate and then receives a signal in which 60% of the sampled members of the H group claim they would make the donation. This signal is perception-inflating. It should be discounted relative to a signal in which 60% of the sampled members of the IC group actually choose to donate because the claims from the H group are likely overstated. But, if the same predictor with the same initial guess instead receives a signal from the H group in which only 40% claim they would make the donation, then this signal is perception-deflating. It should be treated as *even more* informative than a signal from the IC group in which 40% choose to donate: if only 40% claim they would donate despite being able to freely lie, then surely the true choice rate is even lower than that.<sup>27</sup>

To summarize this test in terms of our stylized model, we jointly test if predictors (i) identify whether  $\theta_A$  inflates or deflates hypothetical claims about a given action  $A$  and (ii) understand that this makes perception-deflating signals from the H group less likely and, therefore, more informative about  $IC_A$ .

---

<sup>26</sup>SDR predicts that perception-inflating signals will be more likely from the H group than the IC group. Indeed, perception-inflating signals are 16 percentage points more likely when signals arrive from the H group ( $p < 0.001$ ). For this reason, Hypothesis 1 suggested that the claims of the H group should be discounted relative to the choices of the IC group, *on average*.

<sup>27</sup>This implication of social sophistication relies on the assumption that, for a given action, a predictor knows whether subjects in the H group tend to understate or overstate their demand for it. This means that the predictor does not use their signal to infer whether the action is stigmatized or virtuous. If this were the case, then a sophisticated predictor may use the surprisingly low 40% signal from the H group to conclude that donations to St. Jude are in fact *stigmatized*. We doubt that this alternative explanation drives our findings regarding Hypothesis 2 since results from the Sentiment Stage reveal that our actions have predictable stigma or virtue.

**Hypothesis 3 (Relative Magnitude):** *Predictors with social sophistication will increase the relative weight given to incentive-compatible information as the perceived virtue or stigma of an action becomes more extreme.*

The information content of a signal from the H group is decreasing in the share of subjects falsely claiming socially desirable behaviors,  $\theta_A$ . Consider, for example, the boundary case of  $\theta_A = 1$ : signals from the H group then provide no information and should be ignored. Social sophistication suggests that a predictor should account for the relative magnitude of  $\theta_A$  across actions (i.e. which actions have greater or lesser degrees of virtue or stigma). Thus, as the perceived virtue or stigma of an action grows relatively more extreme, we expect sophisticated predictors to increase their discounting of signals from the H group relative to the IC group.

In terms of our stylized model, this amounts to evaluating whether predictors are better than random at ordering  $\theta_A$  across actions.

#### 4 Data Description and Manipulation Checks

In this section, we provide summary statistics for behavior in each stage of the experiment for each of the eight actions. We then provide results from our manipulation checks, demonstrating that (i) socially desirable responding is present and predictable and (ii) our predictors give positive weight to the signals they receive.

Our manipulation checks serve to establish that the bias from SDR is systematic and predictable. That is, differences between the IC and H groups are not exclusively attributable to noise or random choice errors. Absent this confirmation, no degree of social sophistication would allow a predictor to extract information from the statements of the H group that could improve their guesses about the choices of the IC group because no such information would exist. Thus, by confirming our manipulation checks, we confirm sufficient conditions that allow us to test for the presence of social sophistication.

Table 2 presents descriptive results for each action. The Sentiment Stage and Choice

Stage are captured in Columns 1 and Columns 2–3, respectively. Initial and updated guesses from the Prediction Stage are in Columns 4–6. Columns 7–8 compare predictors’ average accuracy across information sources, where accuracy is measured by the absolute difference between a predictor’s updated guess and the true value.

**Table 2.** Summary statistics for each action

Action	Sentiment ( $V_A$ )	Choice Rate		Initial Guesses	Updated Guesses		Updated ABS Error	
		IC Group	H Group		IC Signal	H Signal	IC Signal	H Signal
<b>St Jude Donation</b>	9.15	61.5%	75.0%	60.8%	63.4%	68.8%	12.9	16.8
<b>NPR Donation</b>	6.14	26.4%	31.3%	26.5%	26.5%	26.8%	11.7	14.4
<b>Steal</b>	2.18	25.3%	19.8%	44.7%	30.9%	29.6%	13.5	15.1
<b>Take Donation</b>	4.07	12.1%	6.3%	24.1%	13.6%	11.1%	9.0	9.1
<b>Trump Donation</b>	3.80	11.0%	17.7%	30.1%	18.4%	21.9%	11.5	13.7
<b>Biden Donation</b>	3.74	3.3%	8.3%	24.2%	10.7%	11.0%	8.4	8.9
<b>RNC Donation</b>	4.72	7.7%	20.8%	33.1%	17.6%	24.5%	11.9	17.7
<b>DNC Donation</b>	4.53	12.1%	25.0%	33.6%	18.2%	26.3%	10.3	16.1

*Notes:* This table does not include data from subjects who were dropped from the analysis because of duplicate entries (see Section 2). Sentiment ( $V_A$ ) is a within-subject average of three responses from 0 to 10 about the social desirability of the action.

#### 4.1 Socially Desirable Responding

In order to conduct valid tests of social sophistication among predictors, we must first establish that SDR is present in the signals they receive. Recall that we defined SDR as the difference in selection rates between the H and IC groups ( $SDR_A \equiv H_A - IC_A$ ). Thus, we must first ensure that the H group overstates (understates) claimed demand for socially desirable (undesirable) behaviors relative to the IC group. Additionally, the inflation of claimed demand for an action must not be random, but rather systematically tied to the action’s perceived virtue, which we measured independently during the Sentiment Stage.

Table 3 presents this analysis at two levels of specificity. Column 1 regresses  $SDR_A$  for each of the eight actions on  $V_A$ , the mean perceived virtue of the action from the Sentiment Stage (see Equation 1). Column 2 follows with an individual-level version of this test that regresses  $SDR_A$  on  $\hat{V}_{i,A}$ , the within-subject normalized index of the perceived virtue of each action (see Equation 2).<sup>28</sup>

<sup>28</sup>See Appendix Section B for details on all of our estimation methods.



**Table 3.** Socially desirable responding and perceived virtue

	Socially Desirable Responding	
Mean Sentiment	2.390*	
	(1.12)	
Standardized Sentiment		3.112***
		(0.48)
Constant	-6.080	5.375***
	(5.814)	(0.00)
Observations	8	312
Clusters	N/A	39

*Notes:* “Mean Sentiment” is aggregated across 39 individual evaluations measured from 0 (Very Negative) to 10 (Very Positive). “Standardized Sentiment” normalizes sentiment ( $V_{i,A}$ ) within each individual to have mean 0 and SD 1. Column 1 presents OLS results. Column 2 presents results of a random-effects linear regression with subject-level random effects and standard errors clustered at the subject level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Our measure of SDR is clearly predicted by the evaluations of virtue and stigma from the Sentiment Stage. Column 2 shows that the H group overstates their claimed demand for socially desirable behaviors by an additional 3.1 percentage points for every one standard deviation increase in perceived virtue.<sup>29</sup>

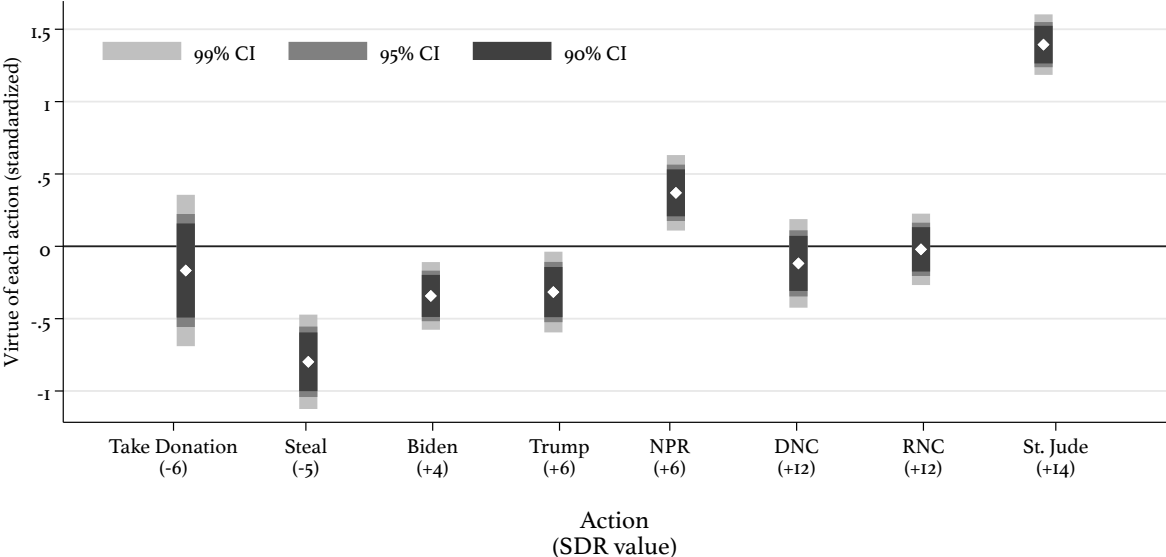
The fact that our subjects’ evaluations of sentiment are predictive of observed SDR demonstrates their “sentiment sophistication:” they have a fairly accurate understanding of the stigma or virtue surrounding an action. Using the knowledge of which actions are more socially desirable—and therefore more likely to inspire dishonest responding from the H group—subjects could tailor their discounting of the H group’s claims to control for SDR. Our tests of Hypotheses 2 and 3 evaluate whether predictors are able to complete this operation and translate knowledge of the relative social desirability of an action—obtained through sentiment sophistication—into an awareness of the resulting bias—a measure of social sophistication.

This sentiment sophistication is presented graphically in Figure 3, which orders each of

<sup>29</sup>Appendix Table A.1 breaks down this association by each of the three components of our sentiment index. The relationships are similar across components, though others’ sentiment and second-order perceptions of sentiment appear to be slightly stronger predictors of SDR than one’s own sentiment.

the eight actions along the horizontal axis according to their observed SDR. For each action, the associated sentiment evaluations are plotted on the vertical axis. There is a clear positive association between the action’s perceived virtue and the SDR in the Choice Stage.

**Figure 3.** Sentiment associated with each action. Actions ordered by *SDR* value.



### 4.2 Accuracy

All predictors were tasked with guessing the behavior of the IC group. Therefore, signals drawn from the choices of the IC group will necessarily be (weakly) more predictive than signals drawn from the claims of the H group. Thus, we can validate that predictors are responsive to signals by testing if higher-quality information (i.e., from the IC group) results in more accurate updated guesses.

Table 4 presents this manipulation check. Columns 1-2 measure accuracy based on the absolute error of a predictor’s guess:  $|IC_A - \text{GUESS}_{i,t,A}|$ , where  $t \in \{1, 2\}$  denotes the initial and updated guess, respectively. Columns 3-4 repeat this analysis using the squared error of a predictor’s guess:  $(IC_A - \text{GUESS}_{i,t,A})^2$ . Since predictors were randomly assigned to receive signals from either the IC or H group after stating their initial guess, the baseline accuracy

was balanced.<sup>30</sup> Therefore, our outcome of interest is the extent to which predictors’ updated guesses become more accurate depending on their information source.

**Table 4.** Improvements in accuracy depending on information source

	Absolute Errors		Squared Errors	
	Updated Error	$\Delta$ Error	Updated Error	$\Delta$ Error
IC Info Source	-2.76*** (0.59)	-2.73*** (1.04)	-115.33*** (30.07)	-98.85 (74.03)
Initial Error	0.24*** (0.02)		0.16*** (0.03)	
Constant	5.02*** (0.68)	-11.41*** (1.24)	113.80*** (33.45)	-576.66*** (90.81)
Mean Initial Error:	21.58		792.99	
Standard Deviation:	(18.09)		(1219.52)	
Observations	2168	2168	2168	2168
Clusters	271	271	271	271

*Notes:* Random-effects linear regression with subject-level random effects. Standard errors clustered at the individual level. Fixed effects are included for each action. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Here, “IC Info Source” is an indicator variable equal to one if the predictor is assigned to receive signals from the IC group. Columns 1-3 show that receiving this higher-quality information causes a large and statistically significant improvement in the accuracy of predictors’ guesses. That is to say, higher-quality signals lead to more accurate updated guesses.

It is important to note that the constant terms estimated in Columns 2 and 4 are negative and significant. Thus, on average, the error in a subject’s guess decreases after receiving information, regardless of the information source. Even the lower-quality signals from the H group improve predictors’ guesses relative to their initial accuracy.

Despite the improvement in accuracy, predictors’ guesses in both groups fall short of a simple benchmark: the accuracy of their signals. Both groups would improve their accuracy by simply making guesses that match their signals exactly.<sup>31</sup> On average, signals from the IC

<sup>30</sup>The p-values for a test of differences in the accuracy of initial guesses are  $p = 0.97$  and  $p = 0.81$  for absolute- and squared-errors, respectively.

<sup>31</sup>Specifically, if one’s signal reveals that  $z$  out of 10 people took the action, then this strategy calls for a guess that the IC choice rate is  $(10 \times z)\%$ .

group have an absolute error of 8.6 percentage points, while the associated updated guesses have an absolute error of 11.1 (test of differences:  $p < 0.001$ ). Signals from the H group have an absolute error of 12.2 percentage points, while the associated updated guesses have an absolute error of 13.9 (test of differences:  $p < 0.001$ ).

## 5 Main Results

Our manipulation checks confirmed that SDR is widespread and predictable and that predictors' guesses are sensitive to their signals. With these prerequisites established, we now proceed to test our hypotheses about social sophistication, exploring the extent to which predictors anticipate and react to SDR. Our analysis closely follows our pre-registration with few amendments. As we test each hypothesis, we will begin with our pre-registered specification before presenting any alternative specifications. Appendix Section B details our empirical estimation and the ways in which we supplement our pre-registered analysis.

### 5.1 Hypothesis 1: Anticipation of SDR

Hypothesis 1 states that social sophistication should cause predictors to give signals from the IC group relatively more weight than those from the H group, on average. This amounts to testing if predictors identify differences in information quality between the two sources and discount hypothetical claims relative to actual choices.

Evaluating a predictor's sensitivity to their idiosyncratic signals from either the IC or H group poses a particular obstacle: participants in the two groups faced different incentives in the Choice Stage, and thus the distribution of signals differs across groups. Therefore, our random assignment of information source is confounded with the assignment of a different mean for the distribution of signals. To resolve this confound and isolate the random sampling variation in signals, we control for the differences in the distributions from which signals are drawn.<sup>32</sup> We accomplish this by including either (i) controls for the mean of the

---

<sup>32</sup>See Kahan (2015) and Thaler (2019) for discussions on why responses to information alone are insufficient

signal distribution or (ii) fixed effects for the distribution. We then causally identify the *differential* impact of signals from the IC group because of our randomly-assigned information source (IC vs. H).

Table 5 presents our test of Hypothesis 1—whether predictors anticipate SDR and accordingly give greater weight to information from the IC group when updating their guesses. Column 1 follows our pre-registration exactly, estimating the updated guess while controlling for the initial guess with additional controls for the mean of the signal distribution. Column 2 examines within-predictor changes in guesses, which increases statistical power. Column 2 also employs a more conservative solution to address the differences in distributions by including fixed effects for each of the 16 combinations of actions and information sources. Columns 3 and 4 replicate the analysis of Column 2 but restrict our sample to newly recruited predictors and experienced predictors, respectively. This allows us explore the role of experience in prompting skepticism toward claims from the H group.

Columns 1 and 2 of Table 5 reveal the skepticism with which predictors treat signals from the H group. Predictors respond to each mechanically-random one-percentage-point increase in a signal from the H group by updating their guesses by 0.55–0.59 percentage points ( $p < 0.001$  for both)—about halfway to the signal. The interaction term “IC Info Source×Signal Value” shows that predictors give signals from the IC group significantly greater weight, confirming Hypothesis 1. A one-percentage-point increase in an IC-group signal results in a *greater* increase in a predictor’s updated guess than an identical increase in an H-group signal. This difference is equal to 0.08–0.17 percentage points ( $p < 0.01$  for both). Equivalently, 14–31% of the updating from IC-group signals is attributable to “extra updating” due to the added weight given to IC-group signals relative to H-group signals.

Concretely, an IC-group signal showing one additional person (out of the 10 sampled) choosing action  $A$  will cause a predictor to increase their guess by 6.7–7.2 percentage points. In contrast, had this signal arrived from the H group, predictors would only increase their

---

to identify differential updating.

**Table 5.** Updated guesses in response to signals from different sources

	Updated Guess		$\Delta$ Guess	
	Full Sample	Full Sample	New Predictors	Experienced Predictors
Signal Value	0.59*** (0.03)	0.55*** (0.05)	0.55*** (0.07)	0.54*** (0.06)
IC Info Source $\times$ Signal Value	0.08*** (0.03)	0.17*** (0.07)	0.11 (0.10)	0.22** (0.09)
Initial Guess	0.30*** (0.02)			
IC Info Source	-1.14 (1.07)			
Observations	2168	2168	760	1408
Clusters	271	271	95	176
Control for Mean Signal:	Yes	N/A	N/A	N/A
Fixed-Effects:	Action	$\times$ Source	$\times$ Source	$\times$ Source

*Notes:* Random-effects linear regression with subject-level random effects. Standard errors clustered at the individual level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

guess by 5.5–5.9 percentage points.

In Appendix Section A.2, we plot each initial and updated guess individually and perform heterogeneity analysis to show that predictors discount signals from the H group at both the intensive and extensive margins. We find suggestive (but not significant) evidence that predictors with signals from the H group are both more likely to entirely ignore their signals, and less likely to submit updated guesses that exactly match their signals.

Predictors who participated in the H group during the Choice Stage may have experienced the temptation to distort their responses, making them more skeptical when receiving signals from the H group. Thus, we use Columns 3 and 4 of Table 5 to test if experience in the Choice Stage is a source of skepticism toward H-group signals. Predictors who previously participated in the Choice Stage give 0.22 percentage points extra weight to each percentage-point increase in IC-group signals relative to H-group signals. On the other hand, newly recruited predictors only give IC-group signals 0.11 percentage points extra weight for

a corresponding increase in signals. The difference between these groups is not significant, though when each prior role—IC group or H group—is analyzed separately, the differential effect of participating in the H group during the Choice Stage approaches marginal significance ( $p = 0.125$ ). These results can be found in Appendix Section A.3.<sup>33</sup>

Our results from Table 5—along with our supplemental analysis in Appendix Section A.2—consistently find that predictors demonstrate a fundamental feature of social sophistication: they anticipate the potential for SDR and respond by discounting the claims of the H group. Full social sophistication, however, involves more complex procedures that we examine next.

## 5.2 Hypothesis 2: Direction of SDR

Here, we test if predictors’ guesses appreciate the direction in which SDR will affect signals from the H group. That is, we ask how accurately predictors recognize whether the H group will tend to overstate or understate their claimed desire to take a given action.

Predictors with social sophistication should increase their discounting of signals from the H group when they are “perception-inflating”—i.e., suggestive of more socially desirable behavior than the predictor’s initial guess—because the H-group signals are drawn from claims that tend to be optimistic exaggerations. Conversely, social sophistication guides predictors to give *more* weight to signals from the H group when they are “perception-deflating”—i.e., suggestive of less socially desirable behavior—because in the rare case that a perception-deflating signal is drawn from the H group’s optimistic claims, it suggests that the initial guess is particularly mistaken and the predictor should respond strongly to the signal. For a concrete example and further details on this logic, see the discussion of Hypothesis 2 in Section 3.

To evaluate this hypothesis, we must first designate which actions are socially desirable. We do so empirically using  $SDR_A$ . If  $SDR_A > 0$ —that is, the H group overstates their demand for action  $A$ —then  $A$  is considered virtuous. Otherwise, if  $SDR_A < 0$ , then  $A$  is

---

<sup>33</sup>Table A.3 contains our pre-registered analysis of the role of experience on social sophistication. The results are qualitatively similar to those in Columns 3 and 4 of Table 5.

considered stigmatized.<sup>34,35</sup> With knowledge of an action’s social desirability, social sophistication will enable predictors to determine if the signal they receive is perception-inflating or perception-deflating. A perception-inflating signal is one that indicates a greater demand for an action with  $SDR_A > 0$  (or a lesser demand for an action with  $SDR_A < 0$ ) *than the predictor initially guessed*. A perception-deflating signal indicates a lesser demand for an action with  $SDR_A > 0$  (or a greater demand for an action with  $SDR_A < 0$ ) *than the predictor initially guessed*. Note that there are no absolute thresholds for perception-inflating (-deflating) signals; they are categorized based on whether they indicate greater (lesser) social-desirability than the predictor’s initial guess.

Our test of Hypothesis 2 modifies the approach of Hypothesis 1 to test if the relative weighting of H-group signals depends on whether they are perception-inflating or perception-deflating. To aid the interpretation of coefficients, we will replicate the analysis of Hypothesis 1 separately for predictors receiving perception-inflating and perception-deflating signals.<sup>36</sup>

Table 6 displays our limited support for Hypothesis 2. Columns 1 and 2 replicate the analysis of Columns 1 and 2 from Table 5 but restrict their focus to perception-inflating signals. In this direction, signals from the H group should be discounted relative to those from the IC group. We find a positive coefficient for “IC Info Source×Signal Value,” revealing that H-group signals receive less weight than IC-group signals. For every one-percentage-point increase in signals, Column 1 shows that predictors’ guesses increase by 0.11 percentage points less when the signals arrive from the H group ( $p < 0.001$ ). Column 2 estimates this diminished weight to be 0.12 ( $p = 0.211$ ). Thus, predictors do appear to recognize that perception-inflating signals are less credible when they come from the H group instead of

---

<sup>34</sup>Note that all of our actions have  $SDR_A > 0$  except for stealing from another subject and taking money from the Make-A-Wish Foundation.

<sup>35</sup>One could consider using our measure of perceived virtue from the Sentiment Stage to identify stigmatized actions. However, this approach presents an issue with units because the Likert scale we used to measure sentiment does not have an obvious cutoff for socially-desirable and socially-undesirable actions.

<sup>36</sup>We conduct similar analysis using a fully-interacted specification in Appendix Section A.4. The results are qualitatively similar, but even more inconsistent with social sophistication.



**Table 6.** Updated guesses in response to perception-inflating and -deflating signals from different sources

	Perception-Inflating		Perception-Deflating	
	Updated Guess	$\Delta$ Guess	Updated Guess	$\Delta$ Guess
Signal Value	0.61*** (0.05)	0.26*** (0.07)	0.71*** (0.03)	0.33*** (0.06)
IC Info Source $\times$ Signal Value	0.11*** (0.04)	0.12 (0.09)	-0.01 (0.03)	0.08 (0.09)
IC Info Source	-2.06 (1.27)		0.67 (1.28)	
Initial Guess	0.29*** (0.04)		0.23*** (0.03)	
Observations	925		1243	
Clusters	267		270	
Control for Mean Signal:	Yes	N/A	Yes	N/A
Fixed-Effects:	Action	Action $\times$ Source	Action	Action $\times$ Source

*Notes:* Random-effects linear regression with subject-level random effects. Standard errors clustered at the individual level. “Perception-Inflating” (“Perception-Deflating”) are defined by whether the signal is in the direction of more (less) social desirability relative to the initial guess. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

the IC group, though the statistical significance of this result depends on the specification.

Social sophistication also requires recognizing that perception-deflating signals should be given relatively *more* weight when they come from the H group. Columns 3 and 4 reveal no evidence for this more complex dimension of social sophistication. In Column 3, we find a near zero and insignificant coefficient for “IC Info Source $\times$ Signal Value,” revealing no difference in the weight given to H-group signals. Note that this coefficient would be negative if predictors put relatively more weight on H-group signals as predicted by social sophistication. Additionally, in the specification of Column 4, this effect is positive, meaning that predictors still discount signals from the H group even when they are perception-deflating, though this effect is not significant. Thus, we conclude that predictors’ guesses demonstrate no recognition that perception-deflating signals from the H group are even stronger indictments of behavior than corresponding signals from the IC group.<sup>37</sup>

<sup>37</sup>In Appendix Section A.4, we present individual guesses in Figures A.3 and A.4 to visualize heterogeneous discounting with respect to the direction of SDR. These figures mirror the approach taken in Figures A.1 and A.2 (which visualize average discounting).

### 5.3 Hypothesis 3: Relative Magnitude of SDR

We now test if predictors appreciate which claims from the H group are more susceptible to SDR and, therefore, more worthy of discounting. Table 3 and Figure 3 show how Sentiment Stage responses can predict which actions generate the strongest image concerns. Here, we examine whether predictors apply such knowledge when interpreting claims from the H group.

Our test of Hypothesis 3 adapts the approach of Hypothesis 1 to include interaction terms for the observed level of  $SDR_A$ . As  $SDR_A$  grows in magnitude, the claims of the H group are increasingly distorted by SDR and social sophistication prescribes greater discounting for such claims. Since the discounting of H signals should increase with the magnitude of  $SDR_A$  regardless of its sign, we use its absolute value,  $|SDR_A| = |H_A - IC_A|$ , as our interaction term.

To demonstrate robustness to an alternative measure of SDR, and to directly connect social sophistication with sentiment sophistication, we repeat the analysis above using responses from the Sentiment Stage as the interaction term. Table 3 and Figure 3 confirm that SDR tends to increase in magnitude as the sentiment group’s evaluations of an action become more extreme; thus, predictors should increasingly discount signals from the H group for such actions. Our interaction term in this case is the absolute value of a normalized measure of sentiment at the action-level:  $|\widehat{V}_A| = \left| \frac{V_A - \bar{V}}{\sigma_V} \right|$ , where  $\bar{V}$  and  $\sigma_V$  are the mean and standard deviation of  $V_A$  (Equation 1) across all eight actions.

Table 7 presents our test of Hypothesis 3. We find no evidence that predictors increase their relative discounting of signals from the H group as either SDR or perceived virtue become more pronounced. In Columns 1 and 2, the additional discounting of the H group is captured by the coefficient for “IC Info Source  $\times$  Signal Value  $\times$   $|SDR|$ .” We find no evidence of increased discounting of H-group claims for actions with greater SDR. In fact, we find point estimates in the wrong direction. In Column 3, the additional discounting of the H group is captured by the coefficient for “IC Info Source  $\times$  Signal Value  $\times$   $|\widehat{V}_A|$ .” As in Columns 1 and 2,

predictors fail to increase their discounting of claims from the H group as  $|\widehat{V}_A|$  grows, with point estimates again in the wrong direction. Thus, Table 7 rejects the notion that predictors tailor their inferences to the relative magnitude of bias from SDR.

**Table 7.** Updated guesses in response to SDR magnitude by information source

	Updated Guess	$\Delta$ Guess	$\Delta$ Guess
Signal Value	0.60*** (0.05)	0.57*** (0.12)	0.63*** (0.06)
IC Info Source $\times$ Signal Value	0.06 (0.07)	0.38** (0.17)	0.24** (0.09)
Signal Value $\times SDR $	-0.00 (0.00)	-0.00 (0.01)	
IC Info Source $\times$ Signal Value $\times SDR $	0.00 (0.01)	-0.02 (0.02)	
IC Info Source $\times SDR $	-0.14 (0.16)	-0.46 (0.47)	
$ SDR $	0.17 (0.13)	-0.40 (0.90)	
Initial Guess	0.32*** (0.02)		
IC Info Source	0.19 (1.58)		
Signal Value $\times \widehat{V}_A $			-0.10** (0.05)
IC Info Source $\times$ Signal Value $\times \widehat{V}_A $			-0.04 (0.07)
IC Info Source $\times \widehat{V}_A $			-4.66 (5.20)
$ \widehat{V}_A $			35.81 (57.97)
Observations		2168	
Clusters		271	
Control for Mean Signal:	Yes	N/A	N/A
Fixed-Effects:	Action	Action $\times$ Source	Action $\times$ Source

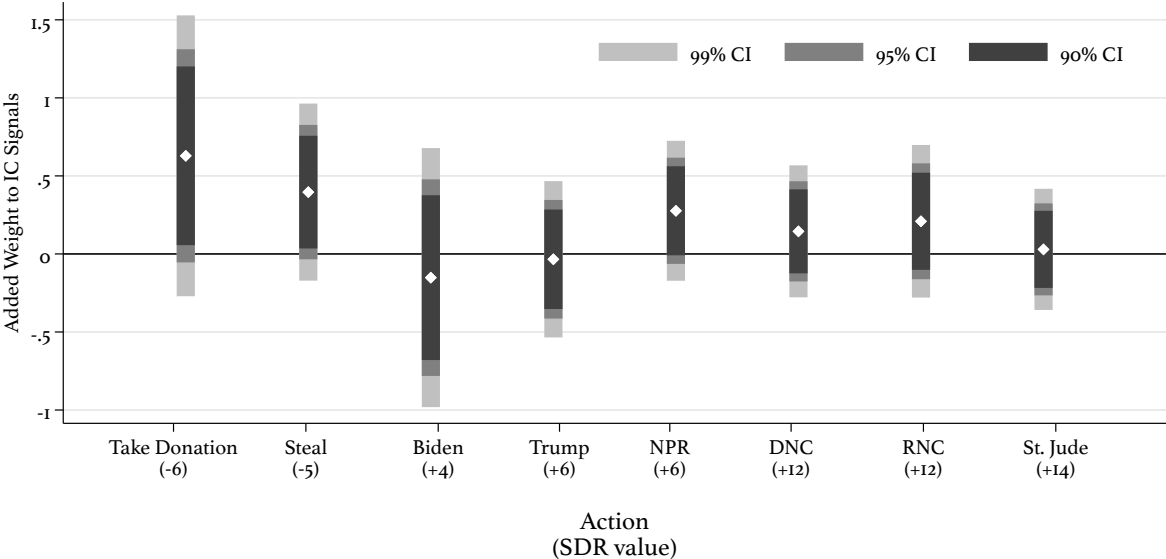
*Notes:* Random-effects linear regression with subject-level random effects. Standard errors clustered at the individual level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Figure 4 visually depicts this lack of social sophistication.<sup>38</sup> Just as in Figure 3, actions

<sup>38</sup>Figure 4 presents coefficients and confidence intervals for “IC Info Source $\times$ Signal Value” separately for each action. The specification is drawn from Column 2 of Table 5 and replicated for each individual action. We include an indicator variable for “IC Info Source,” since we cannot include fixed effects for each

are ordered by their SDR value, with extreme negative values on the left and extreme positive values on the right. With social sophistication, the relative weight given to IC signals (and hence the relative discounting of H signals) should grow at the extremes. We find no such pattern. In fact, for the action with the most extreme SDR—donations to St. Jude Children’s Hospital—predictors do not discount H signals relative to IC signals at all. Thus, in contrast to Figure 3—which demonstrated clear sentiment sophistication—Figure 4 finds no evidence of social sophistication.

**Figure 4.** Weight given to signals from the IC group. Actions ordered by *SDR* value.



Recall from our discussion in Section 3 that Hypothesis 3 provides a rather weak test of social sophistication—it amounts to testing whether predictors’ guesses account for the relative SDR across actions in a way that is better than random. As is evident from Figure 4, predictors fail this test. This failure is striking because the test is a natural extension of the tests from Table 3 and Figure 3. In those, the sentiment group demonstrates a clear understanding of which actions tend to be more virtuous or stigmatized. Thus, it appears that predictors fail to translate the sentiment sophistication that is clearly present in the population into the discounting behaviors prescribed by social sophistication.

---

combination of action and information source. All regressions cluster standard errors at the subject level.

## 5.4 Confidence in Predictions

Immediately after making a guess, we asked predictors to state their confidence in that guess on a scale from 0 to 10. Although these elicitations were not incentivized, they provide further insight on the perceived differences between the two information sources. Table 8 examines the association between confidence and the accuracy of a guess. We specifically focus on how higher-quality information from the IC group influences this relationship. Since IC-group signals are weakly more informative, socially-sophisticated predictors who appreciate this fact should display greater increases in confidence when they receive information from the IC group.<sup>39</sup> With our random assignment of the information source, we can causally identify the relationship between higher-quality information and confidence in predictions.

Our analysis uses absolute errors to measure accuracy, meaning that positive numbers indicate diminished accuracy. Initial confidence and updated confidence are both normalized across all individuals and actions to have a mean of 0 and standard deviation of 1.

**Table 8.** Confidence in predictions

	Initial Confidence	Updated Confidence
Initial Error (Absolute Value)	0.004*** (0.001)	-0.006*** (0.001)
Updated Error (Absolute Value)		0.004** (0.002)
Updated Error×IC Info Source		-0.002 (0.003)
Initial Confidence		0.457*** (0.023)
IC Info Source		-0.015 (0.077)
Constant	-0.298*** (0.070)	0.119* (0.062)
Observations		2168
Clusters		271
Fixed-Effects:	Action	Action

*Notes:* Random-effects linear regression with subject-level random effects. Standard errors clustered at the individual level. Confidence is normalized to mean 0 and standard deviation of 1. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

<sup>39</sup>The analysis of predictor confidence is not included in our pre-analysis plan. However, our framework of analysis mirrors that of the pre-registered hypotheses, and we believe it substantively adds to our understanding of the impact of (perceived) information quality.

Column 1 shows a false confidence by predictors. Similar to the classic result from Kruger and Dunning (1999), there is a negative and significant relationship between the accuracy of a predictor’s initial guess and their initial confidence ( $p < 0.01$ ). Column 2 shows that this false confidence effect persists for the updated guesses of predictors who receive information from the H group ( $p < 0.05$ ). However, the negative association between accuracy and confidence is greatly diminished and statistically insignificant for predictors who receive higher-quality information from the IC group ( $p > 0.40$ ).<sup>40</sup> Interestingly, receiving information from the IC group has a near-zero and not significant level effect on confidence, despite the better information. Taken together, these results suggest not only that predictors fail to account for biased claims from the H group, but that this failure may drive second-order consequences such as the persistence of unfounded confidence in erroneous predictions.

## 6 Discussion and Conclusion

In our experiment, we designed an environment to cleanly identify SDR across several different actions. We then asked subjects to predict choice behavior for the actions. We presented subjects with random subsamples of data from either incentivized choices or unincentivized polls to assist them in their predictions. This novel subsampling approach offers a cleaner causal identification of responses to information than the traditional paradigm of information-revelation or belief-correction experiments. The traditional approach reveals identical information to all subjects, meaning that the direction of updating is endogenous to prior beliefs. We believe our approach can alleviate these endogeneity concerns.

When our subjects were presented with data from unincentivized polls, they showed limited “social sophistication” in controlling for the SDR manifest in those poll data. Our subjects correctly put less weight on what others claimed they would do relative to what others actually did. However, they faltered in calibrating their discounting to the SDR of each action. Despite other subjects from the same population showing a clear ability to identify

---

<sup>40</sup>Gneezy and Serra-Garcia (2019) find similar overconfidence in one’s ability to detect lies by others.

the social desirability of actions, predictors failed to translate this knowledge into sophisticated discounting. While our subjects correctly discounted perception-inflating signals, they incorrectly responded to perception-deflating signals. This is a failure of sophistication: subjects should appreciate that perception-deflating signals are especially informative because few people will lie to make themselves look less socially desirable. Further, when considering actions with more extreme social desirability—which inspired more dishonest hypothetical claims—subjects did not increase their discounting.

Our setting was designed to maximize the control over outside variables in order to cleanly identify biases from SDR. In such an abstract environment where subjects are carefully observed, we might expect that biased reporting due to SDR would be relatively salient. In light of this, the limited evidence we find for social sophistication among predictors is even more striking. We should be skeptical of how well peoples’ inferences will control for more subtle forms of SDR in natural settings if they do not account for the blatant SDR in our contrived environment. However, further research is needed to determine the impact of contextual factors on social sophistication.

Other notions of “sophistication” in behavioral economics typically require the recognition and anticipation of one’s own biases. Such sophistication is rare (Augenblick and Rabin, 2019; Ericson, 2011; Heidhues and Kőszegi, 2010). Although social sophistication in our setting does not require any self-reflection—it only requires participants to recognize that *others* may succumb to social desirability bias—we still find limited evidence of sophistication.<sup>41</sup>

A failure to correct for biases from SDR has significant economic costs. Election results, public-health issues, job-market forecasts, and social-policy preferences are all frequently predicted using unincentivized poll data that are susceptible to SDR.<sup>42</sup> Our study demonstrates

---

<sup>41</sup>A literature on “bias blind spots” finds that people possess a greater ability to recognize others’ biases than their own (Pronin, Lin and Ross, 2002; West, Meserve and Stanovich, 2012). Fedyk (2018) demonstrates this asymmetry in the domain of intertemporal choice.

<sup>42</sup>Polls are used to determine candidate viability and access to debate stages (Fox News, 2016), they influence voter turnout (Agranov, Goeree, Romero and Yariv, 2018; Bursztyn, Cantoni, Funk and Yuchtman, 2021; Großer and Schram, 2010) and reported preferences (Cantú and Márquez, 2021), affect campaign contributions (Adkins and Dowdle, 2002), and may help entrench illiberal regimes (Carlson, 2018). Boukouras, Jennings, Li and Maniadis (2020) find that, even in abstract environments, biased polls inhibit objective

systematic failures in the interpretation of such poll data. Although our poll data should not be interpreted at face value, we find that people do not exhibit the social sophistication necessary to de-bias the data themselves. In this way, biased poll data may carry over into biased inferences and sub-optimal actions.

---

evaluation of candidates and shift electoral outcomes. The influence of polls is so significant that a market has arisen for “fake polls” that manipulate asset prices (Yeargain, 2020). In this way, polling biases have economic costs even absent any biases in how individuals consume and interpret them.



## References

- Abeler, Johannes, Nosenzo, Daniele and Raymond, Collin.** (2019). ‘Preferences for Truth-Telling’, *Econometrica* 87(4), 1115–1153.
- Adkins, Randall E and Dowdle, Andrews J.** (2002). ‘The Money Primary: What Influences the Outcome of Pre-Primary Presidential Nomination Fundraising?’, *Presidential Studies Quarterly* 32(2), 256–275.
- Agranov, Marina, Goeree, Jacob K, Romero, Julian and Yariv, Leeat.** (2018). ‘What makes voters turn out: The effects of polls and beliefs’, *Journal of the European Economic Association* 16(3), 825–856.
- Arnold, Hugh J, Feldman, Daniel C and Purbhoo, Mary.** (1985). ‘The role of social-desirability response bias in turnover research’, *Academy of Management Journal* 28(4), 955–966.
- Augenblick, Ned and Rabin, Matthew.** (2019). ‘An experiment on time preference and misprediction in unpleasant tasks’, *Review of Economic Studies* 86(3), 941–975.
- Bagwell, Laurie Simon and Bernheim, B Douglas.** (1996). ‘Veblen effects in a theory of conspicuous consumption’, *The American Economic Review* pp. 349–373.
- Becker, Gordon M, DeGroot, Morris H and Marschak, Jacob.** (1964). ‘Measuring utility by a single-response sequential method’, *Behavioral Science* 9(3), 226–232.
- Bénabou, Roland and Tirole, Jean.** (2002). ‘Self-confidence and personal motivation’, *The Quarterly Journal of Economics* 117(3), 871–915.
- Benjamin, Daniel J.** (2019), Errors in probabilistic reasoning and judgment biases, in ‘Handbook of Behavioral Economics: Applications and Foundations’, Vol. 2, Elsevier, pp. 69–186.
- Bharadwaj, Prashant, Pai, Mallesh M and Suziedelyte, Agne.** (2017). ‘Mental health stigma’, *Economics Letters* 159, 57–60.
- Boukouras, Aristotelis, Jennings, Will, Li, Lunzheng and Maniadis, Zacharias.** (2020), Can Biased Polls Distort Electoral Results? Evidence From The Lab, Working paper, School of Business, University of Leicester.
- Braghieri, Luca.** (2021), Political Correctness, Social Image, and Information Transmission, Working paper.
- Brownback, Andy and Novotny, Aaron.** (2018). ‘Social desirability bias and polling errors in the 2016 presidential election’, *Journal of Behavioral and Experimental Economics* 74, 38–56.
- Brown-Iannuzzi, Jazmin L, Najle, Maxine B and Gervais, Will M.** (2019). ‘The illusion of political tolerance: Social desirability and self-reported voting preferences’, *Social Psychological and Personality Science* 10(3), 364–373.

- Bursztyn, Leonardo, Cantoni, Davide, Funk, Patricia and Yuchtman, Noam.** (2021), Do Polls Affect Elections? Evidence from Swiss Referenda, Working paper.
- Bursztyn, Leonardo, Ferman, Bruno, Fiorin, Stefano, Kanz, Martin and Rao, Gautam.** (2018). ‘Status Goods: Experimental Evidence from Platinum Credit Cards’, *The Quarterly Journal of Economics* 133(3), 1561–1595.
- Cantú, Francisco and Márquez, Javier.** (2021). ‘The effects of election polls in Mexico’s 2018 presidential campaign’, *Electoral Studies* 73, 102379.
- Carlson, Elizabeth.** (2018). ‘The perils of pre-election polling: Election cycles and the exacerbation of measurement error in illiberal regimes’, *Research & Politics* 5(2), 2053168018774728.
- Charness, Gary, Oprea, Ryan and Yuksel, Sevgi.** (2021). ‘How do people choose between biased information sources? Evidence from a laboratory experiment’, *Journal of the European Economic Association* 19(3), 1656–1691.
- Coffman, Katherine B, Coffman, Lucas C and Ericson, Keith M Marzilli.** (2017). ‘The size of the LGBT population and the magnitude of antigay sentiment are substantially underestimated’, *Management Science* 63(10), 3168–3186.
- Crawford, Vincent.** (1998). ‘A survey of experiments on communication via cheap talk’, *Journal of Economic theory* 78(2), 286–298.
- Crawford, Vincent P and Sobel, Joel.** (1982). ‘Strategic information transmission’, *Econometrica: Journal of the Econometric Society* pp. 1431–1451.
- Del Boca, Frances K and Noll, Jane A.** (2000). ‘Truth or consequences: the validity of self-report data in health services research on addictions’, *Addiction* 95(11s3), 347–360.
- de Quidt, Jonathan, Haushofer, Johannes and Roth, Christopher.** (2018). ‘Measuring and bounding experimenter demand’, *American Economic Review* 108(11), 3266–3302.
- Edwards, Allen L.** (1957), *The social desirability variable in personality assessment and research*, Dryden Press.
- Ericson, Keith M. Marzilli.** (2011). ‘Forgetting We Forget: Overconfidence and Memory’, *Journal of the European Economic Association* 9(1), 43–60.
- Eyster, Erik and Rabin, Matthew.** (2005). ‘Cursed equilibrium’, *Econometrica* 73(5), 1623–1672.
- Farrell, Joseph and Rabin, Matthew.** (1996). ‘Cheap talk’, *Journal of Economic perspectives* 10(3), 103–118.
- Fedyk, Anastassia.** (2018). ‘Asymmetric naivete: Beliefs about self-control’, *Available at SSRN 2727499* .

- Finkel, Steven E, Guterbock, Thomas M and Borg, Marian J.** (1991). ‘Race-of-interviewer effects in a preelection poll Virginia 1989’, *Public Opinion Quarterly* 55(3), 313–330.
- Fischbacher, U and Föllmi-Heusi, F.** (2013). ‘Lies in Disguise—An Experimental Study on Cheating’, *Journal of the European Economic Association* 11(3), 525–547.
- Fox News.** (2016). ‘See Which Candidates Qualified for the Fox News-Google GOP Debates’, *Fox News* <http://insider.foxnews.com/2016/01/26/lineup-republican-candidates-fox-news-google-gop-debates>.
- Gneezy, Uri and Serra-Garcia, Marta.** (2019), Mistakes and Overconfidence in Detecting Lies, Working paper.
- Gonzalez-Ocantos, Ezequiel, De Jonge, Chad Kiewiet, Meléndez, Carlos, Osorio, Javier and Nickerson, David W.** (2012). ‘Vote buying and social desirability bias: Experimental evidence from Nicaragua’, *American Journal of Political Science* 56(1), 202–217.
- Großer, Jens and Schram, Arthur.** (2010). ‘Public opinion polls, voter turnout, and welfare: An experimental study’, *American Journal of Political Science* 54(3), 700–717.
- Haaland, Ingar, Roth, Christopher and Wohlfart, Johannes.** (2020), Designing information provision experiments, Working Paper 20/20, CEBI Working Paper Series.
- Hales, Jeffrey.** (2009). ‘Are investors really willing to agree to disagree? An experimental investigation of how disagreement and attention to disagreement affect trading behavior’, *Organizational Behavior and Human Decision Processes* 108(2), 230–241.
- Heerwig, Jennifer A and McCabe, Brian J.** (2009). ‘Education and social desirability bias: The case of a Black presidential candidate’, *Social Science Quarterly* 90(3), 674–686.
- Heidhues, Paul and Köszegi, Botond.** (2010). ‘Exploiting naivete about self-control in the credit market’, *American Economic Review* 100(5), 2279–2303.
- Holbrook, Allyson L. and Krosnick, Jon A.** (2010). ‘Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique’, *Public Opinion Quarterly* 74, 37–67.
- Hopkins, Daniel J.** (2009). ‘No more wilder effect, never a Whitman effect: When and why polls mislead about Black and feMale candidates’, *The Journal of Politics* 71(3), 769–781.
- Janus, Alexander L.** (2010). ‘The Influence of Social Desirability Pressures on Expressed Immigration Attitudes’, *Social Science Quarterly* 91, 928–946.
- Jones, Ann E and Elliot, Marta.** (2016). ‘Examining Social Desirability in Measures of Religion and Spirituality Using the Bogus Pipeline’, *Review of Religious Research* pp. 1–18.

- Kahan, Dan M.** (2015). ‘The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it’, *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource* pp. 1–16.
- Kane, James G, Craig, Stephen C and Wald, Kenneth D.** (2004). ‘Religion and presidential politics in Florida: A list experiment’, *Social Science Quarterly* 85(2), 281–293.
- Karlan, Dean S and Zinman, Jonathan.** (2012). ‘List randomization for sensitive behavior: An application for measuring use of loan proceeds’, *Journal of Development Economics* 98(1), 71–75.
- Kartik, Navin.** (2009). ‘Strategic communication with lying costs’, *The Review of Economic Studies* 76(4), 1359–1395.
- Kruger, Justin and Dunning, David.** (1999). ‘Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments.’, *Journal of Personality and Social Psychology* 77(6), 1121.
- Krupka, Erin L and Weber, Roberto A.** (2013). ‘Identifying social norms using coordination games: Why does dictator game sharing vary?’, *Journal of the European Economic Association* 11(3), 495–524.
- Krysan, Maria.** (1998). ‘Privacy and the expression of white racial attitudes: A comparison across three contexts’, *Public Opinion Quarterly* pp. 506–544.
- Latkin, Carl A, Edwards, Catie, Davey-Rothwell, Melissa A and Tobin, Karin E.** (2017). ‘The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland’, *Addictive Behaviors* 73, 133–136.
- Maccoby, Eleanor E and Maccoby, Nathan.** (1954). ‘The interview: A tool of social science’, *Handbook of Social Psychology* 1, 449–487.
- Paulhus, Delroy L.** (1984). ‘Two-component models of socially desirable responding’, *Journal of Personality and Social Psychology* 46(3), 598–609.
- Powell, Richard J.** (2013). ‘Social desirability bias in polling on same-sex marriage ballot measures’, *American Politics Research* 41(6), 1052–1070.
- Pronin, Emily, Lin, Daniel Y and Ross, Lee.** (2002). ‘The bias blind spot: Perceptions of bias in self versus others’, *Personality and Social Psychology Bulletin* 28(3), 369–381.
- Raghavarao, Damaraju and Federer, Walter T.** (1979). ‘Block total response as an alternative to the randomized response method in surveys’, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 40–45.
- Reeves, Keith et al.** (1997), *Voting hopes or fears?: White voters, black candidates & racial politics in America*, Oxford University Press on Demand.

- Rosenfeld, Bryn, Imai, Kosuke and Shapiro, Jacob N.** (2016). ‘An empirical validation study of popular survey methodologies for sensitive questions’, *American Journal of Political Science* 60(3), 783–802.
- Samuelson, William and Bazerman, Max.** (1985), The Winner’s Curse in Bilateral Negotiations, in **Vernon Smith.**, ed., ‘Research in Experimental Economics’, Vol. 3, JAI Press, Greenwich, CT, pp. 105–137.
- Stephens-Davidowitz, Seth.** (2014). ‘The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data’, *Journal of Public Economics* 118, 26–40.
- Streb, Matthew J., Burrell, Barbara, Frederick, Brian and Genovese, Michael A.** (2008). ‘Social Desirability Effects and Support for a Female American President’, *Public Opinion Quarterly* 72, 76–89.
- Thaler, Michael.** (2019), The “Fake News” Effect: An Experiment on Motivated Reasoning and Trust in News, Working paper.
- Tourangeau, Roger, Rips, Lance J and Rasinski, Kenneth.** (2000), *The psychology of survey response*, Cambridge University Press.
- Tourangeau, Roger and Yan, Ting.** (2007). ‘Sensitive questions in surveys.’, *Psychological bulletin* 133(5), 859–883.
- Veblen, Thorstein.** (1899), *The theory of the leisure class: An economic study of institutions.*
- Warner, Stanley L.** (1965). ‘Randomized response: A survey technique for eliminating evasive answer bias’, *Journal of the American Statistical Association* 60(309), 63–69.
- West, Richard F, Meserve, Russell J and Stanovich, Keith E.** (2012). ‘Cognitive sophistication does not attenuate the bias blind spot.’, *Journal of personality and social psychology* 103(3), 506–519.
- Yeargain, Tyler.** (2020). ‘Fake Polls, Real Consequences: The Rise of Fake Polls and the Case for Criminal Liability’, *Missouri Law Review* 85(1), 7.

## A Appendix A: Supplemental Analysis

### A.1 Breakdown of Sentiment Measures

Table 3 captures the relationship between SDR and our sentiment index, which is constructed by taking the mean of the three measures of sentiment listed below. In this section, we replicate the analysis of Table 3 after breaking down our sentiment index into these component parts. Below, Table A.1 explores the association between SDR and each of the following sentiment measures:

1. How would you feel about taking this action yourself?
2. How would you feel about other people who take this action?
3. How do you think most other people would feel about people who take this action?

For each action  $A$ , let  $Q_{i,j,A}$  denote subject  $i$ 's response to question  $j \in \{1, 2, 3\}$  above. For each of these three measures, we regress  $SDR_A$  on the sentiment rating averaged over individuals,  $\bar{Q}_{j,A} \equiv \frac{\sum_{i=1}^{N_S} Q_{i,j,A}}{N_S}$ . The results of these regressions are reported in Columns 1, 3, and 5 of Table A.1. We also regress  $SDR_A$  on these same sentiment measures after standardizing them within an individual; that is, we regress  $SDR_A$  on  $\hat{Q}_{i,j,A} \equiv \frac{Q_{i,j,A} - \bar{Q}_{i,j}}{\sigma_{i,j}}$ , where  $\bar{Q}_{i,j}$  and  $\sigma_{i,j}$  are subject  $i$ 's mean and standard deviation of  $Q_{i,j,A}$  for measure  $j$  across all eight actions. The results of these regressions are reported in Columns 2, 4, and 5 of Table A.1. Note that the column headers (e.g., "Measure 1") in Table A.1 indicates which of the three questions above are used to form the regressor.

From these results, we can see consistent relationships between different measures of stigma and the observed socially desirable responding in the Choice Stage. While these relationships are all positive and most are significant, there appears to be a stronger association between anticipation of others' sentiment (Columns 5–6) rather than own-sentiment (Columns 1–2) or sentiment towards others (Columns 3–4). This would suggest that people may be more worried about the virtue or stigma they think others will attach to an action

**Table A.1.** Socially desirable responding and perceived virtue

	Socially Desirable Responding					
	Measure 1		Measure 2		Measure 3	
Mean Sentiment	2.030 (1.28)		2.434** (0.98)		2.504* (1.10)	
Standardized Sentiment	2.156*** (0.48)		3.237*** (0.47)		3.489*** (0.43)	
Constant	-3.448 (6.06)	5.375*** (0.00)	-7.058 (5.39)	5.375*** (0.00)	-6.944 (5.83)	5.375*** (0.00)
Observations	8	312	8	312	8	312
Clusters	N/A	39	N/A	39	N/A	39

*Notes:* “Mean Sentiment” is aggregated across 39 individual evaluations measured from 0 (Very Negative) to 10 (Very Positive). “Standardized Sentiment” normalizes sentiment ( $V_{i,j,A}$ ) within each individual to have mean 0 and SD 1. For each of our three sentiment measures, the first column presents OLS results. The second column presents results of a random-effects linear regression with subject-level random effects and standard errors clustered at the subject level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

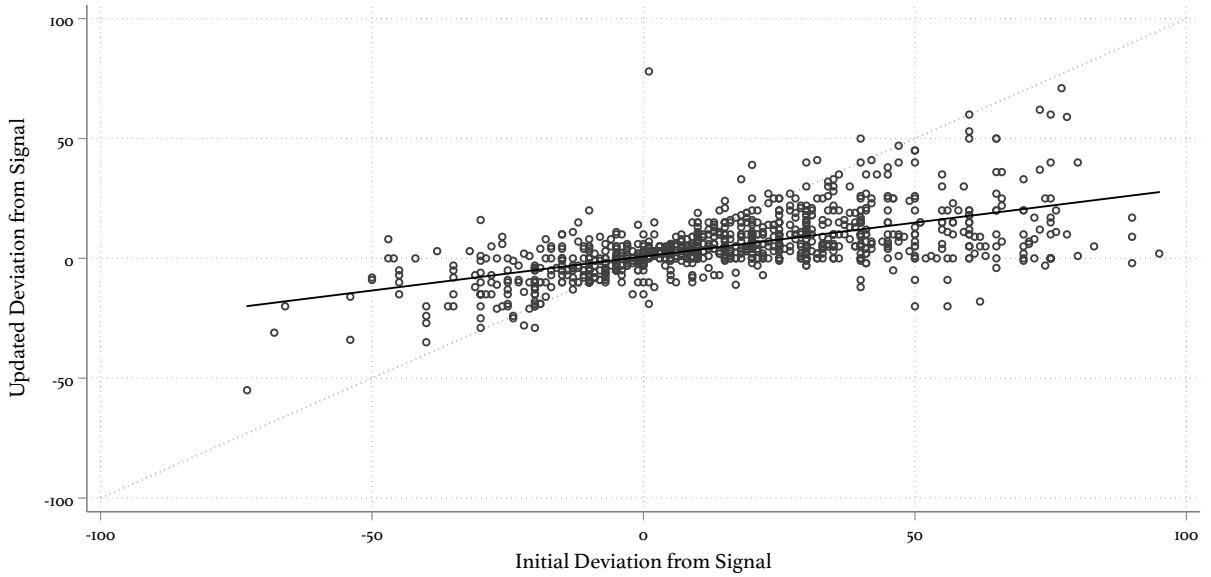
rather than the virtue or stigma they attach to the item themselves, though this would need more targeted research to confirm.

## A.2 Heterogeneous Responses to Signals

The analysis in Table 5 is limited to aggregate updating and could obscure important heterogeneity in updating behavior. Figures A.1 and A.2 add detail to explore this heterogeneity. Each figure shows all predictors’ guesses relative to the signal they received. The x-axis (y-axis) measures the difference between a predictor’s initial (updated) guess and her signal. Since a steeper slope indicates less weight given to the signal, our test of Hypothesis 1 from Table 5 amounts to testing whether the slope is flatter in Figure A.1.<sup>43</sup> These figures demonstrate more subtle responses to signals as well. A predictor who entirely ignores the signal will land on the 45-degree line, while a predictor who fully updates her prediction to match her signal will land on the x-axis. Table A.2 tests whether these behaviors—in addition to partial updating—differ across information sources.

<sup>43</sup>This holds for the region above the x-axis. Below the x-axis would indicate an *overreaction* to the signal.

**Figure A.1.** Predictors receiving signals from the IC group



**Figure A.2.** Predictors receiving signals from the H group

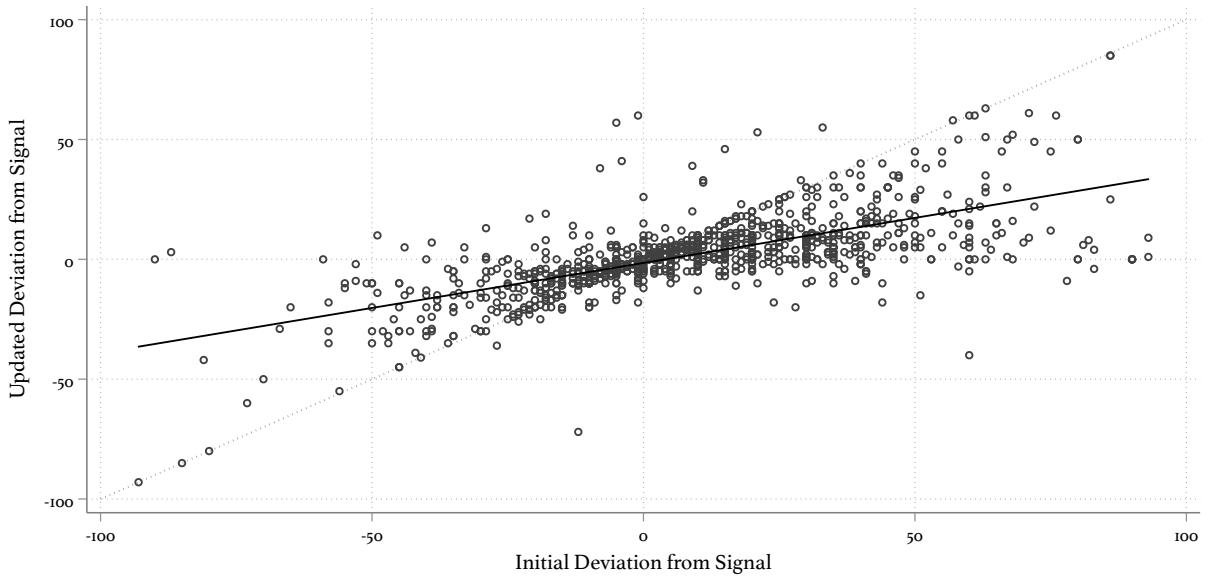


Table A.2 shows that, when a signal comes from the IC group, predictors are 2.7 percentage points less likely to completely ignore it ( $p = 0.175$ ) and 3.2 percentage points more likely to match it exactly ( $p = 0.153$ ). Column 3 shows that predictors who neither completely ignore their signal nor match their signal exactly continue to discount signals from



**Table A.2.** Updated guesses in response to signals from different sources

	Pr[Ignore Signal]	Pr[Match Signal]	$\Delta$ Guess (partial updating)
IC Info Source	-0.03 (0.02)	0.03 (0.02)	
Signal Value			0.61*** (0.05)
IC Info Source $\times$ Signal Value			0.17** (0.07)
Observations	2168	2168	1663
Clusters	271	271	268
Fixed Effects:	None	None	Action $\times$ Source

*Notes:* Columns 1-3: Random-effects linear regression with subject-level random effects and standard errors clustered at the individual level. Column 3 restricts the sample to predictors who neither ignore their signal nor match their signal exactly. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

the H group by 17 percentage points relative to the IC group ( $p = 0.023$ ).

### A.3 Experience with SDR

To examine the mechanisms driving social sophistication, we explore whether predictors who previously participated in the Choice Stage are better at accounting for SDR in poll data than newly recruited predictors. A predictor with experience in the Choice Stage may have felt the impulse to misrepresent their own preferences. This experience may then be transformed into a higher degree of skepticism about signals from the H group. As a natural extension, we also test if this experience makes predictors more accurate in their guesses.

We specifically examine if the discounting of H signals relative to IC signals differs between three types of predictors: (i) those who participated in the IC group in the Choice Stage, (ii) those who participated in the H group in the Choice Stage, and (iii) newly recruited predictors who did not participate in the Choice Stage. To test this, we adapt the approach of Hypothesis 1 to include interaction terms for each of the three groups.

Our results find no significant heterogeneity in the discounting of H signals relative to IC signals. Indeed, a fundamental level of social sophistication seems to be present in all predictors, including those who are newly recruited. However, there is some suggestive evidence that participants from the H group may give greater weight to IC signals. We find a positive point estimate of 0.12 ( $p = 0.125$ ) for the coefficient on “IC Info Source $\times$ Signal Value $\times$ H Group Member”. These predictors, having participated in the H group, may be more aware of the impulse to lie in the hypothetical Choice Stage since they themselves faced this temptation. As a result, they may increase the relative weight they put on choices from the IC group, but this is speculative.

We also find no significant differences in the accuracy of predictors’ guesses based on their experiences. The average absolute errors in first guesses are 21.54, 21.66, and 21.54 for predictors from the IC group, H group, and new recruits, respectively (joint test of equality  $p = 0.99$ ). The corresponding average absolute errors in second guesses are 12.22, 12.61, and 12.58 (joint test of equality  $p = 0.85$ ).

**Table A.3.** Updated guesses by information source across groups with different prior experience

	Updated Guess	$\Delta$ Guess
Signal Value	0.59*** (0.05)	0.56*** (0.06)
Signal Value×IC Group Member	0.05 (0.05)	0.05 (0.06)
Signal Value×H Group Member	-0.04 (0.05)	-0.07 (0.06)
IC Info Source×Signal Value	0.06 (0.05)	0.13* (0.08)
IC Info Source×Signal Value×IC Group Member	-0.02 (0.07)	0.01 (0.08)
IC Info Source×Signal Value×H Group Member	0.07 (0.07)	0.12 (0.08)
Initial Guess	0.30*** (0.02)	
IC Info Source	0.85 (1.98)	
IC Info Source×IC Group Member	-0.94 (2.54)	-0.33 (4.16)
IC Info Source×H Group Member	-4.97* (2.69)	-4.58 (4.03)
Observations		2168
Clusters		271
Control for Mean Signal:	Yes	N/A
Control for IC/H/New Group:	Yes	Yes
Fixed-Effects:	Action	Action×Source

*Notes:* Random-effects linear regression with subject-level random effects. Standard errors clustered at the individual level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

#### A.4 Direction of SDR

In Table A.4, we replicate the analysis of Section 5.2 using a fully-interacted model. This approach produces qualitatively similar results as seen in the coefficients for “Perception-Inflating×IC Info Source×Signal Value” and “Perception-Deflating×IC Info Source×Signal Value.” We again find that predictors discount perception-inflating signals from the H group. However, the incorrect discounting of perception-deflating signals from the H group are now significant in the within-subjects specification. Thus, behavior is even less consistent with social sophistication under this approach.

**Table A.4.** Updated guesses in response to perception-inflating signals from different sources

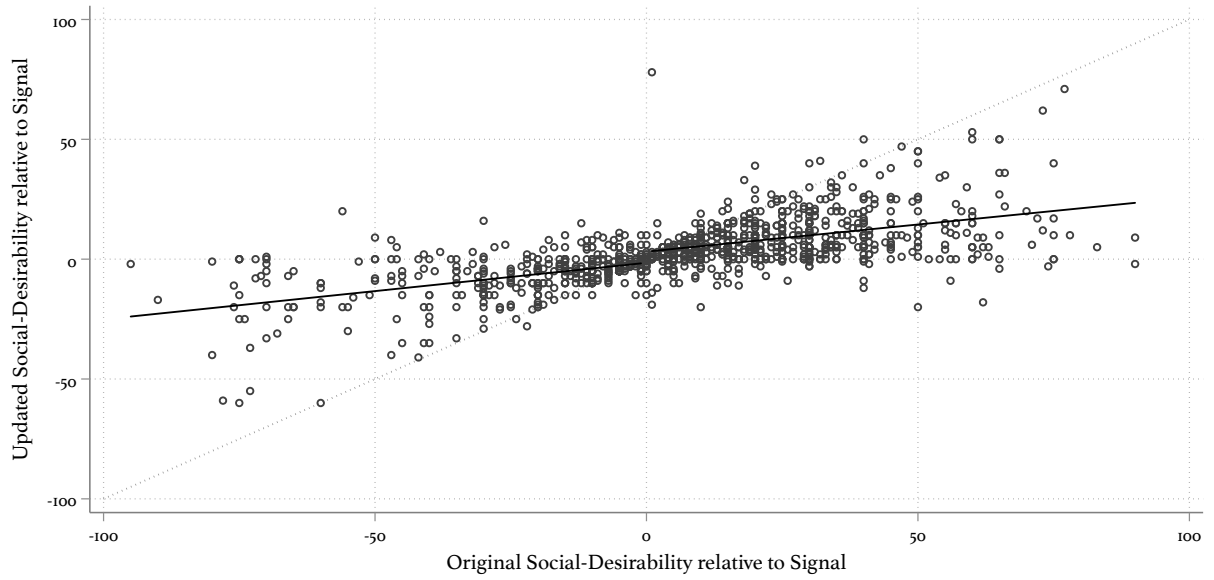
	Updated Guess	$\Delta$ Guess
Perception-Inflating $\times$ Signal Value	0.55*** (0.03)	0.55*** (0.05)
Perception-Inflating $\times$ IC Info Source $\times$ Signal Value	0.16*** (0.04)	0.29*** (0.08)
Perception-Inflating $\times$ IC Info Source	-3.14** (1.41)	-8.60** (3.46)
Perception-Inflating	1.01 (1.11)	1.71 (2.09)
Perception-Deflating $\times$ Signal Value	0.68*** (0.03)	0.41*** (0.05)
Perception-Deflating $\times$ IC Info Source $\times$ Signal Value	-0.01 (0.03)	0.15** (0.08)
Perception-Deflating $\times$ IC Info Source	0.85 (1.15)	-1.51 (2.42)
Initial Guess	0.30*** (0.02)	
Observations		2168
Clusters		271
Control for Mean Signal:	Yes	N/A
Fixed-Effects:	Action	Action $\times$ Source

*Notes:* Random-effects linear regression with subject-level random effects. Standard errors clustered at the individual level. “Perception-Inflating” (“Perception-Deflating”) are indicator variables equal to one if the signal is in the direction of more (less) social desirability relative to the initial guess. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

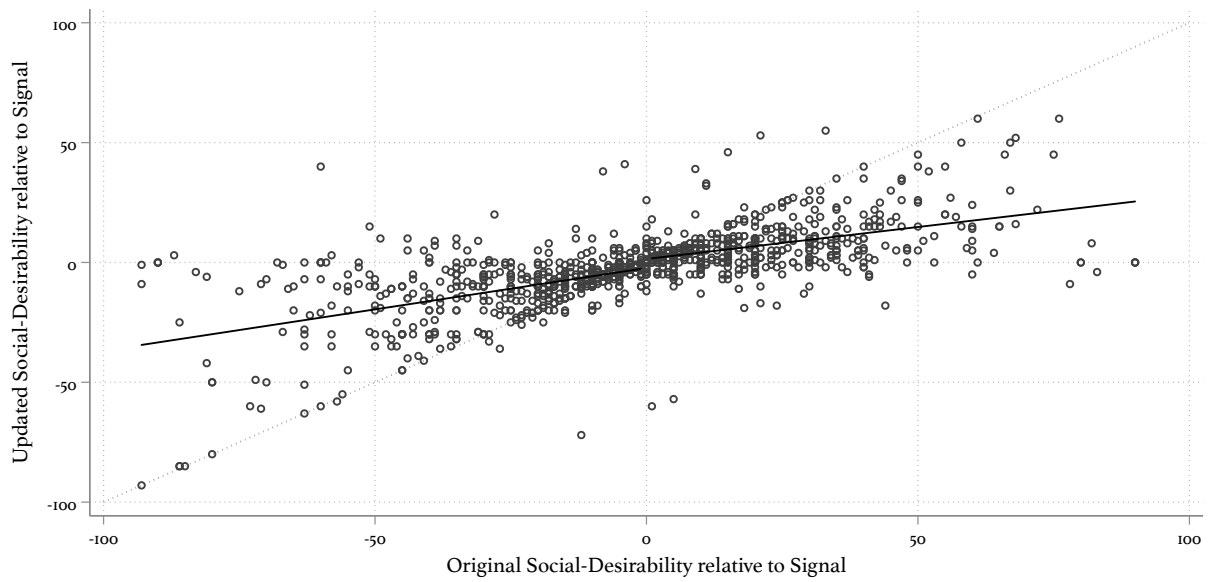
Figures A.3 and A.4 replicate the visualizations from Figures A.1 and A.2 after replacing predictions about the number of subjects taking an action with the number of subjects engaging in the socially-desirable behavior. For example, this transformation replaces predictions about the number of subjects who steal from another subject with the number of subjects who refuse to steal from another subject.

Figures A.3 and A.4 corroborate Table 6 by demonstrating a relatively similar response to information from the IC and H groups when that information suggests less socially-desirability (on the right side of the figures) and a significant discounting of information from the H group when that information suggests greater social-desirability (on the left side of the figures). This can be seen by the flatter slope on the left side of Figure A.3 than on the left side of Figure A.4.

**Figure A.3.** Predictors receiving signals from the IC group



**Figure A.4.** Predictors receiving signals from the H group



## B Appendix B: Details on Analysis

This appendix provides details about the specific regressions underlying our results in Sections 4–???. Our analysis carefully follows our pre-registration, which specifies an analysis of covariance (ANCOVA) framework. The sections below mirror the order of our results in Sections 4–???, and each indicates any changes to the analysis from the pre-registration along with any supplemental analyses that we conduct.

### B.1 Manipulation Check 1: SDR

In Column 1 of Table 3 we run the following pre-registered regression using each of the eight actions as an observation:

$$SDR_A = \beta_0 + \beta_1 \times V_A + \epsilon_A, \quad (4)$$

where  $V_A$  (defined in Equation 1) is the average sentiment for action  $A$  across all participants in the Sentiment Stage.

#### Alternative Specification: Individual-Level Sentiment

Our pre-registered analysis fails to take advantage of the full sample of subjects in the sentiment analysis. Thus, in Column 2 of Table 3, we include a supplementary analysis at the subject-level that increases statistical power without changing the underlying data. Following the standardized index defined in Equation 2, we generate  $\widehat{V}_{i,A} \equiv \frac{V_{i,A} - \bar{V}_i}{\sigma_i}$  and include it on the right-hand side of the random-effects linear regression:

$$SDR_A = \beta_0 + \beta_1 \times \widehat{V}_{i,A} + \nu_i + \epsilon_{i,A}. \quad (5)$$

## B.2 Manipulation Check 2: Accuracy

In Columns 1 and 3 of Table 4, we run pre-registered random-effects linear regressions to test for the impact of the signal source on the accuracy of the guesses:

$$ABS_{i,2,A} = \beta_0 + \beta_1 ABS_{i,1,A} + \beta_2 IC_i + \delta_A + \nu_i + \epsilon_{i,A}, \quad (6)$$

$$SQ_{i,2,A} = \beta_0 + \beta_1 SQ_{i,1,A} + \beta_2 IC_i + \delta_A + \nu_i + \epsilon_{i,A}, \quad (7)$$

where  $IC_i$  is an indicator variable equal to one if subject  $i$  received a signal from the IC group, and  $\nu_i$  are subject random-effects (meaning they will not be individually identified). Standard errors are clustered at the individual level.

### Alternative Specification: Individual Changes in Accuracy

Our pre-registered analysis takes the form of an analysis of covariance (ANCOVA). In Columns 2 and 4 of Table 4, we look at individual-level changes in accuracy to gain statistical power without changing the underlying data:  $\Delta ABS_{i,A} = ABS_{i,2,A} - ABS_{i,1,A}$  and  $\Delta SQ_{i,A} = SQ_{i,2,A} - SQ_{i,1,A}$ . This is equivalent to restricting  $\beta_1 = 1$  in our original equation. We repeat the random-effects linear regression with the new dependent variable:

$$\Delta ABS_{i,A} = \beta_0 + \beta_1 IC_i + \delta_A + \nu_i + \epsilon_{i,A}, \quad (8)$$

$$\Delta SQ_{i,A} = \beta_0 + \beta_1 IC_i + \delta_A + \nu_i + \epsilon_{i,A}. \quad (9)$$

## B.3 Hypothesis 1: Anticipation of SDR

In Column 1 of Table 5, we run the pre-registered random-effects linear regression:

$$GUESS_{i,2,A} = \beta_0 + \beta_1 GUESS_{i,1,A} + \beta_2 S_{i,A} + \beta_3 S_{i,A} \times IC_i + \beta_4 IC_i + \beta_5 \bar{S}_{T,A} + \delta_A + \nu_i + \epsilon_{i,A}, \quad (10)$$

where  $S_{i,A}$  is the signal received by subject  $i$  for action  $A$  (i.e., the fraction of subjects from  $i$ 's random sample of 10 who took action  $A$ ), and  $\bar{S}_{T,A}$  is the mean of the distribution of signals from group  $T$  (either IC or H) for action  $A$ . By controlling for  $\bar{S}_{T,A}$ , we are able to use  $S_{i,A}$  to identify the effect of a change in the signal that is derived only from sampling variation—that is, the mechanically-random change in the signal.  $\delta_A$  are fixed-effects for each action. Again,  $\nu_i$  are subject random-effects, and we cluster standard errors at the individual level.

### Alternative Specification: Individual Changes

Alongside our pre-registered analysis, in Column 2 of Table 5, we include a higher-powered test of individual-level updating:  $\Delta\text{GUESS}_{i,A} = \text{GUESS}_{i,2,A} - \text{GUESS}_{i,1,A}$ . We also modified the specification to use fixed effects for all 16 combinations of actions and choice groups,  $\delta_{T,A}$ , rather than fixed-effects for actions and controls for signal means. Our alternate specification is:

$$\Delta\text{GUESS}_{i,A} = \beta_0 + \beta_1 S_{i,A} + \beta_3 S_{i,A} \times \text{IC}_i + \beta_4 \text{IC}_i + \delta_{T,A} + \nu_i + \epsilon_{i,A}. \quad (11)$$

### Alternative Specification: Extensive- and Intensive-Margin Responses

Table A.2 provides an entirely new analysis of responses to signals. Column 1 estimates the probability of no response to the signal, Column 2 estimates the probability of exactly matching (i.e. perfectly responding to) the signal, and Column 3 explores intermediate responses where the predictor neither ignores nor matches the signal. The three estimating equations are included in sequence below:

$$\Pr(\text{MATCH}_{i,A}) = \Phi(\beta_0 + \beta_1 \text{IC}_i + \delta_A + \nu_i + \epsilon_{i,A}), \quad (12)$$

$$\Pr(\text{IGNORE}_{i,A}) = \Phi(\beta_0 + \beta_1 \text{IC}_i + \delta_A + \nu_i + \epsilon_{i,A}), \quad (13)$$

$$\Delta\text{GUESS}_{i,A} = \beta_0 + \beta_1 S_{i,A} + \beta_3 S_{i,A} \times \text{IC}_i + \beta_4 \text{IC}_i + \delta_{T,A} + \nu_i + \epsilon_{i,A}, \quad (14)$$



where  $\text{MATCH}_{i,A}$  and  $\text{IGNORE}_{i,A}$  are indicators for  $\text{GUESS}_{2,A} = S_{i,A}$  and  $\text{GUESS}_{2,A} = \text{GUESS}_{1,A}$ , respectively. The third equation is estimated on a selected sample of guesses that excludes any where  $\text{MATCH}_{i,A} = 1$  or  $\text{IGNORE}_{i,A} = 1$ .

#### B.4 Hypothesis 2: Direction of SDR

Hypothesis 2 was not included in our pre-registration. All results can be found in Table 6.

To test this hypothesis, we must divide our sample based on whether or not the signal is perception-inflating—that is, if it suggests that the observed behavior is more or less socially desirable than the predictor’s initial guess. The direction of social desirability will be determined based on the relative selection rates for the full sample. An action is socially desirable if  $\text{SDR}_A > 0$ ; thus, a signal is perception-inflating if it suggests that there are more people engaging in (or claiming to engage in) this action than the predictor initially guessed. The opposite is true for actions that are socially undesirable (i.e.  $\text{SDR}_A < 0$ ).

Column 1 of Table 6 presents our first test of Hypothesis 2 using the same random-effects linear-regression specification as in Equation 10, but including a full set of interactions with terms that indicate whether the signal is perception-inflating or perception-deflating:

$$\begin{aligned} \text{GUESS}_{i,2,A} = & \beta_0 + \beta_1 \text{GUESS}_{i,1,A} + \beta_2 S_{i,A} \times \text{PI}_{i,A} \\ & + \beta_3 S_{i,A} \times \text{IC}_i \times \text{PI}_{i,A} + \beta_4 \text{IC}_i \times \text{PI}_{i,A} + \beta_5 \text{PI}_{i,A} + \beta_6 S_{i,A} \times \text{PD}_{i,A} \\ & + \beta_7 S_{i,A} \times \text{IC}_i \times \text{PD}_{i,A} + \beta_8 \text{IC}_i \times \text{PD}_{i,A} + \beta_9 \bar{S}_{T,A} + \delta_A + \nu_i + \epsilon_{i,A}. \end{aligned} \quad (15)$$

Here, we interact all of the relevant terms from Equation 10 with  $\text{PI}_{i,A}$  ( $\text{PD}_{i,A}$ ), indicators for whether the signal is perception-inflating (perception-deflating) relative to  $\text{GUESS}_{i,1,A}$ . We test two aspects of updating: (1) if signals from the IC group are weighted more heavily (relative to signals from the H group) as they indicate greater image inflation (i.e. if  $\beta_3 > 0$ ) and (2) if signals from the H group are weighted more heavily (relative to signals from the IC group) as they indicate image deflation (i.e. if  $\beta_7 < 0$ ).

### Alternative Specification: Individual Changes

Similar to Equation 11, we include our measure of individual-level updating,  $\Delta\text{GUESS}_{i,A}$ , and our fixed-effects for combinations of action and group,  $\delta_{T,A}$ , in place of  $\bar{S}_{T,A}$  and  $\delta_A$ . This analysis is presented in Column 2 of Table 6.

### B.5 Hypothesis 3: Relative Magnitude of SDR

Column 1 of Table 7 conducts our pre-registered test of Hypothesis 3 using the same random-effects linear-regression specification as in Equation 10. However, we now include terms interacted with the absolute value of our measure of SDR:

$$\begin{aligned} \text{GUESS}_{i,2,A} = & \beta_0 + \beta_1\text{GUESS}_{i,1,A} + \beta_2S_{i,A} + \beta_3S_{i,A} \times \text{IC}_i + \beta_4\text{IC}_i + \beta_5S_{i,A} \times |\text{SDR}_A| \\ & + \beta_6S_{i,A} \times \text{IC}_i \times |\text{SDR}_A| + \beta_7\text{IC}_i \times |\text{SDR}_A| + \beta_8|\text{SDR}_A| + \beta_9\bar{S}_{T,A} + \delta_A + \nu_i + \epsilon_{i,A}. \end{aligned} \quad (16)$$

Here, we interact all of the relevant terms from Equation 10 with the absolute value of our measure of SDR for action  $A$ ,  $|\text{SDR}_A|$ . We test if signals from the IC group are weighted more heavily (relative to signals from the H group) as SDR becomes more extreme (i.e. if  $\beta_6 > 0$ ).

### Alternative Specification: Individual Changes and Sensitivity to Sentiment

Similar to Equation 11, we include our measure of individual-level updating,  $\Delta\text{GUESS}_{i,A}$ , and our fixed-effects for combinations of action and group,  $\delta_{T,A}$ , in place of  $\bar{S}_{T,A}$  and  $\delta_A$ . This analysis is presented in Column 2 of Table 7

We also measure how sensitive subjects are to changes in our proxy for social desirability, sentiment. Specifically, we replace  $|\text{SDR}_A|$  with a standardized measure of how extreme sentiment is toward the action,  $|\hat{V}_A| = \frac{|V_A - \bar{V}|}{\sigma_V}$ , where  $\bar{V}$  and  $\sigma_V$  are the mean and standard

deviation of  $V_A$  across all eight actions. This analysis is presented in Column 3 of Table 7.

## B.6 Confidence

Our exploratory analysis on confidence was not pre-registered but adds substantively to our understanding of the implications of poor data-quality on behaviors surrounding inference (in this case, confidence in guesses). Table 8 presents two tests of the impact of a predictor’s accuracy on their confidence. Prior to running this analysis, we normalize confidence measures across all predictors and all actions to generate  $\widehat{\text{CONFIDENCE}}_{i,1,A}$  and  $\widehat{\text{CONFIDENCE}}_{i,2,A}$ , both of which have mean 0 and standard deviation 1. Column 1 presents the association between normalized confidence and the accuracy of initial guesses using the following specification:

$$\widehat{\text{CONFIDENCE}}_{i,1,A} = \beta_0 + \beta_1 \text{ABS}_{i,1,A} + \delta_A + \nu_i + \epsilon_{i,A}, \quad (17)$$

where  $\text{ABS}_{i,1,A}$  is the absolute error in subject  $i$ ’s initial guess,  $\delta_A$  is a vector of action fixed effects, and  $\nu_i$  are subject random-effects. Standard errors are clustered at the individual level.

Column 2 of Table 8 demonstrates how this confidence evolves after receiving information. It uses the following specification:

$$\begin{aligned} \widehat{\text{CONFIDENCE}}_{i,2,A} = & \beta_0 + \beta_1 \text{ABS}_{i,1,A} + \beta_2 \text{ABS}_{i,2,A} + \beta_3 \text{ABS}_{i,2,A} \times \text{IC}_i \\ & + \beta_4 \widehat{\text{CONFIDENCE}}_{i,1,A} + \beta_5 \text{IC}_i + \delta_A + \nu_i + \epsilon_{i,A}, \end{aligned} \quad (18)$$

where  $\text{ABS}_{i,2,A}$  is the absolute error in subject  $i$ ’s updated guess. Standard errors are again clustered at the individual level.

In both tests, we consider how confidence is associated with accuracy ( $\beta_1$  in Equation 17 and  $\beta_2$  in Equation 18). In Equation 18, we also care about how this depends on the randomly-assigned information source ( $\beta_3$ ).

## B.7 Experience with SDR

Column 1 of Table A.3 presents the pre-registered test of our hypothesis about experience. We use the same random-effects linear-regression specification as in Equation 10, but include terms interacted with the role that Predictor  $i$  played in the Choice Stage:

$$\begin{aligned} \text{GUESS}_{i,2,A} = & \beta_0 + \beta_1 \text{GUESS}_{i,1,A} + \beta_2 S_{i,A} + \beta_3 S_{i,A} \times \text{IC}_i + \beta_4 \text{IC}_i + \beta_5 S_{i,A} \times \text{EXP}_i \\ & + \beta_6 S_{i,A} \times \text{IC}_i \times \text{EXP}_i + \beta_7 \text{IC}_i \times \text{EXP}_i + \beta_8 \text{EXP}_i + \beta_9 \bar{S}_{T,A} + \delta_A + \nu_i + \epsilon_{i,A}, \end{aligned} \quad (19)$$

where  $\text{EXP}_i$  is an indicator variable equal to one if the predictor has previous experience participating in the IC or H group. Again, we test for a significant interaction effect by testing if  $\beta_6 > 0$ .

We repeat this analysis looking at members of the H and IC groups separately, which reveals heterogeneity in the learned experience of the two groups.

### **Alternative Specification: Individual Changes**

As with Hypotheses 1–3, we replicate the pre-registered analysis with an alternative specification. As before, we include individual-level updating and fixed-effects for combinations of action and group. This analysis is presented in Column 2 of Table A.3.

## **C Appendix C: Experimental Instructions**

### **C.1 Sentiment-Stage Instructions**

Thank you for your participation today. Just for participating in this study, you will receive \$5 toward your Take-Home Pay. In order to receive your Take-Home Pay, you need to complete the entire survey and then instructions for payment will be emailed to you once all responses have been collected.

All of the choices will be made in private. This means that your responses will be observed by the researchers after-the-fact and no one else.

This is a non-deceptive experiment. That means that, if we say an action has real consequences, those consequences will actually happen. On the other hand, if a choice is hypothetical, we will tell you in advance that it is hypothetical.

#### **C.1.1 Sentiment-Stage Comprehension Question**

We will be asking you to respond to questions about a series of potential scenarios. Your responses will not have any real consequences, we are simply asking for your feelings on each scenario.

To ensure that you understand, please answer the following question. Will your choices have real consequences?

- Yes, all of them will counts.
- Yes, on will be chosen at randomly-chosen.
- No, you are just asking my opinion.

### Figure C.1. Sentiment-Stage Decision Screen

Here, we would like for you to tell us how you feel about Donating \$1 to St. Jude Children's Hospital.

Specifically, consider the following:

You can privately donate \$1 to St. Jude Children's Hospital. St. Jude is a pediatric treatment and research facility focused on children's catastrophic diseases, particularly leukemia and other cancers. If you choose to donate, St. Jude will receive \$1 and it will cost you \$1 of your payment. Nobody besides the researchers will know if you donated.

---

How would you feel about taking this action yourself?

Very Negative 0 1 2 3 4 5 6 7 8 9 10 Very Positive

Feelings



How would you feel about other people who take this action?

Very Negative 0 1 2 3 4 5 6 7 8 9 10 Very Positive

Feelings



How do you think most other people would feel about people who take this action?

Very Negative 0 1 2 3 4 5 6 7 8 9 10 Very Positive

Feelings



## **C.2 Choice-Stage Instructions: Hypothetical Group**

Thank you for your participation today. Just for participating in this part of the experiment, you will receive \$5 toward your Take-Home Pay. In order to receive your Take-Home Pay, you must complete the second part of the experiment that we will email to you after you complete this. The second part of the experiment will pay you between \$5 and \$10. So, you will receive between \$10 and \$15 for completing both parts of the study.

All of the choices will be made in private. This means that your choice will be observed by the researchers after-the-fact and no one else.

This is a non-deceptive experiment. That means that, if we say an action has real consequences, those consequences will actually happen. On the other hand, if a choice is hypothetical, we will tell you in advance that it is hypothetical.

### **C.2.1 Choice-Stage Comprehension Question: Hypothetical**

We will be asking you to make a series of choices and answer a few questions. All of your choices will be hypothetical. Meaning that none of your choices will have real consequences.

We simply want to know how you would respond if you were asked to make a choice in these hypothetical situations.

To ensure that you understand, please answer the following question. Will your choices have real consequences?

- Yes, one randomly selected choice will count
- Yes, all of them will count.
- No, they are hypothetical.

## **C.3 Choice-Stage Instructions: Incentive Compatible Group**

Thank you for your participation today. Just for participating in this part of the experiment, you will receive \$5 toward your Take-Home Pay. In order to receive your Take-Home Pay,

you must complete the second part of the experiment that we will email to you after you complete this. The second part of the experiment will pay you between \$5 and \$10. So, you will receive between \$10 and \$15 for completing both parts of the study.

All of the choices will be made in private. This means that your choice will be observed by the researchers after-the-fact and no one else.

This is a non-deceptive experiment. That means that, if we say an action has real consequences, those consequences will actually happen. On the other hand, if a choice is hypothetical, we will tell you in advance that it is hypothetical.

### **C.3.1 Choice-Stage Comprehension Question: Incentive-Compatible**

We will be asking you to make a series of choices and answer a few questions. Your choices will have real consequences.

At the end of the study, we will randomly select one of your choices to be the Choice That Counts. The Choice That Counts will determine your outcome today. Since any choice can be selected as the Choice That Counts, you should treat every choice like it is the Choice That Counts.

To reiterate, only one of your choices will be randomly chosen as the Choice That Counts. So, treat each choice as a separate, meaningful choice.

To ensure that you understand, please answer the following question. Will your choices have real consequences?

- Yes, on randomly selected choice will count
- Yes, all of them will count.
- No, they are hypothetical.



## C.4 Prediction-Stage Instructions

### C.4.1 Prediction-Stage General Instructions

Just for participating, you will be guaranteed to receive \$5. You may earn significantly more money depending on how you perform your tasks in this study.

In this study, you are a "Predictor." Your task today will be to make predictions about the behavior of other participants in the study. The more accurate your predictions are, the more money you will earn.

We recruited students at the University of Arkansas to be "Real-Deciders." Real-Deciders made a series of private choices and entered them confidentially into a computer.

The Real-Deciders knew that their choices would never be individually observed by anyone but the researchers.

The choices that the Real-Deciders made had real consequences. One choice made by each Real-Decoder was randomly selected to be carried out by the experimenters.

Key Points: Real-Deciders made private decisions without anyone watching. Their decisions had real consequences and really determined their payment.

### Figure C.2. Hypothetical Decision

Suppose you could privately donate \$1 to St. Jude Children's Hospital. St. Jude is a pediatric treatment and research facility focused on children's catastrophic diseases, particularly leukemia and other cancers. If you chose to donate, St. Jude would receive \$1 and it would cost you \$1 of your payment. Nobody besides the researchers would know if you donated.



I would DONATE

I would NOT DONATE

### Figure C.3. IC Decision

You can privately donate \$1 to St. Jude Children's Hospital. St. Jude is a pediatric treatment and research facility focused on children's catastrophic diseases, particularly leukemia and other cancers. If you choose to donate, St. Jude will receive \$1 and it will cost you \$1 of your payment. Nobody besides the researchers will know if you donated.

I will DONATE

I will NOT DONATE

#### C.4.2 Prediction-Stage Comprehension Question

What is your role in this study?

- Make decisions
- Guess what decisions the Real-Deciders made
- Help the Real-Deciders make their decisions

Did the Real-Deciders' choices have consequences?

- Yes, their choices mattered
- No, their choices were hypothetical

#### C.4.3 Prediction-Stage Predictions Instructions

The Real-Deciders made decisions about several different actions. We described these actions to the Real-Deciders before they made their choices. We will describe them to you in exactly the same way.

For each action, there were only two options: Option 1: Take the action Option 2: Do not take the action

Your job is to predict **P** – the number of the Real-Deciders out of 100 who chose to take the action (the first option). You will report your best guess about **P**.

There is a true percentage of Real-Deciders who chose to take each action. We'll call this value "**True-P**". The closer you get to guessing the **True-P**, the more money you can earn.

It is important that you think carefully about your prediction for **P** because we will offer you a chance to win money based on your accuracy.

You will make 16 predictions in this study. We will randomly select one of these predictions to be the Prediction That Counts. Your money will depend on how accurate you are on the Prediction That Counts. Since each prediction could be the Prediction That Counts, you should treat each prediction like it is the Prediction That Counts.

#### C.4.4 Payment Comprehension Question

How many of your 16 predictions will determine your payment?

- All of them collectively
- One selected at random: "the Prediction That Counts"
- The first one
- The last one

#### C.4.5 Prediction-Stage Lottery Draw Instructions

You will have a chance to earn an extra \$5 lottery bonus at the end of the study (in addition to the \$5 you are already guaranteed). You will earn lottery tickets if your guess about **P** is close to the **True-P**. At the end of the session, we will randomly draw a lottery number between 1 and 100; if that number matches one of your lottery tickets, you will win the bonus payment. So it's best to get as many lottery tickets as possible to maximize your chance of a bonus.

On the next page, we will describe how you can earn tickets based on your guess of **P**. The precise method we use to calculate your lottery tickets may sound complicated, but you will always earn the most if you simply answer truthfully.

#### C.4.6 Prediction-Stage Lottery Draw Comprehension Questions

What is the easiest way to earn the most lottery tickets?

- Guess the largest number as the **True-P**
- Guess the smallest number as the **True-P**
- Guess your honest beliefs about the **True-P**

#### C.4.7 Prediction-Stage Lottery Ticket Instructions

The number of lottery tickets you will receive will be one of the following: *Option A*: The number of lottery tickets you will receive is equal to the **True-P**. *Option B*: The number of lottery tickets you will receive is equal to your "Random Draw," which is a random number between 0 and 100.

The option you receive depends on how your Random Draw compares to your guess about **P**. If your Random Draw is below your guess, then you will get Option A (lottery tickets equal to the **True-P**). If your Random Draw is above your guess, then you will get Option B (lottery tickets equal to your Random Draw).

Here are two examples:

If your guess is that **P=50**, and your Random Draw is 25, then your Random Draw is less than your guess about the **True-P**. So, you will get Option A (lottery tickets equal to the **True-P**).

If your guess is that **P=50**, and your Random Draw is 75, then your Random Draw is more than your guess about the **True-P**. So, you will get Option B (lottery tickets equal to your Random Draw).

### C.4.8 Prediction-Stage Lottery Ticket Comprehension Questions

If your guess about  $P$  is that  $P=23$  and your Random Draw is 17, how many lottery tickets will you receive?

- 50
- Option A: you will receive a number of lottery tickets equal to the **True-P**
- Option B: you will receive a number of lottery tickets equal to your Random Draw, 17.

If your guess about  $P$  is that  $P=43$  and your Random Draw is 73, how many lottery tickets will you receive?

- 50
- Option A: you will receive a number of lottery tickets equal to the **True-P**
- Option B: you will receive a number of lottery tickets equal to your Random Draw, 73.

You might think you can “game the system” and earn more lottery tickets by reporting a higher guess for  $P$  than you really believe. That won’t help you. It will only increase the chance that you pass up your Random Draw when it is a high number.

On the other hand, you also can’t game the system by reporting a lower guess for  $P$  than you really believe. If you do that, then you will increase the chance that you accept your Random Draw when it is a low number.

**Figure C.4.** First-Prediction Choice

Real-Deciders chose between:

- Option A: **Pay \$1** to Donate \$1 to St. Jude Children's Hospital.
- Option B: Do not donate \$1.

How many of the 100 Real-Deciders do you think chose to donate?

0    10    20    30    40    50    60    70    80    90    100

Real-Deciders



**Figure C.5.** First-Prediction Choice Confidence

How confident are you in your prediction?

Very Uncertain    0    1    2    3    4    5    6    7    8    9    10    Very Confident

Confidence:



#### C.4.9 2nd Prediction Instructions (Hypothetical Information)

Your task is to predict the behavior of the 100 Real-Deciders that we recruited from the University of Arkansas to participate in the study. Before you make these predictions for a second time, we will show you the decisions of 10 "Hypothetical-Deciders."

We recruited 100 Hypothetical-Deciders at the same time that we recruited the 100 Real-Deciders for the study. Both were recruited out of the same subject pool at the University of Arkansas.

For every one of the decisions that the Real-Deciders made, the Hypothetical-Deciders reported what they would have chosen if they had been asked to choose. But, the statements

made by Hypothetical-Deciders did not have any real consequences.

If a Hypothetical-Decider reported that they would take an action, the Hypothetical-Deciders never actually had to take the action. These responses were entirely hypothetical.

We have randomly selected 10 of the 100 Hypothetical-Deciders. We will show you their responses on all 8 actions.

The Hypothetical-Deciders did not make the exact same choices as the Real-Deciders. But this information may be useful in revising your predictions about the choices that the 100 Real-Deciders made.

While you are revising your predictions about the Real-Deciders, we will remind you of the responses of the Hypothetical-Deciders. So, you do not need to memorize their choices now.

#### **C.4.10 2nd Prediction Comprehension Question (Hypothetical Information)**

Did the Hypothetical-Deciders make choices with actual consequences?

- Yes, their choices mattered
- No, their choices were hypothetical

#### **C.4.11 2nd Prediction Instructions (IC Information)**

Your task is to predict the behavior of the 100 Real-Deciders that we recruited from the University of Arkansas to participate in the study. Before you make these predictions for a second time, we will show you the decisions of 10 of the Real-Deciders.

These 10 Real-Deciders were randomly selected from among the 100 Real-Deciders you are making predictions about. They were all recruited from the same subject pool at the University of Arkansas.

Recall that all choices made by the Real-Deciders had real consequences.

We have randomly selected 10 of the 100 Real-Deciders. We will show you their choices on all 8 actions.

The 10 randomly chosen Real-Deciders that we will show you did not make the exact same choices as the other 90 Real-Deciders. But this information may be useful in revising your predictions about the choices that all 100 Real-Deciders made.

While you are revising your predictions about the 100 Real-Deciders, we will remind you of the responses of the 10 randomly chosen Real-Deciders. So, you do not need to memorize their choices now.

#### C.4.12 2nd Prediction Comprehension Question (IC Information)

Did the 10 randomly selected Real-Deciders make choices with actual consequences?

- Yes, their choices mattered
- No, their choices were hypothetical

**Figure C.6.** Second Prediction Choice

Real-Deciders chose between:

- Option A: **Pay \$1** to Donate \$1 to St. Jude Children's Hospital.
- Option B: Do not donate \$1.

Recall that you can change your predictions however you like.

- Your original prediction was that 53 Real-Deciders chose to donate.
- 70% of the 10 Hypothetical-Deciders said they would donate \$1.

How many of the 100 Real-Deciders do you think chose to donate?

0    10    20    30    40    50    60    70    80    90    100  
Real-Deciders

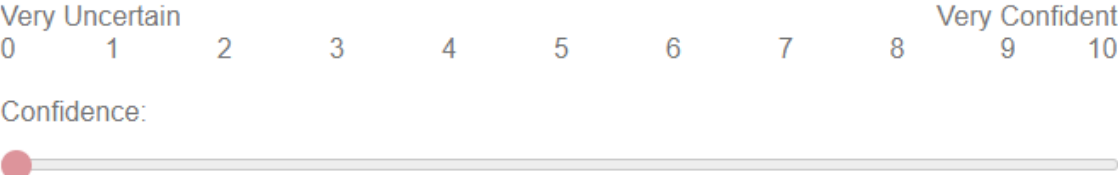




Figure C.7. Second Prediction Choice Confidence

How confident are you in your prediction?

- Your original confidence level was: 6.



## Chapter 2: Group Identity and Opportunity Cost

Nathaniel Burke<sup>1</sup> Sherry Li<sup>2</sup>

### Abstract

People place real value on the identities that they hold and membership in groups of people that hold the same or similar identities. While not easily quantifiable, individuals consciously and subconsciously take actions to signal membership and loyalty to specific identity groups over others. At times, the required signals may come at a personal cost that prevents opportunities for improved individual outcomes such as deciding between staying in a small community that you grew up in and moving away to attend a high ranked university or for job opportunities. We examine group identity loyalty in an experimental setting using parallel public goods games with two teams of homogeneous gender identities. We give participants opportunities to switch teams, leaving their identity group, in order to increase their earnings. There is some loyalty to identity groups in early stages of the game, particularly amongst women. We find that the biggest positive determinant of group-oriented contribution behavior is communication and identity priming, respectively.

**JEL classification:** D91, C92, D71.

**Keywords:** Group Identity, Public Goods Game, Gender Bias, Opportunity Cost.

---

<sup>1</sup>West Virginia University: nathaniel.burke@mail.wvu.edu

<sup>2</sup>University of Arkansas: sli@walton.uark.edu

## 1 Introduction and Background

Group identity has become more prominent in behavioral economics as a key determinant of individual behavior and decision making (Akerlof and Kranton, 2000; Chen and Li, 2009). In educational settings, group identity has been suggested to have impacts on individual utility functions where the respective individuals consider how they signal membership to their identity group and weigh those behaviors against individual long term utility maximizing behavior, often resulting in personal sacrifices to appease their group identity (Akerlof and Kranton, 2002; David and Fryer, 1995; Fryer and Torelli, 2010). This project will test the opportunity cost threshold of preexisting group identities by using a team based public goods game where participants have the ability to switch teams for higher payout with varying levels of peer pressure and knowledge that their leaving their team will adversely affect their team.

The motivating setting behind this approach is drawn from the real life trade-offs observed by high performing members of lower or limited opportunity communities. The limited opportunity nature of one of these communities does not necessarily imply a lack of skill or priority but could be a range of possibilities from a lack of opportunities due to changing markets such as West Virginia's coal communities in the modern age or perhaps an area that specializes in a particular market or trade, such as an agricultural community in Arkansas that specializes in rice or soy production. Through a combination of economic and cultural factors, individuals may find themselves strongly identifying with a certain place they grew up or the community they group up with. This is also found to be true in many areas with cohesive minority and/or immigrant communities. A commonly understood dilemma is the story of a high achieving community member applying to universities. They have to negotiate the internal trade-off between going to a high ranking university or perhaps one that just offers something that is not available in their local area to develop their personal human capital and staying in their community for reasons including identity loyalty, not contributing to a local brain drain, etc. This decision carries risk if they do decide to leave

their community or even just apply to leave their community. In this context there are two points where there is a probability of failure which carries the weight of failing on personal goals and negatively signaling identity membership. The first is when the student applies to the given university and has to wait on an admissions decision and the second is when the student actually attends the school and has to maintain an academic standard and manage to integrate into a new group identity. In our experimental setting, these two points of potential failure are collapsed into a single decision point.

This motivating example is of a specific decision for a student leaving their community and identity group to engage in higher level human capital accumulation opportunities. Though this is the original motivation for the design, this type of decision making process is repeated across many different scenarios, not all requiring a large geographic move. Another scenario to consider is the team member in a multi-department organization that wants to switch to another team or department whether it is to increase pay, benefits, or for personal utility. In work environments, there are often subgroup identities such as gaining utility from being a member of the sales team or working in research and development. Even if financial compensation across opposing teams is equal, individual utility from the type of work or preferences over work environment or conditions may make one more attractive to the individual decision maker than the other while being at odds with their group signaling preferences.

In this paper, we use an experimental approach using a virtual experiment. Participants were put into two parallel, gender-homogeneous teams to participate in standard public goods games with the ability to see the opposite team's contributions and payouts and the opportunity to switch teams, subject to being approved by a vote, every three rounds. We find that women are more likely to stay loyal to their identity group than men by way of not trying to switch teams as often as men when there is more money to be made by switching. We also find that there is significant impact on contribution decisions based on how well participants can coordinate. Regardless of gender, participants increase their team

contributions when they have the ability to chat with each other. We also find that people are more likely to make their switching decisions in the first half of the experiment, with attempted and successful switches dropping off after the third opportunity to switch out of the scheduled six.

The rest of this paper follows in this fashion: Section 2 literature looking at previous work, Section 3 going over the experimental design, Section 4 presenting and discussing our estimation results, and Section 5 going over our conclusions and directions for future work.

## **2 Literature Review**

Much of the work that economics has done regarding group identity has been predicated by a wealth of social psychology literature. While economics does not make as significant of a distinction between identity theory and social identity theory, social psychologists have been working between these parallel identities since the 1970s. While social identity theory focuses more on how the individual behaves as part of a group in group processes and intergroup behavior (Hogg and Abrams, 1988; Hogg and Turner, 1985; St Claire and Turner, 1982; Tajfel, Turner, Austin and Worchel, 1979; Turner, Hogg, Oakes, Reicher and Wetherell, 1987). Identity theory has been more focused on the individual's role-related behaviors and reflexive sense of self. (Burke, 1980; McCall and Simmons, 1968; Turner, 1978). Stets and Burke (2000) argue that the integration of the of social identity and reflexive view of the "self" identity help to make the complete sense of self that a person identifies with and presents themselves. Previous literature had already shown that self-identity is impacted by social identity and the affects the way intention and behavior are constructed. This is especially true when it comes to group norms and behaving within a set group expectation (Terry, Hogg and White, 1999).

Early work in economics laying the foundation for identity in economic models was initiated by Akerlof and Kranton (2000), which introduced identity as a way to explain some

group conforming individual behaviors by gender, socioeconomic class, and the way tasks are distributed in the household. This work is built upon by the same authors in different field focused papers including education Akerlof and Kranton (2002), labor based organizations (Akerlof and Kranton, 2005), and the workplace dynamic between employees and supervisors (Akerlof and Kranton, 2008). The field of education has had great insights into understanding how identity impacts race based learning outcomes beyond Akerlof and Kranton's 2002 paper. One of the more notable analyses is from Austen-Smith and Fryer Jr (2005), which breaks down the social pressures on Black and Hispanic students to avoid being seen as "acting white". The "acting white" problem has been discussed in sociology (Horvat and Lewis, 2003; Tyson, Darity Jr and Castellino, 2005), anthropology (Fordham and Ogbu, 1986; Ogbu, 2004), and education circles. This follows from the social psychology theory of social identity that prescribes the idea that individuals will adjust behaviors to match group norms. In the case of the "acting white" burden, Black and Hispanic students signal group membership by avoiding certain behaviors in the school setting that may inadvertently signal them to want to be accepted by their white peers. In a follow up to their analysis paper, Fryer and Torelli (2010) does an empirical study to measure which activities were avoided to maintain popularity under this burden including high levels of continuous academic effort and certain extracurricular activities. Others have found that there is a limited amount of empirical basis to justify weighting so much explanatory power towards a negative framing of Black culture (Horvat and Lewis, 2003; Wildhagen, 2011) the underlying identity value is generally supported.

Gender identity has been studied as a determinant for behavior and habit formation. Previous literature has found that there is an aversion by women to certain roles within the household that is attributed to gender identity norms, particularly pertaining to earning potential relative to their husband's income (Bertrand, Kamenica and Pan, 2015). Strong sense of gender norms also impacts how women signal their attributes on the marriage market and the probability that they are found to be desirable, particularly with respect to income

and how household labor is distributed (Bertrand et al., 2015; Greenstein, 2000; Salland, 2018). These identity norms also impact happiness in married couples after the marriage market (Akerlof and Kranton, 2000; Booth and Van Ours, 2009) and the kind of behavior that households engage in based on who is the breadwinner (Ke, 2021).

Experimental work in economics has been very useful in detecting group identity impacts on public goods and production where identity cooperation is a factor. Our experimental design is motivated by the previous experiment that Charness, Cobo-Reyes and Jimenez (2014) used to examine identities in the public goods setting. Charness et al. (2014) uses a public goods game with a 2x2 design with endogenous group formation. The primary treatment dimensions vary whether participants participate in a team building exercise and whether some participants receive an endowment twice as much as other participants. Where the authors allowed for endogenous group formation alongside natural identities, in our experiment we are isolating only natural identities (men and women) and our priming reinforces those natural identities while the authors of this previous experiment are inducing subtle endogenous identities through priming. Another notable difference is that our experimental design limits the number of team members in a way that it is impossible for anybody to ever be a singleton. We also put everyone through a team building exercise before the public goods game starts to promote team behavior. The results of Charness et al. (2014) have a lot of focus on the impacts of the endogenous group identity aspect and how it positively impacts group cooperation. In particular, the word task that the authors used to induce group behavior was highly effective.

### **3 Experimental Design**

The goal of our experiment is to address two primary objectives: Evaluate participants' willingness to maintain their group identity through contributions to the group and evaluate participants' loyalty to their groups when facing opportunity costs to identity loyalty.

Our experiment consists of a 2x2 between-subject treatment factorial design in which we

change the salience of gender identity through identity prime on the one hand and the possibility of communication during the public goods game on the other (Table 1). All treatments contain two parts: a pre-game anagram challenge to enhance the gender-team identity in Part I and a 20-period public goods provision game with the possibility of switching groups in Part II.

**Table 1.** Treatment Participation

Treatments	Part I	Part II	# of Sessions
Identity: Salient Communication: Yes	Identity prime questions Anagram challenge	Team communication	2 sessions; 16 participants
Identity: Salient Communication: No	Identity prime questions Anagram challenge	Without communication	2 sessions; 16 participants
Identity: Not Salient Communication: Yes	Anagram challenge	Team communication	2 sessions; 16 participants
Identity: Not Salient Communication: No	Anagram challenge	Without communication	2 sessions; 16 participants

### 3.1 Participants

Participants were recruited from the University of Arkansas undergraduate population. Recruitment was done using a combination of Sona Systems subject pool recruitment within the College of Business and recruiting from large principle/intro level classes from the Colleges of Liberal Arts & Science, Engineering, Business, and Education. Participants were 50% female and 50% male. The experiment was conducted online using oTree (Chen, Schonger and Wickens, 2016). Participants accessed the experiment through their web browsers on their personal internet connected devices using participant-specific dedicated links. They logged into a Zoom call with their cameras on and microphones muted so they can see all other participants in their session. Participant links were given to participants upon joining the Zoom call.



### 3.2 Part 1: Anagram Game

All sessions are gender balanced with 8 participants each, 4 men and 4 women. Each session starts with participants being assigned to two gender-homogeneous teams. The all-women team is named Team A and is assigned the color purple, and the all-men team is named Team B and is assigned the color green. The colors are only used when displaying the team names "Team A" and "Team B" at the top of each page of the experiment. This is done so that the team name stands out, and the individual's current team is salient on each page. The participants are made aware of the team names and the initial team gender compositions.

In the treatments with the salient gender identity, participants are asked to enter their gender information and answer several questions that are designed to prime and make salient their gender identity.<sup>3</sup> These questions include asking participants about the gender composition of their living arrangements and preferred living arrangements. Participants continue to the Anagram challenge after answering these questions. In the treatments without the salient gender identity, participants proceed to the Anagram challenge without the gender prime questions<sup>4</sup>.

In all the treatments, every session starts with having participants engage in an incentivized pre-game anagram challenge. Each participant is given 20 anagrams to solve and they receive a bonus for each correctly solved anagram. The anagrams are all five to seven letters long with the first 10 being five letters, next five being six letters, and the last five being seven letters. All payouts and bonuses in the experiment are expressed in points and the participants are paid using a conversion of 600 points= \$1 or \$0.0017*perpoint*. Each anagram is worth a bonus of 50 points for a total potential bonus of 1000 points. During this pre-game, participants must submit their own responses for grading but have the ability to use a chat function to communicate with their teammates and help each other. Participants self-report their team sentiment—attachment to their gender-team at the end of the

---

<sup>3</sup>Appendix B

<sup>4</sup>Appendix C

anagram challenge. This is to instill a sense of team identity in the participants before moving on to the public goods game part of the experiment. All teammates have the exact same anagrams in the exact same order and all participants are aware of this in order to make coordination easier. Participants are told that they can coordinate with their teammates for assistance on successfully completing the anagrams and are encouraged to communicate with each other. The payoffs of Part I were not revealed to participants until the end of the experiment after the survey.

### 3.3 Public Goods Game with Team Switching

Part II, the public goods games, consists of 20 total periods, which are broken down into 6 segments of 3 periods and 2 rounds of end game play. In each period, participants decide how much to contribute to a public goods game within their respective teams. The main differences between our study and a standard public goods game are asymmetric endowment distributions across the two teams, the possibilities of team switching, and the dependence of MPCR on the team size, as detailed below.

#### 3.3.1 Asymmetric Distributions of Endowment

Teams A and B are each randomly assigned to a uniform distribution of endowment,  $E_i \sim Uniform(a_g, b_g)$ , where  $a$  and  $b$  indicate the upper and lower bounds of the endowment range, and  $g$  indicates the team, with one distribution being better than the other.<sup>5</sup> At the beginning of each period, each team member receives a random draw from the predetermined team endowment range of their respective uniform distribution. The two distribution ranges overlap by  $\frac{1}{4}$ , i.e., there is a 75% chance that a member from the “wealthier” team receives a higher endowment than a member from the “poorer” team.

The endowment range stays fixed for each team throughout the experiment once it is randomly assigned at the beginning of the public goods game. However, all the participants

---

<sup>5</sup>Whether  $\{a_1, b_1\}$  or  $\{a_2, b_2\}$  is a better endowment range is predetermined by a pre-randomized schedule at the session level.

receive a fresh endowment draw from their respective range in each period and decide how much to keep in their private accounts and how much,  $c_{ig}$ , to contribute to their respective team account. The total contributions to their team account,  $\sum_i c_{ig}$ , is multiplied by the team's MPCR  $M_g$  and shared equally by all team members.  $M_g$  is positively correlated with team size as presented in Table 2. Individual's payoff per period is  $\pi_{ig} = E_{ig} - c_{ig} + (M_g \sum_i c_{ig})/T_g$ , where  $T_g$  is the number of people on Team  $g$ . At the end of each period, participants receive a reminder on their individual endowment, contribution, the number of their team members, and the associated MPCR. They also receive feedback information on their team's contributions, the portion of their earnings from their team's public good account, and their individual total earnings in that period. In addition, they are also given information on the other team, including their endowment range, total contributions to their public account, the number of team members, the associated MPCR, and each team member's portion of earnings from their team's public account (see the screenshot in the Appendix). The feedback information on the other team is provided to facilitate one's decision making in possible team switching.

### 3.3.2 Teams and Switching

As mentioned above, Teams A and B start as an all-men and all-women team, respectively, with one team randomly receiving a better range of endowment distribution than the other. After every three rounds of public goods contribution decisions, participants will have an opportunity to stay or switch teams. For example, after playing the public goods game in periods 1-3, participants will be given a chance to switch teams at the end of the third period. With the new (or original) teams if some participants (don't) successfully switch team, participants will continue to play the public goods game in periods 4-6, and will be given another chance to switch teams at the end of the sixth period. This cycle repeats with team switching and voting occurring after public goods contribution decisions in periods 9, 12, 15, and 18. Periods 19 and 20 only contain contribution decisions without team switching

and voting occurring any more. All these rules are public information to participants before the public goods games start. If a participant chooses to switch team, the desired team will see her average contributions, in tokens and as percentage of her endowment, to her team account during (the previous three periods?) and will then vote on the acceptance. If the votes do not reach a majority, the requester is not accepted and will stay with the original team, but her team members will know, by the ID number, she tries to leave but fails. After each period with possible team switching, the MPCR will be updated to reflect the team size according to Table 2 before the next period of team switching. No team is allowed to have fewer than two or more than six members.

**Table 2.** MPCR by Team Size

n (Team Members)	1	2	3	4	5	6
Multiplier Factor	1.000	1.250	1.500	1.750	2	2.25
MPCR	1.000	0.625	0.500	0.438	0.400	0.375

In the treatments with communication, team members can use the built-in chat program to communicate at the public goods contribution stage and at the voting stage.

Overall, 96 participants participated in 12 sessions on the one-hour long online experiment from April 2021 to December 2021. All participants participated in the experiment only once. They received payments based on their cumulative earnings in the public goods games plus their earnings in the anagram challenge and a \$5 participation fee. The exchange rate was 600 points for \$1. The average earnings were \$21.08 per participant including the anagram, public goods game and \$5 participation fee.

The experimenter read the experimental instructions aloud through the beginning of the public goods games while participants were able to follow along on their personal devices. Participants were given practice questions for comprehension check. After the games and before the payoff, participants filled a post-experiment survey that gathered information on demographics, strategy, and background.<sup>6</sup>

---

<sup>6</sup>Survey question list is outlined in Appendix A.

### 3.4 Treatments and Hypotheses

This experiment has two primary treatments: chat and prime. Sessions with the chat treatment have the ability to chat during all contribution decisions as well as during voting on a new member to join a team. Sessions with the *prime* treatment receive a priming survey before the anagram stage.

**3.4.1  $H_1$  (Chat and Priming): When team members are able to communicate, there will be an increase in contributions and a decrease in attempts to switch teams.**

Both the chat and prime treatments are expected to have similar directional effects on participant behavior with respect to their contribution decisions. More specifically, previous experiments have found that chat improves coordination and efficiency in public goods games (Haruvy, Li, McCabe and Twieg, 2017; Oprea, Charness and Friedman, 2014; Palfrey, Rosenthal and Roy, 2017; Palfrey and Rosenthal, 1991). In a similar fashion we expect that priming will improve the desire to coordinate among group members through the mechanism of increasing the value of their identity signaling and therefore increasing the cost of signaling "selfish" behavior to the identity group by holding on to more of one's endowment each round.

**3.4.2  $H_2$  (Team Switching by Round): Participants will be more likely to switch in earlier rounds of play.**

The primary incentive for individuals to switch groups are centered around the improved returns from being on a higher endowed team. It is not expected that many individuals will want to switch to the lower endowed team since initial play is more likely to have higher returns within the higher endowed team. In the beginning of each session, the teams are balanced at 4 each but the experimental design restricts the number of team members to 6, which also leads to a much higher team multiplier. The individuals on a team with 6 members

are highly incentivized to stay on that team due to much higher potential earnings.

**3.4.3  $H_4$  (Feedback): Participants will be more likely to switch teams when they see the opposing team with higher levels of contributions.**

One of the key aspects of our experiment is the transparency of the success of each team between teams. This improves the information set that participants make their decisions with and lets participants know whether their team is higher endowed or lower endowed quickly. The feedback information also informs participants of the relative cooperation levels between their team and the opposing team. This improved information about the potential advantages of the opposing team for individual gain will increase incentives to switch if the other team is doing better than if they did not have the feedback information and had to develop an information asymmetry guessing strategy.

**3.4.4  $H_5$  (Female): Females will have a higher contribution level than male participants.**

Based on previous literature studying the behavior of different genders in public goods games, we expect females to contribute a higher portion of their endowment in the earlier rounds of the public goods game (Cadsby and Maynes, 1998; Charness et al., 2014; Nowell and Tinkler, 1994). There is also supporting work previously done showing women more likely to be voluntarily cooperative in a variety of different settings even at their own risk of receiving a lower benefit than the rest of the group (Babcock, Recalde, Vesterlund and Weingart, 2017; Charness and Rustichini, 2011; Goeree, Holt and Smith, 2017).

## **4 Results**

### **4.1 Contribution Behavior**

Contribution behavior is measured using random effects tobit model where contribution is measured as a percentage of an individual's endowment sorted by which round the contribu-

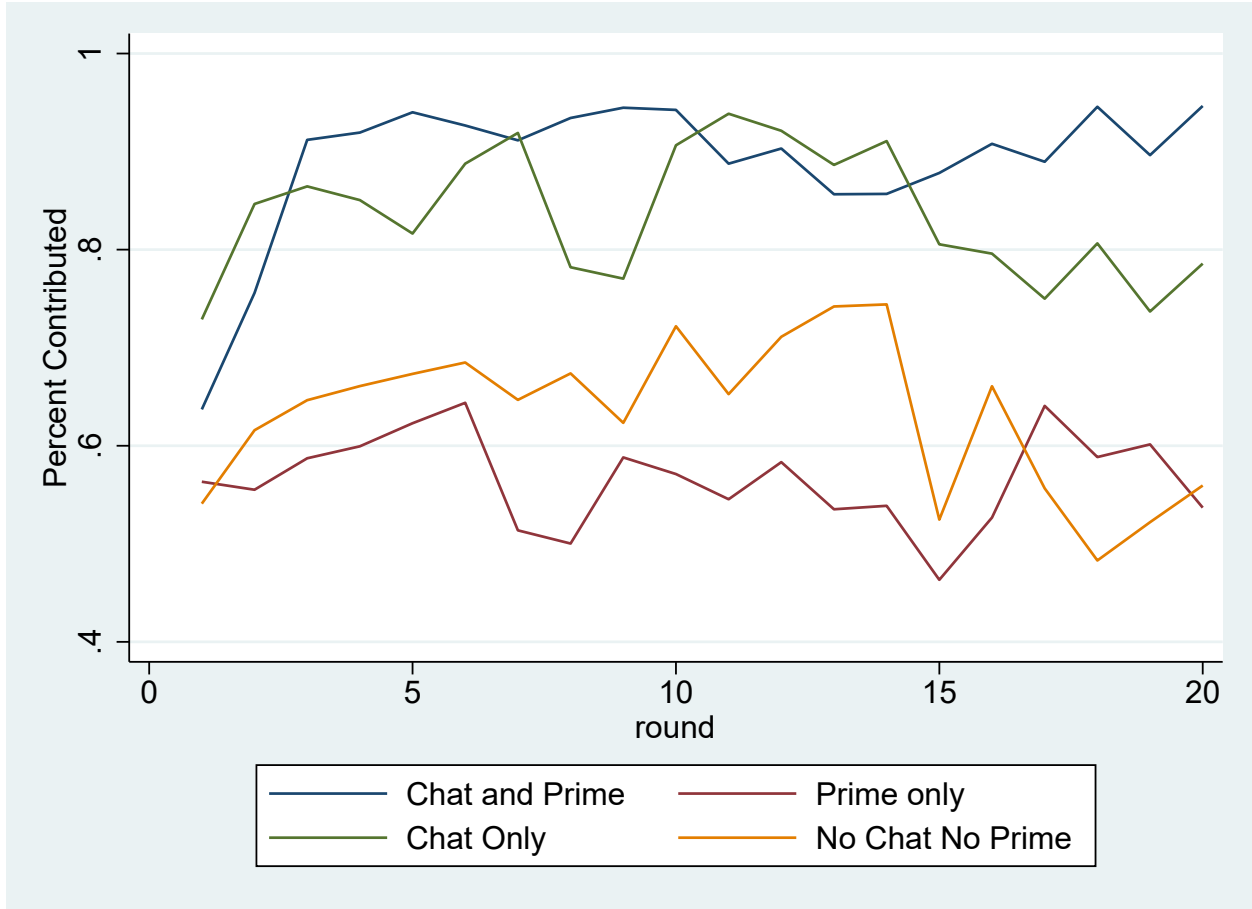
tion decision was made. The dependent percentage of contribution is censored to the range  $C_{i,t} = [0, 1]$  where  $C$  is the percent of endowment contributed by individual  $i$  in round  $r$ . This is expressed by the model

$$C_{i,t} = \beta_0 + \beta_{1,i}Chat_s + \beta_{2,i}Prime_s + \beta_{3,i}Chat_s * Prime_s + \beta_{3i}female_i + \beta_{4i}zEndowment_{i,t} + \psi_{t-1} + \theta_i + \mu_i \quad (1)$$

where Chat and Prime describe the 2x2 treatment design categories,  $zEndowment_{i,t}$  describes the standardized endowment an individual,  $i$ , receives in round  $r$ , while  $\psi_{t-1}$  describe the vector of effects caused by the vector of feedback variables presented to individuals in the previous round including the other team's contribution and the other team's individual returns from the team account. Finally,  $\theta_i$  is the vector of demographic controls.

The baseline estimation between contribution behavior and the treatments indicates that only chat has a significant impact on the percentage of an individual's endowment that they contribute to their team accounts. Having access to chat increases an individual's contribution percentage by 20.3 percentage points ( $p < .01$ ), while priming has no statistically significant effect, nor does the interaction between priming and chat as illustrated in Figure 2. When consideration is added for feedback variables regarding the other team's performance, the effect of chat is reduced from .203 to .155 ( $p < 0.01$ ) This result does not show any statistically significant changes for gender, same as the simpler specifications, implying that an individual's gender does not greatly influence their contribution strategy. This is at odds with other literature on public goods games that have found that women tend to have higher initial contribution levels that taper off over multiple rounds (Cadsby and Maynes, 1998) or have an aptitude to participate in public goods in the field (Greig and Bohnet, 2009) but is not unique in this regard (Brown-Kruse and Hummels, 1993) different time horizons and settings have found. While not statistically significant, there is a negative effect from

**Figure 1.** Contribution Behavior by Treatment Group



priming on percentage of endowment contributed.

One interesting result is that the standardized endowment consideration has a negative impact on the percentage of the gross endowment that an individual contributes ( $-0.0274$ ,  $p < 0.01$ ). The implication of this result is that for every experimental unit that a participant is endowed with, they will reduce their average contribution to the team account by 2.7 percentage points of their gross endowment. This is not completely surprising and can be simply understood when considering that even though the percentage of their endowment is decreasing, we observe that their gross contribution is rising and due to the higher endowment, participants can rationalize providing a higher total amount while retaining a larger portion of their points. This is similar to someone who is wealthy emphasising the total amount of money they contribute to taxes or charity rather than the percentage of their



wealth.

There is a statistically strong but practically insignificant result for understanding how the seeing the other team's feedback impacts a participants decision making. Notably, in the fourth specification, we observe highly significant ( $p < .01$ ) estimates for the impact on seeing the other team's individual payouts ( $\beta = 6.99e^{-5}$ ) and the total amount contributed to their group account ( $\beta = 1.77e^{-4}$ ). Both of these have an average impact on percentage points of less than 0.0018 percentage point increase of endowment contributed in a given round. The behavior here suggests that when participants are unhappy with their team's contributions, they may consider switching teams rather than trying to convince their own team to change contribution strategies after just a few rounds.

Further, the feedback that participants get each round informs the individual about the opposing team's performance, which may impact the way they think about their own contribution strategy. The specification that controls for this does observe strong statistical significance around the contributions of the other team in the previous round but the effect is a small, albeit tightly estimated, nearly 0 effect of 0.00007 ( $p < .01$ ). This is a similar effect as displayed by the lagged impact of the individual's own group contribution from the previous round of 0.00018 ( $p < .01$ ). This implies that that the impact of an individual's own group and the opposite group have a predictable but nearly 0 positive impact on contribution decision making.

## 4.2 Team Switching Behavior

Team switching behavior is measured using a random effects logit model estimating the log odds that an individual attempts to switch teams. This is specific to whether an individual selects to try to switch teams during a voting period and does not imply that they are

**Table 3.** Percent of Endowment Contributed

VARIABLES	(1) % Contributed	(2) % Contributed	(3) % Contributed	(4) % Contributed
Chat	0.203*** (0.0567)	0.202*** (0.0571)	0.203*** (0.0561)	0.155*** (0.0505)
Prime	-0.0670 (0.0567)	-0.0699 (0.0571)	-0.0601 (0.0562)	-0.0378 (0.0502)
Chat×Prime	0.121 (0.0802)	0.125 (0.0807)	0.115 (0.0794)	0.0804 (0.0711)
female		0.00718 (0.0404)	0.00482 (0.0397)	0.0137 (0.0353)
Endowment		-0.0166** (0.00669)	-0.0244*** (0.00682)	-0.0274*** (0.00686)
MPCR			0.135*** (0.0275)	-0.00276 (0.0427)
OtherTeamCont <sub>t-1</sub> <sup>a</sup>				0.0699*** (0.0267)
OwnTeamCont <sub>t-1</sub> <sup>b</sup>				0.177*** (0.0239)
Constant	0.632*** (0.0401)	0.630*** (0.0451)	0.378*** (0.0676)	0.445*** (0.0910)
Observations	1,920	1,920	1,920	1,824
Number of id	96	96	96	96

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>a</sup>in thousands of points

<sup>b</sup>in thousands of points

successful in making the switch. This is measured by the model

$$S_{i,t} = \beta_0 + \beta_{1,i}Chat_s + \beta_{2,i}Prime_s + \beta_{3,i}Chat_s * Prime_s + \beta_{3i}female_i + \beta_{4i}HE_{g,t} + \psi_{t-1} + \theta_i + \mu_i \quad (2)$$

where  $HE_{g,t}$  indicates if an individual is in the higher endowment group during a given round.

Unlike contribution strategies, team switching behavior shows to be influenced by both

**Table 4.** Logit estimating attempts to switch

VARIABLES	(1) wantswitch	(2) wantswitch	(3) wantswitch	(4) wantswitch
Chat	3.023** (1.410)	2.726** (1.145)	2.705** (1.122)	2.965** (1.171)
Prime	3.100** (1.413)	2.734** (1.147)	2.706** (1.124)	2.656** (1.161)
Chat×Prime	-5.679*** (2.032)	-5.193*** (1.635)	-5.141*** (1.605)	-5.147*** (1.658)
Female		-5.610*** (0.919)	-5.504*** (0.908)	-5.717*** (0.941)
HighEndowment			0.449 (0.486)	-0.392 (0.679)
OtherTeamCont <sub>t-1</sub> <sup>a</sup>				-1.16* (0.645)
Constant	-2.486** (0.974)	0.499 (0.890)	0.226 (0.920)	1.51 (1.196)
Observations	576	576	576	576
Number of id	96	96	96	96

Standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

<sup>a</sup>in thousands of points

chat coordination and priming treatments, making the desire to switch 2.810 and 2.541 times more likely, respectively. This is counter to the expected outcomes outlined in hypotheses 2 and 4 since coordination via chat and priming make individuals more likely to attempt team switching when one or the other is present. Testing the interaction between the two delivers an opposite effect with a coefficient of -5.079, significant at the 1% level. This makes the combined effect of both chat and prime lower than the individual treatments but still positive at 0.272.

Table 6 shows the likelihood of a successful switch given an attempt to switch.<sup>7</sup> Endowment showed to have strong impacts on the contribution behavior for individuals but at the same time, there is no statistical evidence that endowment impacts an individual's decision

<sup>7</sup>To see the effects of total switchers leaving one group for another, see Table 7 in Appendix E.

to stay with their team, specifically whether the individual was a member of the high endowment team during a given round. Gender does have very strong impacts on switching behavior. That is, females in the study demonstrated that they were over five times less likely (-5.311) to switch teams, robust to whether they were in a high endowment team or not. This result was also robust across treatments. This result was further tested by round to understand the switching behavior by gender identities over the experimental time horizon as displayed in Table 5.

Across specifications, it is clear that the likelihood an individual participant attempts to switch teams is inversely related with how many rounds have passed. Again, this is expected since each team starts with 4 members but a maximum possible team size of 6. It would be expected that those who are going to switch would attempt to in the earlier part of the experiment, which by Table 5 is shown to be in the first half of the experiment. By the second half of the experiment, there are still people wanting to switch teams, but the opposite team would be full and participants would be aware of that, discouraging them from even indicating that they wanted to switch teams. This is even more likely when considering that with only 2 members in a public goods game, retaliation for trying to switch teams can be applied directly compared to when there are  $n \geq$  members and retaliation cannot be well directed since defection play would signal general uncooperative action taking that would adversely impact the entire team.

MPCR and feedback variables did not have enough power to show any statistically significant results.

### **4.3 Voting Behavior**

Voting behavior in this experiment is only observed every third round, meaning there are six voting periods that participants will engage in. Whether or not a given individual will have to vote or be voted on in any voting period is dependant on them either trying to switch teams or someone trying to join their team and there being a vacant slot on the

respective team to vote on. Votes were made one participant at a time and were cast in a binary fashion. This means that if two participants were both trying to switch into a team with only one vacant opening, any individual already on that team would be allowed to vote "yes" for both participants trying to join the team. In other words, votes were not exclusive between participants trying to join a given team with insufficient openings and a tie would be resolved by a random pick from the experimental application.

The outcome of a respective participant's attempt to switch teams was designed as a simple binary where they either were successfully voted into the new team ( $switch = 1$ ) or they were rejected by the new team ( $switch = 0$ ). Voting success is modeled in this binary fashion.<sup>8</sup> This outcome is modeled as:<sup>9</sup>

$$logit(switch_{i,t}) = \beta_0 + \beta_{1,i}female_i + \beta_{2,i}t + \beta_{3,i}Chat_s + \beta_{4,i}Prime_s + \beta_{5,i}Chat_s * Prime_s + \theta_i + \mu_i \quad (3)$$

It is important to note that the round an individual attempts to switch teams is needs to be controlled for due to the inverse relationship between switching and round number. In later rounds of the experiment, teams tend to stabilize due to one of the teams reaching the maximum number of members and MPCR being sufficiently large to discourage anyone from leaving, as shown in Table 5.

Table 6 gives the results of the logit models estimating the outcomes of an attempt to switch teams. We start with a simplified model in specification (2) to test for a difference between male and female participants attempting to switch. The log odds of a female participant trying to switch teams are slightly better than that of a male participant (1.068 vs constant of -1.842) but no detectable significance based on the percent of contribution

---

<sup>8</sup>If there are adequate slots available a participant is successfully voted in by a simple majority. In the event that there are more participants voted in than there are spots available, the participants with the highest number of votes are voted in first and then any ties are decided by a random number draw from a uniform distribution.

<sup>9</sup>The odds of a successful switch are conditional on an individual attempting to switch in the first place such that  $logit(switch_{i,t}) = \{logit(switch_{i,t})|wantswitch_{i,t} = 1\}$ .

that the participant has given in previous rounds. In the preferred specification (3), we see a slightly stronger advantage of women vs men in trying to gain access to a new group with log odds of success increasing 1.153 for women compared to the baseline -1.269. The likelihood of success does not appear to be strongly differentiated by treatments other than the joint treatment when participants are primed and have the ability to chat, which shows slight significance of improving the log odds of a successful switch (1.937) over the baseline constant when neither treatment is employed (-1.269).

## 5 Discussion and Conclusion

The purpose of this experiment was to explore how group identity impacts an individual's decision to stay with their team and improve their team's overall position when faced with the opportunity cost of their individual payout. This was achieved by setting up an environment where participants were participating in two separate public goods games in respectively gender homogeneous teams. The two teams can see the each other's performances and were given 6 opportunities to switch teams. Through this process we are able to see what kind of patterns there are in contribution strategies within identity homogeneous teams, the preferences of identity groups to accept out-group members, and the impacts of chat and identity priming.

Chat was found to be one of the biggest contributors to team account contribution strategies and this was robust across genders. This is consistent with previous literature and matches the intuition that improved coordination capabilities between team members would lead to more group-centered behavior to everyone's benefit. The effect of gender priming is not as clear and statistically insignificant as is the interactive effect between priming and chat capabilities. Of note, however is that MPCR does have a positive impact on percentage of endowment contributed, meaning participants were more likely to give on average 13.5 percentage points more of their endowment for each increase in the MPCR. In other words, people give more when they get more out of it. Other specifications of the model from 3 show

MPCR being insignificant when we looked at lagged effects on feedback from the opposing team. This does not discount the importance of the MPCR in contribution strategies though due to the set of the experiment creating a negative correlation between a team's MPCR and the other team's individual payouts. That is, since the number of participants in given session is capped at 8, then the higher a team's MPCR is, the lower the MPCR of the other team and thus lower individual payouts. Overall, these results match what previous public goods games experiments have found, save the other team impacts.

Identity played a much bigger role in the decision to stay or leave a given team rather than how much to contribute. Participants contribute what they will contribute regardless of their gender identity but their strategy with respect to loyalty to their identity-based team is a separate decision. We see that while chat still has a strong effect a participant's decision to attempt a team switch, it is in the opposite direction than expected, same with priming. The idea here is mostly based around how participants deal with defection. When chat is enabled, participants are able to more quickly and surely ascertain if their teammates playing a "selfish" strategy was just optimization or defecting from a team focused plan. When identity priming along is introduced we see a similar impact where participants are more likely to switch and this can be partially attributed to "betrayal" from their team. The interaction between priming participants gender identity and allowing them to chat with each other essentially counteracts the additive effects of chat and prime on average inducing switching for men. Women in the experiment still show a five time less likely chance to try and switch teams, leaving us with our conclusion that women are more affected by their identity group, regardless of the priming.

Even though women were less likely to try to switch teams during this experiment, they had a slight advantage in being successful at switching when they did attempt from 6. This result has two potential mechanisms. One is that women made themselves more attractive candidates when they decided to switch. This is possible because participants are aware of what information is portrayed to the opposing team when voting on potential switchers

and the participants know the schedule of switching opportunities. This mechanism is not favorable because it does not hold up without evidence that females at any point contribute a higher percentage on average. The randomization of endowment ranges also guarantees that there is an even split between high endowment females and males across sessions. The other mechanism to explain the marginally higher female success is the behavior of men in the experiment. Rather than the assumption that the women make themselves more attractive via their contributions, men may be more willing to take on additional team members in order to increase their team account multiplier.

The results of this experiment at the current stage warrant further investigation into the role that identity plays in public goods games and using this experimental design to continue exploring the effect of group loyalty. This work contributes to the existing literature by demonstrating how gender identity has value in a multiple time period setting when considered against individual monetary gain. In particular, women demonstrate a much lower likelihood to try and change teams even when they are on a lower endowed team. We also find that there is an interesting consideration that participants, when given the ability to switch teams, do not necessarily try to improve their own team's cooperation through action if they get feedback that the opposing team is doing better at cooperation. Rather, they continue at their current endowment levels waiting for an opportunity to switch.



## References

- Akerlof, George A and Kranton, Rachel E.** (2000). ‘Economics and Identity’, *The Quarterly Journal of Economics* 115(3), 715–753.  
URL: <https://academic.oup.com/qje/article-abstract/115/3/715/1828151>
- Akerlof, George A and Kranton, Rachel E.** (2002). ‘Identity and schooling: Some lessons for the economics of education’, *Journal of Economic Literature* 40(4), 1167–1201.
- Akerlof, George A and Kranton, Rachel E.** (2005). ‘Identity and the economics of organizations’, *Journal of Economic Perspectives* 19(1), 9–32.
- Akerlof, George A. and Kranton, Rachel E.** (2008). ‘Identity, Supervision, and Work Groups’, *American Economic Review: Papers Proceedings* 98(2), 212–217.  
URL: <http://www.aeaweb.org/articles.php?doi=10.1257/aer.98.2.212>
- Austen-Smith, David and Fryer Jr, Roland G.** (2005). ‘An economic analysis of “acting white”’, *The Quarterly Journal of Economics* 120(2), 551–583.
- Babcock, Linda, Recalde, Maria P, Vesterlund, Lise and Weingart, Laurie.** (2017). ‘Gender differences in accepting and receiving requests for tasks with low promotability’, *American Economic Review* 107(3), 714–47.
- Bertrand, Marianne, Kamenica, Emir and Pan, Jessica.** (2015). ‘Gender identity and relative income within households’, *The Quarterly Journal of Economics* 130(2), 571–614.
- Booth, Alison L and Van Ours, Jan C.** (2009). ‘Hours of work and gender identity: Does part-time work make the family happier?’, *Economica* 76(301), 176–196.
- Brown-Kruse, Jamie and Hummels, David.** (1993). ‘Gender effects in laboratory public goods contribution: Do individuals put their money where their mouth is?’, *Journal of Economic Behavior & Organization* 22(3), 255–267.
- Burke, Peter J.** (1980). ‘The self: Measurement requirements from an interactionist perspective’, *Social psychology quarterly* pp. 18–29.
- Cadsby, C Bram and Maynes, Elizabeth.** (1998). ‘Gender and free riding in a threshold public goods game: Experimental evidence’, *Journal of economic behavior & organization* 34(4), 603–620.
- Charness, Gary, Cobo-Reyes, Ramon and Jimenez, Natalia.** (2014). ‘Identities, selection, and contributions in a public-goods game’, *Games and Economic Behavior* 87, 322–338.
- Charness, Gary and Rustichini, Aldo.** (2011). ‘Gender differences in cooperation with group membership’, *Games and Economic Behavior* 72(1), 77–85.
- Chen, Daniel L, Schonger, Martin and Wickens, Chris.** (2016). ‘oTree—An open-source platform for laboratory, online, and field experiments’, *Journal of Behavioral and Experimental Finance* 9, 88–97.

- Chen, Yan and Li, Sherry Xi.** (2009). ‘Group Identity and Social Preferences’, *American Economic Review* 99(1), 431–457.  
 URL: <http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.1.431>
- David, Austen-Smith and Fryer, Roland G.** (1995), An Economic Analysis of ”Acting White”, Technical report.  
 URL: <https://academic.oup.com/qje/article-abstract/120/2/551/1933943>
- Fordham, Signithia and Ogbu, John U.** (1986). ‘Black students’ school success: Coping with the “burden of ‘acting white’”’, *The urban review* 18(3), 176–206.
- Fryer, Roland G. and Torelli, Paul.** (2010). ‘An empirical analysis of ‘acting white’’, *Journal of Public Economics* 94(5-6), 380–396.
- Goeree, Jacob K, Holt, Charles A and Smith, Angela M.** (2017). ‘An experimental examination of the volunteer’s dilemma’, *Games and Economic Behavior* 102, 303–315.
- Greenstein, Theodore N.** (2000). ‘Economic dependence, gender, and the division of labor in the home: A replication and extension’, *Journal of Marriage and family* 62(2), 322–335.
- Greig, Fiona and Bohnet, Iris.** (2009). ‘Exploring gendered behavior in the field with experiments: Why public goods are provided by women in a Nairobi slum’, *Journal of Economic Behavior & Organization* 70(1-2), 1–9.
- Haruvy, Ernan, Li, Sherry Xin, McCabe, Kevin and Twieg, Peter.** (2017). ‘Communication and visibility in public goods provision’, *Games and Economic Behavior* 105, 276–296.
- Hogg, Michael A and Abrams, Dominic.** (1988), *Social identifications: A social psychology of intergroup relations and group processes.*, Taylor & Frances/Routledge.
- Hogg, Michael A and Turner, John C.** (1985). ‘Interpersonal attraction, social identification and psychological group formation’, *European journal of social psychology* 15(1), 51–66.
- Horvat, Erin McNamara and Lewis, Kristine S.** (2003). ‘Reassessing the” burden of ‘acting White’”: The importance of peer groups in managing academic success’, *Sociology of education* pp. 265–280.
- Ke, Da.** (2021). ‘Who Wears the Pants? Gender Identity Norms and Intrahousehold Financial Decision-Making’, *The Journal of Finance* 76(3), 1389–1425.
- McCall, G.J. and Simmons, J.L.** (1968). ‘Identities and Interactions’, *American Journal of Sociology* 74(1).
- Nowell, Clifford and Tinkler, Sarah.** (1994). ‘The influence of gender on the provision of a public good’, *Journal of Economic Behavior & Organization* 25(1), 25–36.
- Ogbu, John U.** (2004). ‘Collective identity and the burden of “acting White” in Black history, community, and education’, *The Urban Review* 36(1), 1–35.

- Oprea, Ryan, Charness, Gary and Friedman, Daniel.** (2014). ‘Continuous time and communication in a public-goods experiment’, *Journal of Economic Behavior & Organization* 108, 212–223.
- Palfrey, Thomas, Rosenthal, Howard and Roy, Nilanjan.** (2017). ‘How cheap talk enhances efficiency in threshold public goods games’, *Games and Economic Behavior* 101, 234–259.
- Palfrey, Thomas R and Rosenthal, Howard.** (1991). ‘Testing for effects of cheap talk in a public goods game with private information’, *Games and economic behavior* 3(2), 183–220.
- Salland, Jan.** (2018). ‘Income comparison, gender roles and life satisfaction’, *Applied Economics Letters* 25(20), 1436–1439.
- St Claire, Lindsay and Turner, John C.** (1982). ‘The role of demand characteristics in the social categorization paradigm.’, *European Journal of Social Psychology* .
- Stets, Jan E and Burke, Peter J.** (2000). ‘Identity theory and social identity theory’, *Social psychology quarterly* pp. 224–237.
- Tajfel, Henri, Turner, John C, Austin, William G and Worchel, Stephen.** (1979). ‘An integrative theory of intergroup conflict’, *Organizational identity: A reader* 56(65), 9780203505984–16.
- Terry, Deborah J, Hogg, Michael A and White, Katherine M.** (1999). ‘The theory of planned behaviour: self-identity, social identity and group norms’, *British journal of social psychology* 38(3), 225–244.
- Turner, John C, Hogg, Michael A, Oakes, Penelope J, Reicher, Stephen D and Wetherell, Margaret S.** (1987), *Rediscovering the social group: A self-categorization theory.*, Basil Blackwell.
- Turner, Ralph H.** (1978). ‘The role and the person’, *American journal of Sociology* 84(1), 1–23.
- Tyson, Karolyn, Darity Jr, William and Castellino, Domini R.** (2005). ‘It’s not “a black thing”: Understanding the burden of acting white and other dilemmas of high achievement’, *American sociological review* 70(4), 582–605.
- Wildhagen, Tina.** (2011). ‘Testing the ‘acting White’ hypothesis: A popular explanation runs out of empirical steam’, *The Journal of Negro Education* pp. 445–463.

**Table 5.** Switching Attempts by round and Gender

VARIABLES	(1) wantswitch	(2) wantswitch	(3) wantswitch	(4) wantswitch
round 3		1.284** (0.553)	3.638*** (0.888)	3.977*** (0.926)
round 6		1.284** (0.553)	3.638*** (0.888)	3.802*** (0.893)
round 9		1.147** (0.551)	2.433*** (0.806)	2.586*** (0.816)
round 12		0.150 (0.549)	0.803 (0.742)	0.824 (0.743)
round 15		-0.153 (0.553)	0.269 (0.734)	0.197 (0.729)
female			-4.807*** (1.265)	-4.675*** (1.249)
round 3*female			-5.402*** (1.361)	-5.290*** (1.335)
round 6*female			-5.402*** (1.361)	-5.311*** (1.349)
round 9*female			-2.799** (1.181)	-2.816** (1.187)
round 12*female			-1.577 (1.162)	-1.546 (1.181)
round 15*female			-1.043 (1.155)	-0.919 (1.171)
round	-0.110*** (0.0322)			
highendowment				1.168* (0.608)
Constant	0.227 (0.659)	-1.570** (0.725)	0.981 (0.882)	0.221 (0.950)
Observations	576	576	576	576
Number of id	96	96	96	96

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 6.** Likelihood of Successful Switch

VARIABLES	(1) switch	(2) switch	(3) switch	(4) switch
Female	1.069** (0.522)	1.068** (0.489)	1.153** (0.547)	1.179** (0.570)
Round	-0.100** (0.0495)	-0.0996** (0.0487)	-0.0926* (0.0504)	-0.0933* (0.0507)
Chat			-1.043 (0.715)	-1.018 (0.730)
Prime			-1.249 (0.762)	-1.358 (0.842)
Chat×Prime			1.937* (1.098)	2.103* (1.228)
% Contributed		0.185 (0.778)		-0.373 (1.080)
Constant	-1.842*** (0.674)	-1.916** (0.786)	-1.269** (0.645)	-1.008 (0.991)
Observations	261	261	261	261
Number of id	63	63	63	63

Standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

## A Appendix A: Post-Experiment Questionnaire

1. How close do you feel to your team from the anagram task?
  - Very Close
  - Somewhat Close
  - Neutral
  - Somewhat Distant
  - Very Distant
2. Why did you choose to switch or not switch teams?
3. How did you choose how much to contribute?
4. How old are you?
5. What is your gender?
6. What year are you?
7. Are you a varsity athlete? If so, which team are you on?
8. Are you a member of a Greek Life Organization? If so, which one?
9. Which college are you matriculated in?
  - Dale Bumpers College of Agricultural, Food and Life Sciences
  - Fay Jones School of Architecture and Design
  - J. William Fulbright College of Arts and Sciences
  - Sam M. Walton College of Business
  - College of Education and Health Professions
  - College of Engineering
  - School of Law
10. What is your major?
11. Are you a first generation college student?

- Yes
- No

12. Are you of latinx/e descent?

13. If so, which do of the following do you ethnically identify with:

- Mexican
- Central American
- South American
- Caribbean

14. Which racial categories do you identify with:

- Black
- White
- Asian
- Native American/Alaska Native

15. What was your childhood zipcode?

## B Appendix B: Priming Questionnaire

1. Do you normally live on campus or off campus?
  - Off Campus
  - On Campus
2. Do you normally have a roommate?
  - I do not have a roommate
  - I have a roommate
3. Is your floor single sex or co-ed?
  - Single Sex
  - Co-ed
4. Do you prefer a single sex floor or co-ed?
  - Single Sex
  - Co-ed
5. Why is this your preference?
6. Is your living environment single sex or co-ed?
  - Single Sex
  - Co-ed
7. Why is this your preference?
8. Do you prefer a single sex living environment or co-ed?
  - Single Sex
  - Co-ed
9. Why is this your preference?

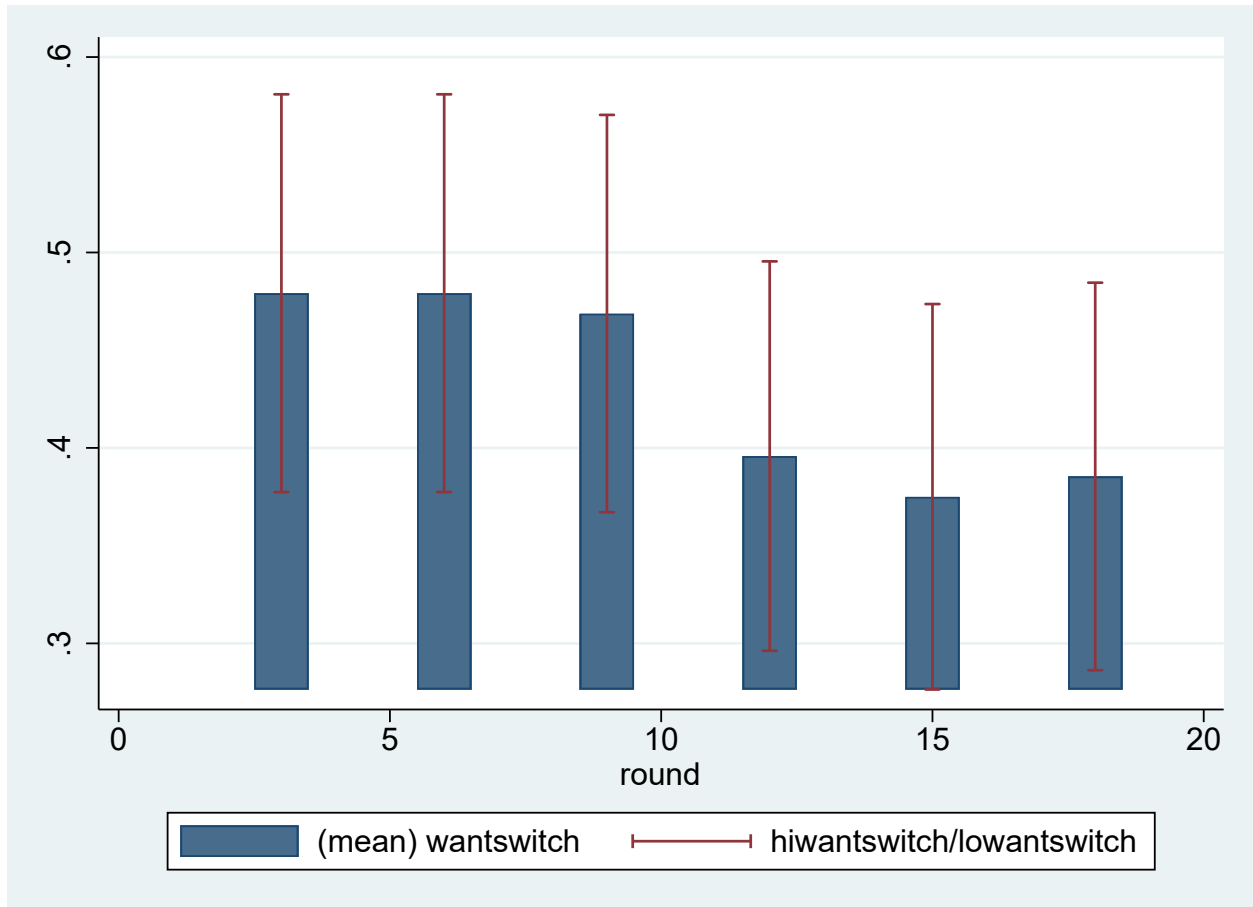


## C Appendix C: Identity-Neutral Questionnaire

1. Which smartphone operating system do you prefer?
  - Android
  - iPhone
2. Do you prefer voice calls or video calls?
  - Voice calls
  - Video calls
3. Do you prefer using a laptop computer or a desktop computer?
  - Laptop computer
  - Desktop computer
4. Do you prefer a mobile app or a mobile website?
  - Mobile app
  - Mobile website
5. Do you prefer using a keyboard/keypad or a touchscreen?
  - Keyboard/keypad
  - Touchscreen
6. Do you prefer using a remote control or your smartphone to control smart devices such as a smart TV?
  - Remote control
  - Smart device

## D Appendix D: Rate of Switching Requests by Round

Figure 2. Rate of Switching Requests by Round



## E Appendix E: Total Switching by Group

Table 7 measures how many individuals leave a team clustering on the session level. The interpretation should be how many individuals leave the indicated team, on average, in a given session.

**Table 7.** Number of Switchers Leaving the Team

VARIABLES	(1) Switchers	(2) Switchers	(3) Switchers	(4) Switchers
Female Team	-0.308** (0.122)	-0.308** (0.121)	-0.308** (0.119)	-0.308*** (0.118)
High Endowment Team	-0.240*** (0.0554)	-0.240*** (0.0554)	-0.246*** (0.0553)	-0.248*** (0.0552)
Chat			-0.260 (0.169)	-0.275* (0.167)
Prime			-0.379** (0.169)	-0.375** (0.167)
Chat×Prime			0.394* (0.239)	0.385 (0.236)
%Contribution		0.0693 (0.0540)		0.0717 (0.0544)
Constant	0.766*** (0.0924)	0.716*** (0.100)	0.992*** (0.138)	0.947*** (0.141)
Observations	1,920	1,920	1,920	1,920
Number of id	96	96	96	96

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## Chapter 3: Identity Based Risk and Time Preference Predictions

Nathaniel Burke<sup>1</sup>

### Abstract

Assessing the risk and time preferences of other people is a valuable skill in strategic decision making situations such as negotiations, debate, sports, or even gambling. The accuracy of these assessments can vary greatly based on the predictors experience and what information they know about their opponents, which often times may only be immutable characteristics such as presented gender or race. This paper uses a two phase experiment to test how well individuals can predict the race and time preferences of a sample of people with only gender or demographic information to make their predictions. Specifically, this paper explores if there is an information advantage to making in-group predictions about someone who shares the same gender or political identity. In phase 1, participants participate in a modified double multiple price list (DMPL) to elicit their respective preferences. In phase 2, predictors make predictions about decisions made in the DMPL and then adjust their predictions for different gender and political subgroups.

**JEL classification:** D91, D84, C92.

**Keywords:** Risk Preference, Time Preference, Identity Bias

---

<sup>1</sup>West Virginia University: [nathaniel.burke@mail.wvu.edu](mailto:nathaniel.burke@mail.wvu.edu)

## 1 Introduction

Information about time and risk preferences is usually obtained by incentivized elicitation or through unincentivized questioning. How these preferences are formed can be attributed to many things such as experience or how they were taught to think about risk and preferences and timing, or even cultural implications towards risk and patience. Regardless of where an individual's preferences come from, others can observe these preferences in action based on decisions made.

This paper outlines an experimental approach to evaluate subject perceptions regarding the risk and time preferences of others based entirely on certain demographic information. In this paper, individuals make predictions about risk and time preferences using only reported gender and political identities. The idea behind testing the kinds of predictions that people make based on identity measures is there are incentivized decisions that are often made based on someone's perception or prediction of another person's risk and time preferences. These decisions could look like anticipating counters in negotiations, it comes into play in athletics or other tactical environments such as chess, gambling, or even business or legal decisions. When entering into one of these scenarios and having to come up with strategies, someone will have to make a decision about how aggressive or risky their opponent is when trying to come up with their strategy, whether it is an opposing coach making a judgement based on an individual's gender, or a gambler going off of a gut feeling.

The question specifically explored in this methodology is whether there is an advantage when making these judgements about someone if they are in your own identity group vs an opposing group, or are women better at predicting the preferences of other women than they are with other men? Beyond just looking at if there is improved information about in-group, this experiment also explores if identity groups have any advantage over their respective out-groups when comparing how much better they are at predicting in-group preferences. This approach also allows for testing of how much bias individuals have regarding the preferences

of respective identity groups.

The within subject experimental design to test these perceptions starts with a first phase of deciders participating in incentivized time and risk preference elicitation using multiple price lists. The individuals making these decisions are incentivized by being told that one of their decisions will be randomly chosen to count. As noted in Azrieli, Chambers and Healy (2018) and Laury (2005), it is as effective to randomly select one task to incentivize participants as it is to incentivize multiple or all tasks<sup>2</sup>. The multiple price lists were set up so that for the risk preference elicitation, participants chose between two lotteries, one "safer" and one "riskier". The lotteries stayed the same across ten choices while the probability of getting the higher payout for the chosen lottery monotonically increased, therefore increasing the expected value of the riskier lottery. In similar fashion, the time preferences used two fixed time periods, now vs 2 weeks later, where the "get paid now" option monotonically decreased and was always lower than the "get paid in 2 weeks" option. Participants had to make decisions about 10 risk choices and 6 time based choices. Next, a second phase of predictors was asked to submit their predictions about what percentage of deciders chose to be risky in each risk preference lottery decision and what percentage chose to be patient in each time preference decision. Once they made their initial predictions, predictors were asked to submit their predictions about specific subgroups including male, female, conservative, and liberal deciders. The results show that while there is not a significant difference in the in-group:out-group error ratios between identity group predictions across the entire distribution of decisions, there are some significant differences in the higher risk portions of the risk preference decisions. This meaning that women are more in tune with other women regarding their risk preferences when the probability of receiving a more favorable outcome is low than men are with other men. There are also statistically insignificant trends with similar results among political identity groups.

The rest of this paper follows an order as listed: Section 2 goes over a review of re-

---

<sup>2</sup>Azrieli et al. (2018) further argues that it is preferable to actually choose one task to incentivize and may not be proper to incentivize every decision and task performed by participants

lated previous literature, Section 3 outlines the experimental design used in this paper, Section 4 discuss the results found in the experiment, and 5 gives conclusions based on the experimental results and discusses follow on avenues of research on this topic.

## 2 Literature Review

There has been a lot of previous work using multiple price lists to elicit risk preferences (Andersen, Harrison, Lau and Rutström, 2006; Drichoutis and Lusk, 2016; Holt and Laury, 2002, 2005). The simple idea behind the multiple price list format is that there is a monotonic change risk level or patience level respectively required to choose the column B option. Once the risk level is within an individual's tolerance, they will switch to the second column lottery. Observing this decision in expectation, the expected payout of column B gets incrementally higher and at some point will switch to having a higher expected payout than the "safer" Column A. Since the change in probabilities and therefore expected value is monotonic, a rational individual should have one switching point from Column A to Column B as outlined in Holt and Laury (2002). While there are many other methods to examine risk elicitation which boast simpler analysis and implementation (Dave, Eckel, Johnson and Rojas, 2010; Eckel and Grossman, 2002; Gneezy and Potters, 1997; Lejuez, Read, Kahler, Richards, Ramsey, Stuart, Strong and Brown, 2002), multiple price list approach has seen some a lot of popularity in economics for its ability to identify a single "switching point" (Charness, Gneezy and Imas, 2013).

While this paper looks at the perceived differences in risk preferences between gender and political groups, this does leave the initial condition that there is a reason for individuals to believe there is a difference in the preferences and risk tolerance between different groups. There has not been substantial work done on the differences between political groups and their associated preferences but gender differences have been well studied. The current state of the literature suggests that if there is any difference it is that males tend to have a higher risk tolerance than females (Charness and Gneezy, 2012; Holt and Laury, 2002; Hunt, Hopko,

Bare, Lejuez and Robinson, 2005; Lejuez et al., 2002; Weber, Blais and Betz, 2002). Many of the significant differences where males are found to have higher risk tolerances than females are found in elicitation designs that are simpler for the participants. Part of this has been thought to be that the complex MPL design is easier for participants to properly understand the instructions and thus behave more true to their incentivized preference (Charness et al., 2013). Other literature has found that males and females have similar risk preferences across different elicitation techniques (Dohmen, Falk, Huffman and Sunde, 2010; Fecteau, Pascual-Leone, Zald, Liguori, Théoret, Boggio and Fregni, 2007; Gneezy, Leonard and List, 2009; Jacobson and Petrie, 2009; Laury, 2005).

Time preference elicitation is often done separately from risk preference elicitation to ensure there is not a conflation when time preferences are observed under risk (Andreoni and Sprenger, 2012). Andersen, Harrison, Lau and Rutström (2008) describes using a double multiple price list strategy to elicit time and risk preferences in one study but for this study it was important that later predictors would be able to clearly indicate their predictions for risk and time preferences, respectively so this paper makes a modification on the double multiple price list . Similar to risk preference elicitation methods, there are a variety of approaches for eliciting time preferences. Much of the difference in which methods are most advantageous is dependent on the setting (field vs lab) and time horizon available.

Previous work in understanding the differences between gender groups and their perceived preferences was done in Eckel and Grossman (2002) where the authors implemented a laboratory experiment using the Zuckerman Sensation-Seeking Scale (Zuckerman, 1994) and the now known Eckel-Grossman risk elicitation method. The paper finds a positive difference between men and women's risk preferences, with women tending to be more risk adverse, but they also explored the ability of participants to make guesses about the risk choices of their fellow participants with rewards given for a correct answer. It should be noted that in this elicitation method, there are 5 fixed choices in the risk preference elicitation, so it is much more realistic for a participant to get it precisely correct than in the set up described in Sec-



tion 3 of this paper where participants must predict a value from 0-100. The authors found that women tended to overestimate the risk aversion of others, especially other women, while all participants did better than what would occur in expectation from random guessing. My work extends beyond the original scope of Eckel and Grossman (2002) by using a within subjects design to get more observations between different matchups between genders, it also looks at the effects of political groups. The two other major differences are, as already stated, participants make predictions about what proportion of individuals are choosing each option in a MPL, rather than making a guess of which option an individual chooses, and the framing of my study consistently reminds the participant of their predictions when given no information, strengthening the prediction about the whole group as a baseline for analysis of group predictions.

### 3 Experimental Design

This study consisted of two stages: the Elicitation (Decision) Stage and the Prediction Stage. Each stage took place online with subjects recruited from the University of Arkansas. Both stages refer to the same 16 decisions but with the second stage observing these decisions broken down into four demographic subgroups as well as the whole group.

55 subjects were recruited for the Decision Stage and 155 subjects were recruited for the Prediction Stage. Payments are described in Table 1

**Table 1.** Average Payments by Stage

Payment	N	SUF	Avg Total
Decision Stage	55	\$2.00	\$6.497
Prediction Stage	155	\$5.00	\$8.065

#### 3.1 Decisions Stage: Risk and Time Preference Elicitation

Subjects made all of their decisions in the context of two multiple price lists designed to elicit risk and time preferences, respectively.

The time preference elicitation uses a multiple price list with 6 decisions to be made over two fixed periods of time, now vs two weeks later. Participants in the Decision Stage were asked to choose between option A and B where option A is to receive some sum of money today and option B is to receive a fixed 20 EUs in two weeks. The option A earnings start at 19 EU and monotonically decrease to 11 EU as shown in Table 2.

**Table 2.** Multiple Price List: Time Preference Elicitation

	Choice A	Choice B
1.	Today: 19 EU 2 Weeks: 0 EU	Today: 0 EU 2 Weeks: 20 EU
2.	Today: 18 EU 2 Weeks: 0 EU	Today: 0 EU 2 Weeks: 20 EU
3.	Today: 17 EU 2 Weeks: 0 EU	Today: 0 EU 2 Weeks: 20 EU
4.	Today: 16 EU 2 Weeks: 0 EU	Today: 0 EU 2 Weeks: 20 EU
5.	Today: 14 EU 2 Weeks: 0 EU	Today: 0 EU 2 Weeks: 20 EU
6.	Today: 11 EU 2 Weeks: 0 EU	Today: 0 EU 2 Weeks: 20 EU
Exchange: 1 EU = \$0.33		

The risk preference elicitation also used a multiple price list but with 10 decisions rather than 6. Participants in the Decision Stage were asked to choose from Lottery A and Lottery B. Each lottery has a "low payout" and a "high payout" The probability of receiving the high payout monotonically and linearly increases with each decision in the list. Lottery A is a "safer bet" with a larger low payout but smaller high payout and Lottery B is a "riskier bet" with a smaller low payout and a larger high payout which are outlined in Table 3. The set of probabilities in order of their presentation to the subjects is:  $(P_l, P_h) = \{(90, 10), (80, 20), (70, 30), (60, 40), (50, 50), (40, 60), (30, 70), (20, 80), (10, 90), (0, 100)\}$ , where  $P_h$  is the probability of receiving the high payout and  $P_l$  is the probability of receiving the low payout.

All subjects received \$2 for completing the stage successfully as a show up payment.

**Table 3.** Risk Preference Elicitation Payouts

	Lottery A	Lottery B
High Payout	20 EUs	38 EUs
Low Payout	16 EUs	2 EUs
Exchange: 1 EU = \$0.33		

Additional payments were determined by randomly selecting a "choice that counts". Participants were informed that there was an equal opportunity that any of their sixteen decisions to be drawn as the "choice that counts".

### 3.2 Prediction Stage

None of the subjects that participated in the Prediction Stage participated in the Decision Stage. Prediction stage subjects were asked to indicate their beliefs about what percentage of Decision Stage participants would choose a given option for each decision. For the time preference elicitation, predictors were submitting their prediction of how many deciders would take the option to wait two weeks for a higher payout in each decision, Choice B, and in the risk preference predictions, predictors were asked to submit what percentage of deciders they believed would take the "riskier" lottery, Lottery B.

Once predictors responded with their beliefs about the entire group of deciders in the 16 decisions, they were asked to submit their beliefs about 4 different demographic subgroups within the whole group: female deciders, male deciders, liberal deciders, and conservative deciders. This made for an additional 64 predictions for a total of 80 predictions. The order these subsequent predictions were made in was randomized by subgroup, maintaining the integrity of each MPL decision set.

All subjects received a \$5 participation payment for completing this stage as well as a chance at an additional lottery payment of \$5, incentivized by the accuracy of a randomly chosen "prediction that counts" using an adapted Binarized Scoring Rule (Burdea and Woon, 2022; Danz, Vesterlund and Wilson, 2022; Hossain and Okui, 2013).

## 4 Results

### 4.1 Decision Phase Outcomes

Phase 1 Deciders made all of their choices under incentive that one of their choices would be randomly chosen to actually impact their payment and therefore they should be properly motivated to act like every decision could potentially matter. Table 4 reports the percentage of deciders in each demographic subgroup, as well as the overall Phase 1 group, that opted for the "riskier" Lottery B in the risk preference multiple price list elicitation. Due to the monotonic nature of the success probabilities across the different risk choices, we would expect that the percentage of deciders choosing lottery B would also follow a monotonic nature. This is largely holds in all subgroups except in risk choice 1 for the female subgroup where one participant chooses Lottery B in choice 1 and then going back to A in following choices. There are not many statistically significant differences in the average decisions by choice between the demographic subgroups, though there are some consistent trends, such as liberal deciders tend to have a later switching point than conservative deciders, though not significantly as seen in Table A.1 of Appendix A. Similarly, the same table also shows that females tend to also have a later switching point, but not significantly and with less consistency that liberals vs conservatives, implying that female participants may be less risk tolerant than male participants, as with liberals versus conservatives, respectfully.

The deciders in the time preference choices do display the expected monotonic behavior expected in the given decision environment as shows in Table 5. This monotonicity is consistent across all subgroups as well as the decider group as a whole. Similar to the risk decisions by Phase 1 participants, there is not a statistically significant difference in decision making between the difference subgroups with respect to their time decisions which can be seen in Table A.2<sup>3</sup>.

---

<sup>3</sup>The lack of statistical significance in the subgroup differences is not meant to imply a disagreement with previous literature regarding gender differences in risk and time preferences. The goal of this paper was not to prove a gender difference but to have a a group from the same population that the Phase 2 predictors were from so they can make predictions. This paper is agnostic towards the gender and political differences

**Table 4.** Percent of Risk Deciders Choosing Lottery B

Choice	All	Female	Male	Liberal	Conservative
1	1.82	3.03	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	18.18	15.15	22.73	10.53	31.58
5	54.55	51.52	59.09	47.37	63.16
6	65.45	60.61	72.73	52.63	73.68
7	74.55	69.70	81.82	57.89	89.47
8	87.27	81.82	95.45	78.95	100
9	94.55	90.91	100	89.47	100
10	98.18	96.97	100	100	100
n	55	33	22	29	19

**Table 5.** Percent of Time Deciders Choosing Option B

Choice	All	Female	Male	Liberal	Conservative
1	40	36.36	45.45	26.32	36.84
2	52.73	54.55	50	47.37	47.37
3	65.45	66.67	63.64	63.16	63.16
4	72.73	75.76	68.18	73.68	73.68
5	89.09	81.82	100	89.47	94.74
6	92.73	87.88	100	94.74	94.74
n	55	33	22	29	19

## 4.2 Prediction Error by Subgroups

The main question when looking at the prediction outcomes of the predictors and their relative accuracy was focused on how group predictions compared to whole sample predictions. In order to achieve this, the ratio of errors between the group predictions and whole sample predictions are calculated and implemented a specification that accounts for the subgroup a predictor belongs to and which subgroup they are making predictions about. This is the described with the simple specification as follows:

---

in risk and time preferences

$$\varepsilon_{c,g}^r = \hat{P}_{c,g}^r - P_{c,g}^r \quad (1)$$

$$ER_c = \left| \frac{\varepsilon_{c,g}}{\varepsilon_{c,0}} \right| = \beta_0 + \beta_1 female_i + \beta_2 GuessingFem_i + \beta_3 female_i \times GuessingFem_i + \mu_i \quad (2)$$

where  $\varepsilon_{c,g}^r$  refers to the risk preferences ( $r$ ) prediction error for choice  $c$  within subgroup  $g$ .  $g = 0$  indicates the entire sample, rather than a subgroup. The indicator variables of *liberal* and *conservative* are demographic variables according to how predictors self-identified their political preferences on the following 5 point scale:

1. Strongly Conservative
2. Moderately Conservative
3. Moderate
4. Moderately Liberal
5. Strongly Liberal

Those who identified with a 1 or 2 were coded as "conservative" and those who identified with 4 or 5 were coded as "liberal". Those who identified with 3 were coded as "moderate".

Approaching the analysis of the subgroup prediction errors for the multiple price lists presents an option to implement the data in a pooled manner. However, this would not be as informative to the applicable intention of the problem. When individuals have to make predictions about the risk or time preferences of somebody else, they are usually doing so in a partially informed scenario, where they are roughly aware of the probability of a favorable outcome for their opponent. This means that real life predictions are not made about the entire distribution of someone's risk or time preferences but at specific points along that distribution in any given moment. An illustrating example is in gambling such

as a poker where there is some amount of chance involved. A poker player may need to attempt predicting their opponents willingness to take risk based on what cards are already on the table. Since some cards are already displayed and the player has cards in their hand, they can roughly estimate the probability of a favorable outcome for their opponent if they make a risky play such as bluffing or going all in, which is necessary for the opponent to inform their decision on how to play a hand. They would not make a prediction considering all of the potential points along the probability distribution of a favorable outcome because it is not necessary to solve the their problem in the moment. To make this analysis more applicable to actual strategic situations, the analysis will mainly continue by looking at the individual choices in each MPL with pooled results being primarily used for illustration of distributional differences.

#### 4.2.1 Gender Subgroups

The results in Table 6 show us how an individual's error in predictions changes based on additional information and whether that individual is part of the same gender group as who they are making predictions about. These coefficients report changes to a ratio comparing the error in risk prediction of a certain subgroup divided by an individual's error in prediction about the entire sample of Deciders. In the following tables,  $Guess_f$  is an indicator variable signaling whether a predictor is making a prediction about a group of female deciders or male if ( $Guess_f = 0$ ). This interaction model allows us to understand the error ratio of four groups: Men predicting men, men predicting women, women predicting men, and women predicting women. Specifically, this means that the constant value in this output represents the ratio errors for men making predictions about other men. For example, in  $ER_1$ , the constant is indicated as 1.968. This is interpreted as the error that men make when making predictions about other men is almost double the error than when they are making predictions about the general group. There are not any significant differences in this ratio for the other three prediction groups, but if we look at  $ER_2, ER_3, ER_4$ , there is a case to be made that men

**Table 6.** Risk Prediction Error Ratio ( $ER_c$ ) by Gender

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$	(7) $ER_7$	(8) $ER_8$	(9) $ER_9$	(10) $ER_{10}$
female	-0.0999 (0.772)	0.260 (0.161)	-0.0108 (0.270)	0.518 (0.651)	-0.171 (0.564)	-1.021 (0.743)	-1.321 (1.131)	-0.126 (0.956)	0.639 (1.051)	-2.054 (2.313)
$Guess_f$	-0.949 (0.644)	-0.288*** (0.0686)	-0.303*** (0.0758)	-0.366* (0.190)	-0.144 (0.308)	-0.384 (0.683)	-0.405 (0.836)	-0.204 (0.417)	-0.763 (0.582)	1.869 (1.693)
female $\times$ $Guess_f$	0.0750 (0.732)	-0.0946 (0.144)	-0.0276 (0.113)	-0.174 (0.607)	0.0776 (0.365)	0.136 (0.723)	-0.0456 (0.939)	-0.214 (0.531)	-0.0519 (0.732)	-0.935 (1.995)
Constant	1.968*** (0.678)	1.060*** (0.0913)	1.357*** (0.244)	1.353*** (0.220)	1.727*** (0.535)	2.991*** (0.688)	3.994*** (1.010)	2.732*** (0.899)	2.414*** (0.596)	5.086*** (2.070)
Observations	294	303	307	291	302	300	299	308	290	235 <sup>a</sup>
R-squared	0.021	0.045	0.016	0.008	0.001	0.018	0.015	0.002	0.007	0.014

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>a</sup>Due to the set up in the survey, there are missing observations from the last prediction, which was adjacent to the "next" button in the experimental survey



making predictions about women is actually more accurate than when making predictions within their own group<sup>4</sup>. That is, men are better off in general with less information than when they try to account for additional gender information if their opponent is another man but if it is a woman, they can make a slightly better prediction than if they did not have any information. This is accounted for by using the  $Guess_f$  coefficient of  $-0.288$  in  $ER_2$ , for example, and adding it to the constant  $1.060$ , yielding  $0.772$ . This is interpreted by saying that when a man makes a prediction about a women, the expected error in that prediction is  $0.772$ , or  $77.2\%$ , of the error made when making a prediction with no information about gender.  $ER_4$  has a similar result with an informed ratio of  $0.988$ , while  $ER_3$  has an informed ratio just over 1. In all 10 choice outcomes, men do worse when they have gender information about another man<sup>5</sup>.

**Table 7.** Time Prediction Error Ratio by Gender (In Group/Out Group)

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$
female	0.156 (0.188)	0.479 (0.457)	0.764*** (0.252)	0.503 (0.317)	0.899* (0.516)	0.688 (0.776)
$Guess_f$	-0.216 (0.135)	0.101 (0.174)	0.0511 (0.128)	0.254 (0.235)	-0.446*** (0.134)	-0.953* (0.555)
female $\times$ $Guess_f$	-0.129 (0.171)	-0.320 (0.404)	-0.309 (0.189)	-0.501* (0.298)	-0.950** (0.382)	-0.528 (0.708)
Constant	1.259*** (0.133)	1.185*** (0.223)	1.078*** (0.0758)	1.218*** (0.155)	1.644*** (0.291)	2.076*** (0.595)
Observations	286	301	305	304	305	307
R-squared	0.025	0.004	0.026	0.006	0.040	0.037

Robust standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Gendered groups do not have as clear of a difference in time prediction error ratios com-

<sup>4</sup>The robustness model in presented in Table C.1 of Appendix E shows a similar trend where women making predictions about women are more accurate when compared to men making predictions about men in the choices that have a lower probability of a favorable outcome.

<sup>5</sup>There is a notable change in observations for  $ER_{10}$  of Table 8 due to the high number of perfect predictions about this particular risk choice. This is expected considering that the probability of a favorable result at this point of the distribution is  $P = 1$ .

pared to risk prediction. Table 7 shows that there is not as clear of a trend in any particular part of the distribution regarding improved accuracy for one group nor an advantage in male vs female predictors for improved relative in-group accuracy. The constants are informative and highly significant though and tell a similar story as the constants in Table 6. If we look at the constant for time choice 6 (2.076), then we see that when men are making predictions about other men, the average error in their predictions is actually twice as much as if they just made a prediction without any gendered information. In general, this holds across the other prediction matching with few coefficients showing there is much of statistically significant difference. Primarily, choice 5 has an interaction coefficient of -0.950 and discount when making predictions about females, regardless of the predictor’s gender of -0.446.

Political preferences are not considered to be immutable characteristics (Enriquez, 2013; Hoffman, 2010; Sen and Wasow, 2016) so it is reasonable that there may not be as strong of trends for prediction error analysis based on political information compared to gender information. Table 8 does give a strong baseline ratio on the value of learning political information about someone when making predictions about their risk preferences. The interpretation of these coefficients is similar to Table 6 but for illustration, the constant in the estimation for choice 1 (1.947) indicates that the error in predictions when conservative have information that their opponent is a conservative is nearly twice as large as when they do not have any information at all. There is no other significant result for the other guessing group match ups, indicating that across all groups, they are twice as inaccurate once they have gender information compared to when they do not and are making predictions about a generic Decider. On the other extreme, Choice 10 shows that individuals with political information have errors five times as large compared to when they have no information and are making a general prediction.<sup>6</sup>

Table 9 depicts the results that examine prediction errors based on politically identifying

---

<sup>6</sup>Just as in Table 6, there is a notable change in observations for  $ER_{10}$  of Table 8 due to the high number of perfect predictions about this particular risk choice. This is expected considering that the probability of a favorable result at this point of the distribution is  $P = 1$ .

**Table 8.** Risk Prediction Error Ratio by Politics

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$	(7) $ER_7$	(8) $ER_8$	(9) $ER_9$	(10) $ER_{10}$
liberal	-0.136 (0.803)	0.531 (0.428)	0.361 (0.352)	1.030 (1.390)	0.683* (0.396)	-0.671 (0.796)	0.765 (0.823)	2.274*** (0.863)	-0.180 (0.603)	-0.179 (3.334)
$Guess_l$	-0.447 (0.558)	0.00761 (0.108)	0.143 (0.153)	0.148 (0.351)	-0.0247 (0.275)	0.194 (0.593)	0.817 (0.590)	0.171 (0.334)	1.258 (1.339)	-1.236 (1.111)
liberal $\times$ $Guess_l$	-0.180 (0.790)	-0.490 (0.410)	-0.510 (0.354)	-1.452 (1.023)	0.00101 (0.356)	0.0536 (0.655)	-0.795 (0.905)	-1.012 (0.870)	-1.524 (1.386)	-0.292 (1.893)
Constant	1.947*** (0.546)	1.015*** (0.106)	1.052*** (0.0978)	2.157*** (0.495)	1.300*** (0.211)	2.232*** (0.754)	2.006*** (0.515)	1.576*** (0.280)	2.100*** (0.485)	5.296*** (2.243)
Observations	197	194	198	198	193	195	193	198	186	155 <sup>a</sup>
R-squared	0.010	0.018	0.011	0.008	0.030	0.009	0.004	0.046	0.017	0.003

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>a</sup>Due to the set up in the survey, there are missing observations from the last prediction, which was adjacent to the "next" button in the experimental survey

**Table 9.** Time Prediction Error Ratio by Politics

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$
liberal	-0.0550 (0.189)	-0.00186 (0.267)	-0.944 (0.906)	-0.0931 (0.222)	0.140 (0.246)	0.333 (0.572)
$Guess_l$	0.217 (0.158)	-0.0447 (0.215)	-0.474 (0.780)	0.818** (0.330)	0.541 (0.327)	2.365* (1.267)
liberal $\times$ $Guess_l$	-0.250 (0.211)	0.279 (0.264)	0.402 (0.785)	-0.773** (0.353)	-0.800** (0.337)	-2.903** (1.325)
Constant	1.118*** (0.163)	1.132*** (0.195)	1.998** (0.892)	1.204*** (0.175)	1.141*** (0.146)	1.495*** (0.315)
Observations	190	195	198	196	198	200
R-squared	0.010	0.006	0.018	0.057	0.023	0.043

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

information compared to general prediction about the entire Decider sample group. The results in this estimation show a similar result as the other time and risk predictions already presented, at a base point, across prediction subgroups, there are smaller prediction errors when individuals are making predictions about a general Decider without any politically identifying information. The main difference with these results is that the interaction between the indicator term for a liberal and the indicator for whether someone is making a prediction about a liberal shows an improvement on the constant. In other words, when a conservative is making a prediction about another conservative, their error ratio on average would be 1.204 in Choice 4. When a liberal is making predictions about another liberal, the value of the error decreases by 0.773. Another way to illustrate this is going group by group where conservatives predicting conservatives have an error ratio of 1.204, there is no significant difference when liberals are making predictions about conservatives, conservatives making predictions about liberals are 0.818 worse or 2.022 meaning a twice as high error compared to a general and uninformed prediction, and a liberal making a prediction about another liberal is  $1.204 + 0.818 + (-0.773) = 1.249$ . This means that a liberal making a

prediction about another liberal carries a 24.9% larger error than an uninformed prediction.

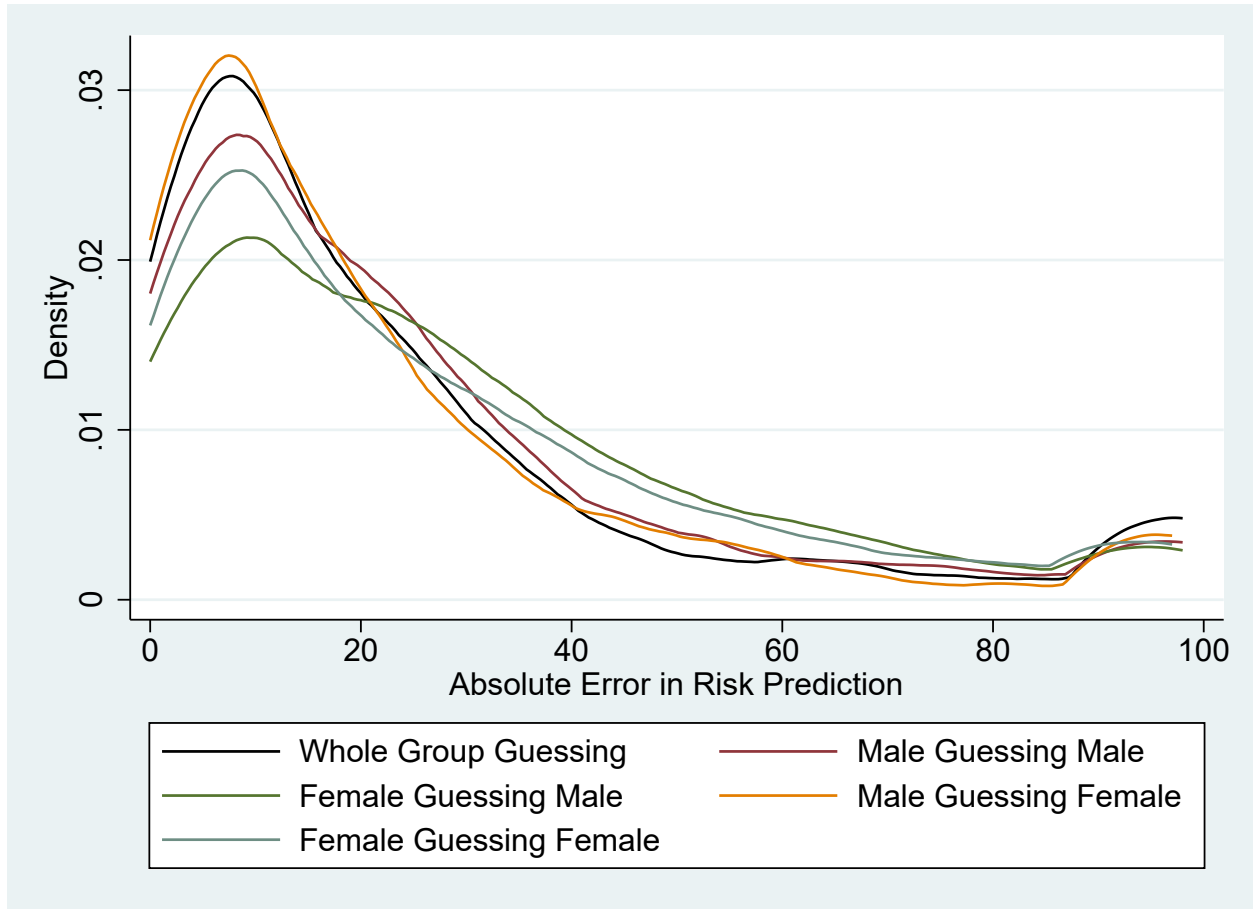
#### 4.2.2 Pooled Distributions and Fixed Effects Model

The pooled error distributions for the four different predictions made are not as helpful in understanding the specific situations that individuals encounter when applying predictive behaviors in the real world but do help illustrate the differences in how different subgroup prediction scenarios are in general across a range of probabilities for a favorable outcome by the decider that the predictions are being made about. Figures 1, 2, 3, 4 give a kernel density smoothing of a pooled histogram depicting the distribution of absolute errors for each of the four predictions types (risk/time vs gender/politic). The interpretation of these plots should consider that these are pooled across the entire distribution of errors, meaning that if a certain prediction matching is of better quality than another, we would expect to see a higher density in the lower regions of the plot, where absolute value is less, and a lower density when absolute value is higher. This gives us the ability to somewhat rank the effectiveness of different prediction matching in reducing prediction errors. Whole group guessing is the uninformed guesses that predictors make without any identifying gender or political information and can be used as a reference for how well information benefits certain group match ups.

In Figure 1 it is clear which matching has the worst outcomes with respect to minimizing absolute error due to female guessing female having the lowest density of low errors and the highest density of high errors other than at the extreme high end. This is complemented by male guessing female having the highest density of low errors with no clear distinction past errors at a level of 20 or higher.

Figure 2 shows a clear advantage in women making predictions about other women with regards to density of small errors versus a primarily low density of large error relative to the other matchings. This is particularly true when comparing to the uninformed guessing baseline and provide further evidence that this improves the outcome in reducing error. The

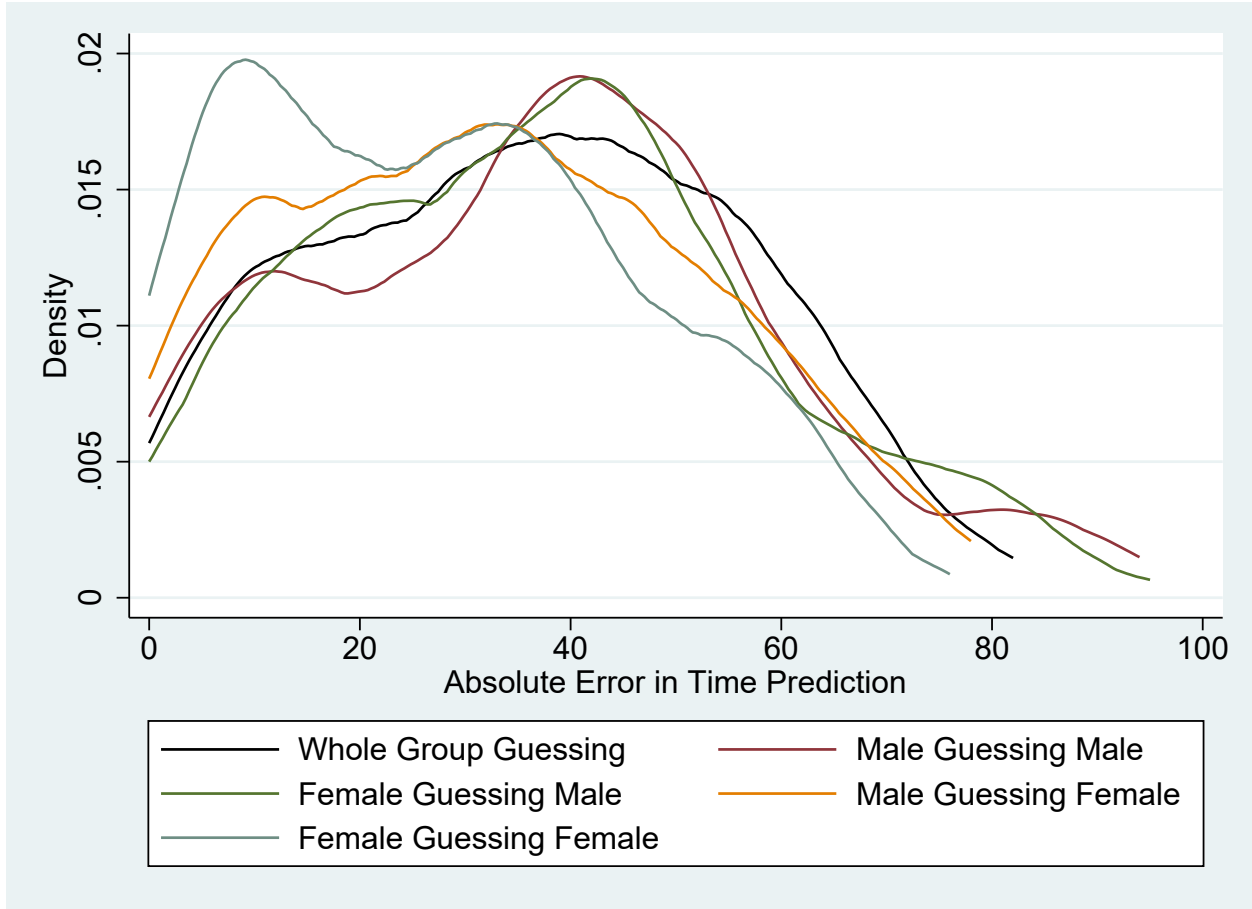
**Figure 1.** Smoothed Pooled Distribution of Risk Prediction Errors across Genders



other result that is identifiable here is that female guessing men appears to skew towards having higher errors over all and men guessing men are relatively poor at accuracy, though not quite as bad.

Figure 3 depicts the pooled distribution of risk prediction errors by political groups. This plot shows a very tight band between all four prediction matches and the baseline, showing that there is not much difference in what the match up, additional information does not appear to help improve the accuracy of predictions. This means that someone making a prediction is no better off with additional political information. The pooled distribution of time prediction errors by politics on the other hand has relatively well defined differences in effectiveness. Namely, we see a clear dominance of conservative guessing about other conservatives and inferiority of liberals making predictions about other liberals. The distri-

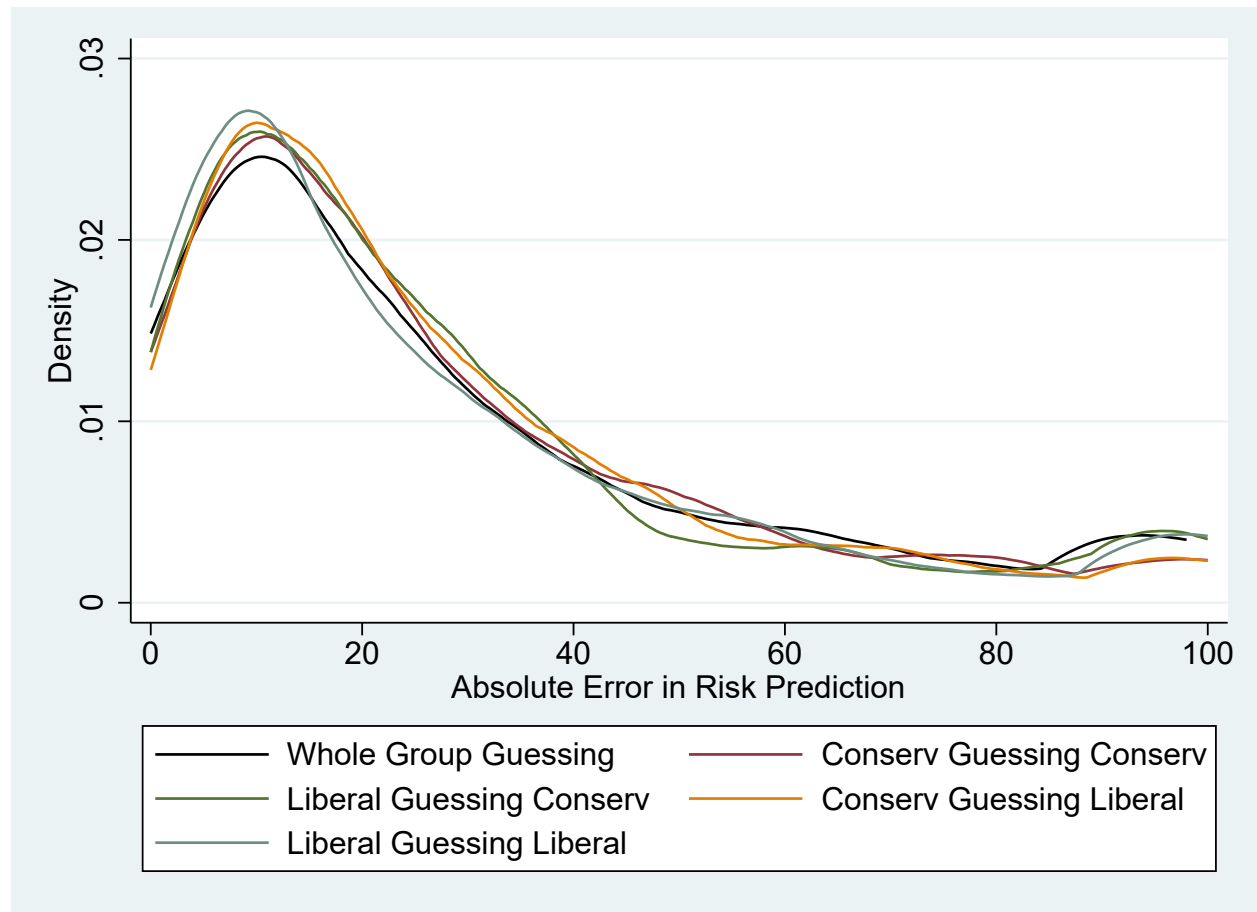
**Figure 2.** Smoothed Pooled Distribution of Time Prediction Errors across Genders



butional switching point from low errors to high errors occurs just under an absolute error of 30, meaning that below 30 there is a higher density among the more effective prediction matches and above 30, the ordinal rankings invert.

The results of Table 10 show the effects on the respective subgroup:general error ratios that gender and political alignment show. These models take into consideration choice fixed effects regarding the multiple price list choice predictions that the phase two predictors made. Model (1) looks at the gender based risk prediction error ratio. While neither the gender nor the political identity coefficients show any significant results, we do see that there is consistency in the coefficient signs with the trends seen in Table 6. The interpretation is that there is a suggestion uninformed predictions have a lower error than when there is gender information, where the error is over double. Model (2) also has a consistent with the

**Figure 3.** Smoothed Pooled Distribution of Risk Prediction Errors across Political Identities

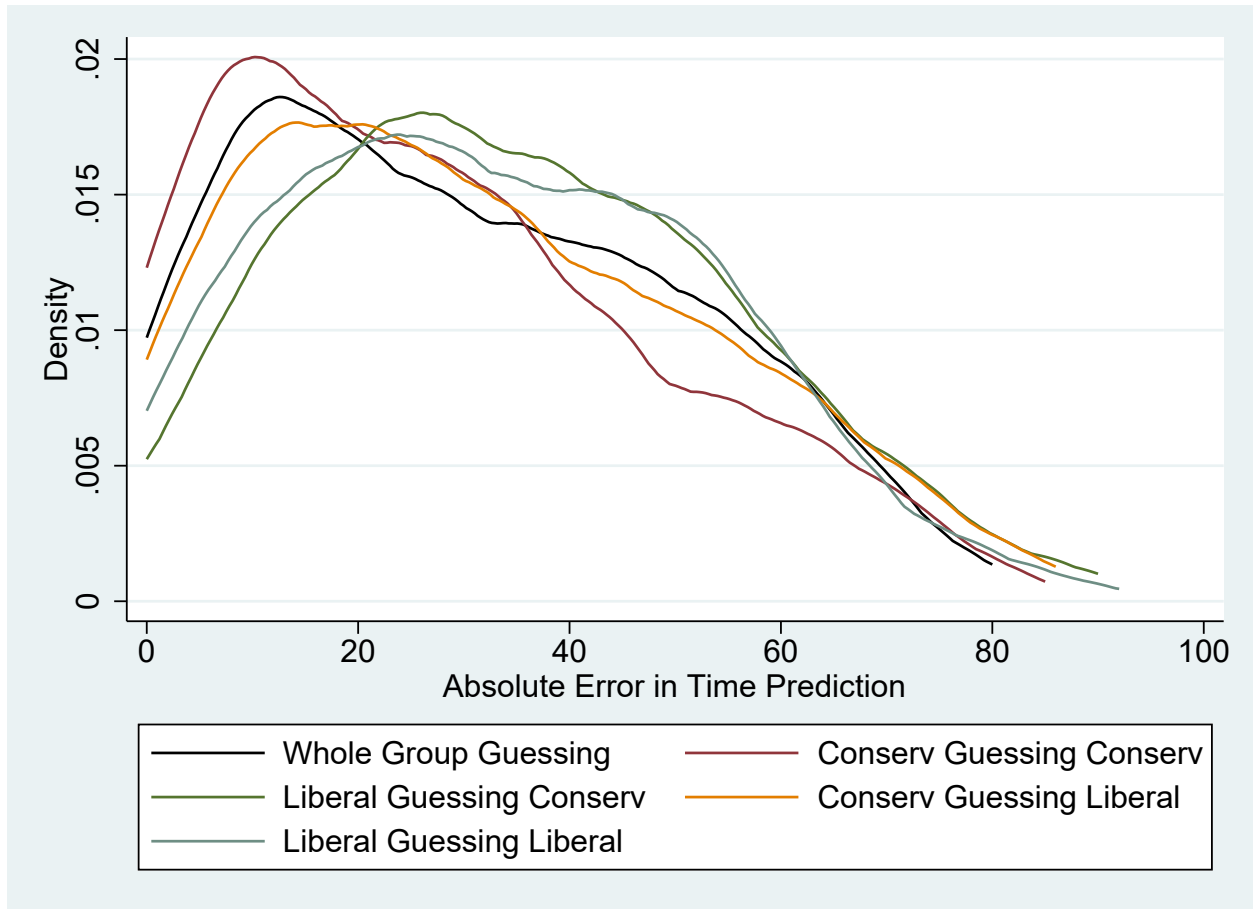


relevant non-fixed effects model from Table 8 with respect to how baseline uninformed error ratios compare to subgroup prediction match ups where predictors are more accurate when they have less information rather than more for men making predictions about other men. When the predictor is a woman making predictions about other men, they are actually less accurate by almost 40% larger group based error. Men making predictions about women is no different than when they make predictions about men but finally, when women make predictions about other women, they have a ratio of approximately 1.5, meaning they are slightly less accurate with more information than when men are predicting about other men.

Model (3) depicts a near double error in political group level predictions compared to uninformed whole sample predictions when making predictions about risk preferences across all guessing groups, similar to what we saw in Table 8 while model (4) shows group level



**Figure 4.** Smoothed Pooled Distribution of Time Prediction Errors across Political Identities



prediction differences. More specifically, conservatives making predictions about other conservatives have an error ratio of 1.35, which is not significantly different than liberals making the predictions about conservatives. Predictions about liberals by conservatives though have an even higher error rate, nearly double the error compared to having no information about the decider being liberal, and when a liberal makes a prediction about another liberal, they are actually more accurate on average than when they have no information with an error only two-thirds that of their uninformed error.<sup>7</sup>

<sup>7</sup>As described in Section 3, the moderates are left out of this model due to the within subject design not asking subjects about their preference beliefs about moderates.

**Table 10.** Fixed Effects Error Ratio Models

VARIABLES	(1) Risk Gender	(2) Time Gender	(3) Risk Politics	(4) Risk Politics
female	-0.209 (0.331)	0.588** (0.234)		
$Guess_f$	-0.0380 (0.299)	-0.201 (0.188)		
female $\times$ $Guess_f$	-0.191 (0.367)	-0.461** (0.233)		
liberal			0.465 (0.424)	-0.0931 (0.259)
$Guess_l$			0.134 (0.349)	0.574*** (0.216)
liberal $\times$ $Guess_l$			-0.636 (0.470)	-0.686** (0.291)
Constant	2.197*** (0.268)	1.408*** (0.189)	1.998*** (0.314)	1.347*** (0.192)
Observations	2,929	1,808	1,907	1,177
Number of id	155	154	100	100

Standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

## 5 Conclusions

This paper presented an experimental design that captured beliefs about time and risk preferences in an incentivized prediction experiment. The initial preferences were elicited in a first stage of Deciders who were incentivized to participate in a modified double multiple price list elicitation activity. In addition to eliciting beliefs about time and risk preferences, the within subject design also allowed for estimation of in-group and out-group biases by gender and political identity. Moreover, the difference in predictions by groups allow for estimation of perceived bias by different subgroups about other subgroups. All of the predictions were made in the second stage of the experiment by predictors.

The first phase elicitation of decider risk and time preferences resulted in some nominal differences between identity group preferences but not statistically significant differences.

Previous literature has been split between finding females have a lower risk preference and men and women have equal risk tolerances. Since the purpose of this project was to gain an understanding about how much identity information improves the predictions about risk and time preferences and informs biases, the actual difference between the identity groups is not the primary concern and is measured in order to gauge accuracy of the phase 2 predictions. The preference elicitation did reveal mostly rational behavior in the multiple price list decisions. The primary indicator of this is the mostly monotonic trends in the proportion of participants choosing the riskier and more patient option in the risk and time preference elicitations respectively. The only exception to this is one observation<sup>8</sup> having two switching points to lottery B. The time preference elicitation had monotonically increasing trends across all subgroups and in aggregate.

The results from the subgroup error analysis did not show significant differences for many of the choice options but there were some interesting results that came out of this primary analysis. In particular, the risk preference prediction analysis by gender identity groups showed a significant difference in the improved accuracy bonus that was realized when participants made in-group predictions vs out-group predictions. In the first four choices of the risk preference predictions, female participants enjoyed improved accuracy when making predictions about other women compared to men making predictions about other men from Table C.1. The interpretation of this is that women are more in touch with how other women feel about risk than men are about other men when compared to when men and women make predictions about the opposite gender. This improved perception of risk preference is specific to when the odds of a good outcome are relatively low. A real life application is when two individuals are engaged in some sort of strategy based competition such as athletics, chess, card game, etc, and two individuals do not know anything about each other other than gender identity. When the two individuals are both women, they have a relatively more

---

<sup>8</sup>The one participant was a female observation that chose Lottery B in Choice 1 and then switched back to Lottery A in Choice 2. This can either be attributed to the individual not understanding the instruction initially but is more likely exploring the option space.

accurate perception of each other's risk tolerances than if it was a woman trying to make a judgement about an opponent who is a man. This advantage does not carry over well for men judging other men vs if they are judging woman however nor does this advantage hold as the probability of a favorable outcome in the lottery choices increases towards near certainty of preferential payout.

Politically based advantages in the in:out group error ratios were not consistent across either risk or time preference predictions. There is some indication at the lower risk levels (choices 7-10) that liberals are better at intuiting the risk tolerances of other liberals compared to their predictions about conservatives than conservatives are about making in-group predictions vs out-group, but these results only hold statistical significance in choices 8 and 10, where choice 10 is a corner solution <sup>9</sup>. Over the time preference predictions, there was no significant difference between conservative or liberal in-group accuracy advantages.

Testing for differences in the way subgroups made predictions about other subgroups showed only slight differences. There is not a noticeable difference in how subgroups perceive each other vs within their own group with regards to risk preferences. There is a difference in how the different genders perceive the difference between in time preferences however. Women in the predicting sample tended to believe that the difference between men and women was smaller than men believed. Women also had a smaller gap in their predictions regarding conservatives and liberals. This means that men display larger biases about the difference in the patience of subgroups than women.

Overall, the effects of knowing gender or political identity about an individual in a risk or time preference scenario does not have a very strong impact on the accuracy of predictions about the deciding individual. While there is not a universal effect for all types of subgroup identities or expected value of a risky decision, there are some more specific inferences that can be made and some marginal advantages to making predictions within identity groups.

---

<sup>9</sup>Choice 10 in the risk preference elicitation has a probability of receiving the higher payout of  $p = 1$ , therefore all rational actors in the experiment should choose the "riskier" lottery. Predictors are aware of this probability level

There is further room to explore these types of identity based prediction accuracy problems, specifically looking into other types of identities that have large cultural impacts such as race, ethnicity, and and regional identities.

## References

- Andersen, Steffen, Harrison, Glenn W, Lau, Morten Igel and Rutström, E Elisabet.** (2006). ‘Elicitation using multiple price list formats’, *Experimental Economics* 9(4), 383–405.
- Andersen, Steffen, Harrison, Glenn W, Lau, Morten I and Rutström, E Elisabet.** (2008). ‘Eliciting risk and time preferences’, *Econometrica* 76(3), 583–618.
- Andreoni, James and Sprenger, Charles.** (2012). ‘Risk preferences are not time preferences’, *American Economic Review* 102(7), 3357–76.
- Azrieli, Yaron, Chambers, Christopher P and Healy, Paul J.** (2018). ‘Incentives in experiments: A theoretical analysis’, *Journal of Political Economy* 126(4), 1472–1503.
- Burdea, Valeria and Woon, Jonathan.** (2022). ‘Online belief elicitation methods’, *Journal of Economic Psychology* p. 102496.
- Charness, Gary and Gneezy, Uri.** (2012). ‘Strong evidence for gender differences in risk taking’, *Journal of Economic Behavior & Organization* 83(1), 50–58.
- Charness, Gary, Gneezy, Uri and Imas, Alex.** (2013). ‘Experimental methods: Eliciting risk preferences’, *Journal of economic behavior & organization* 87, 43–51.
- Danz, David, Vesterlund, Lise and Wilson, Alistair J.** (2022). ‘Belief elicitation and behavioral incentive compatibility’, *American Economic Review* .
- Dave, Chetan, Eckel, Catherine C, Johnson, Cathleen A and Rojas, Christian.** (2010). ‘Eliciting risk preferences: When is simple better?’, *Journal of Risk and Uncertainty* 41(3), 219–243.
- Dohmen, Thomas, Falk, Armin, Huffman, David and Sunde, Uwe.** (2010). ‘Are risk aversion and impatience related to cognitive ability?’, *American Economic Review* 100(3), 1238–60.
- Drichoutis, Andreas C and Lusk, Jayson L.** (2016). ‘What can multiple price lists really tell us about risk preferences?’, *Journal of Risk and Uncertainty* 53(2), 89–106.
- Eckel, Catherine C and Grossman, Philip J.** (2002). ‘Sex differences and statistical stereotyping in attitudes toward financial risk’, *Evolution and human behavior* 23(4), 281–295.
- Enriquez, Anthony R.** (2013). ‘Assuming Responsibility for Who You Are: The Right to Choose Immutable Identity Characteristics’, *NYUL Rev.* 88, 373.
- Fecteau, Shirley, Pascual-Leone, Alvaro, Zald, David H, Liguori, Paola, Théoret, Hugo, Boggio, Paulo S and Fregni, Felipe.** (2007). ‘Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making’, *Journal of Neuroscience* 27(23), 6212–6218.

- Gneezy, Uri, Leonard, Kenneth L and List, John A.** (2009). ‘Gender differences in competition: Evidence from a matrilineal and a patriarchal society’, *Econometrica* 77(5), 1637–1664.
- Gneezy, Uri and Potters, Jan.** (1997). ‘An experiment on risk taking and evaluation periods’, *The quarterly journal of economics* 112(2), 631–645.
- Hoffman, Sharona.** (2010). ‘The importance of immutability in employment discrimination law’, *Wm. & Mary L. Rev.* 52, 1483.
- Holt, Charles A and Laury, Susan K.** (2002). ‘Risk aversion and incentive effects’, *American economic review* 92(5), 1644–1655.
- Holt, Charles A and Laury, Susan K.** (2005). ‘Risk aversion and incentive effects: New data without order effects’, *American Economic Review* 95(3), 902–912.
- Hossain, Tanjim and Okui, Ryo.** (2013). ‘The binarized scoring rule’, *Review of Economic Studies* 80(3), 984–1001.
- Hunt, Melissa K, Hopko, Derek R, Bare, Robert, Lejuez, CW and Robinson, EV.** (2005). ‘Construct validity of the balloon analog risk task (BART) associations with psychopathy and impulsivity’, *Assessment* 12(4), 416–428.
- Jacobson, Sarah and Petrie, Ragan.** (2009). ‘Learning from mistakes: What do inconsistent choices over risk tell us?’, *Journal of risk and uncertainty* 38(2), 143–158.
- Laury, Susan.** (2005). ‘Pay one or pay all: Random selection of one choice for payment’, *Andrew Young School of Policy Studies Research Paper Series* (06-13).
- Lejuez, Carl W, Read, Jennifer P, Kahler, Christopher W, Richards, Jerry B, Ramsey, Susan E, Stuart, Gregory L, Strong, David R and Brown, Richard A.** (2002). ‘Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART).’, *Journal of Experimental Psychology: Applied* 8(2), 75.
- Sen, Maya and Wasow, Omar.** (2016). ‘Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics’, *Annual Review of Political Science* 19, 499–522.
- Weber, Elke U, Blais, Ann-Renee and Betz, Nancy E.** (2002). ‘A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors’, *Journal of behavioral decision making* 15(4), 263–290.
- Zuckerman, Marvin.** (1994), *Behavioral expressions and biosocial bases of sensation seeking*, Cambridge university press.

## A Appendix A: Phase 1 Logit Models

**Table A.1.** Logit on Risk Choices by Gender and Political Identity

VARIABLES	risk4	risk5	risk6	risk7	risk8	risk9
female	-0.369 (0.834)	0.248 (0.688)	-0.109 (0.719)	-0.332 (0.844)	-0.405 (1.278)	
liberal	-1.314 (0.904)	-0.686 (0.674)	-0.908 (0.703)	-1.778** (0.886)		
Constant	-0.585 (0.643)	0.410 (0.592)	1.087* (0.649)	2.325*** (0.900)	1.609 (1.095)	1.705** (0.769)
Observations	38	38	38	38	19	13

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note: Risk Choices 1, 2, 3, and 10 omitted due to no variance in outcome

**Table A.2.** Logit on Time Choices by Gender and Political Identity

VARIABLES	time1	time2	time3	time4	time5	time6
female	-0.574 (0.719)	0.0478 (0.674)	-0.258 (0.704)	0.604 (0.757)		
liberal	-0.406 (0.718)	-0.00754 (0.658)	0.0405 (0.683)	-0.0988 (0.754)	-0.492 (1.305)	0.288 (1.481)
Constant	-0.247 (0.595)	-0.131 (0.581)	0.677 (0.611)	0.733 (0.628)	2.197** (1.054)	2.197** (1.054)
Observations	38	38	38	38	23	23

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1



## B Appendix B: Prediction Biases between Groups

### B.1 Differences in Predictions by Gender

These results analyze how much of a difference there is in the beliefs about two subgroups' preferences. Higher differences suggest increased beliefs that the demographic subgroups behave differently with respect to their time or risk preferences. These differences are calculated by simply taking

$$\Delta \hat{P}_c^r = \hat{P}_{c,g}^r - \hat{P}_{c,g^{-1}}^r \quad (3)$$

or

$$\Delta \hat{P}_c^t = \hat{P}_{c,g}^t - \hat{P}_{c,g^{-1}}^t \quad (4)$$

where  $r$  represents a risk preferences and  $t$  represents time preferences, and then standardizing each difference within each respective choice.

### B.2 Differences in Predictions by Politics

These results analyze how much of a difference there is in the beliefs about two subgroups' preferences. Higher differences suggest increased beliefs that the demographic subgroups behave differently with respect to their time or risk preferences. The variable coding follows from earlier sections measuring political differences.

**Table B.1.** Standardized Difference in Risk Predictions between Gender Groups

VARIABLES	(1) $P_{\Delta g 1}$	(2) $P_{\Delta g 2}$	(3) $P_{\Delta g 3}$	(4) $P_{\Delta g 4}$	(5) $P_{\Delta g 5}$	(6) $P_{\Delta g 6}$	(7) $P_{\Delta g 7}$	(8) $P_{\Delta g 8}$	(9) $P_{\Delta g 9}$	(10) $P_{\Delta g 10}$
female	-0.0659 (0.177)	-0.105 (0.172)	-0.0939 (0.172)	0.0330 (0.174)	0.0350 (0.170)	-0.0755 (0.170)	-0.0107 (0.170)	-0.101 (0.170)	-0.331* (0.168)	0.0333 (0.202)
Constant	0.0443 (0.145)	0.0695 (0.140)	0.0624 (0.140)	-0.0210 (0.139)	-0.0230 (0.138)	0.0497 (0.138)	0.00702 (0.138)	0.0664 (0.138)	0.218 (0.136)	-0.0229 (0.167)
Observations	146	152	152	143	155	155	154	155	155	115
R-squared	0.001	0.002	0.002	0.000	0.000	0.001	0.000	0.002	0.025	0.000

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.2.** Standardized Difference in Time Predictions between Gender Groups

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta\hat{P}_1^r$	$\Delta\hat{P}_2^r$	$\Delta\hat{P}_3^r$	$\Delta\hat{P}_4^r$	$\Delta\hat{P}_5^r$	$\Delta\hat{P}_6^r$
female	0.171 (0.176)	0.168 (0.172)	0.237 (0.170)	0.281* (0.169)	0.485*** (0.166)	0.353** (0.168)
Constant	-0.115 (0.144)	-0.111 (0.140)	-0.156 (0.138)	-0.184 (0.137)	-0.318** (0.134)	-0.231* (0.136)
Observations	146	150	152	153	154	154
R-squared	0.006	0.006	0.013	0.018	0.053	0.028

Standard errors in parentheses  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.3.** Standardized Difference in Risk Predictions between Political Groups

VARIABLES	(1) $P_{\Delta c1}$	(2) $P_{\Delta c2}$	(3) $P_{\Delta c3}$	(4) $P_{\Delta c4}$	(5) $P_{\Delta c5}$	(6) $P_{\Delta c6}$	(7) $P_{\Delta c7}$	(8) $P_{\Delta c8}$	(9) $P_{\Delta c9}$	(10) $P_{\Delta c10}$
liberal	0.0851 (0.217)	0.215 (0.225)	-0.0168 (0.225)	0.0341 (0.227)	-0.0575 (0.229)	-0.121 (0.222)	-0.289 (0.219)	-0.00133 (0.222)	0.0601 (0.210)	-0.531* (0.274)
Constant	-0.114 (0.163)	-0.135 (0.168)	-0.000985 (0.167)	-0.0318 (0.169)	0.00867 (0.171)	0.0452 (0.165)	0.106 (0.162)	-0.0403 (0.165)	-0.0424 (0.156)	0.329 (0.209)
Observations	98	99	98	100	99	99	99	100	100	77
R-squared	0.002	0.009	0.000	0.000	0.001	0.003	0.018	0.000	0.001	0.048

Standard errors in parentheses  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.4.** Standardized Difference in Time Predictions between Political Groups

VARIABLES	(1) $P_{\Delta lc1}$	(2) $P_{\Delta lc2}$	(3) $P_{\Delta lc3}$	(4) $P_{\Delta lc4}$	(5) $P_{\Delta lc5}$	(6) $P_{\Delta lc6}$
liberal	0.0962 (0.230)	0.196 (0.230)	0.383 (0.231)	0.398* (0.221)	0.437** (0.217)	0.494** (0.210)
Constant	-0.104 (0.171)	-0.168 (0.170)	-0.253 (0.172)	-0.329** (0.164)	-0.282* (0.161)	-0.278* (0.156)
Observations	96	97	98	98	100	100
R-squared	0.002	0.008	0.028	0.033	0.040	0.053

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## C Appendix C: In:Out Group Ratios in Gendered Risk Prediction

The results in describe the ratio between in:out group predictions according to the following specification.

$$\varepsilon_{c,g}^r = \hat{P}_{c,g}^r - P_{c,g}^r \quad (5)$$

$$ER_c = \left| \frac{\varepsilon_{c,g}}{\varepsilon_{c,g^{-1}}} \right| = \beta_0 + \beta_1 female_i + \beta_2 liberal_i + \phi_i D_i + \mu_i \quad (6)$$

where  $\varepsilon_{c,g}^r$  refers to the risk preferences ( $r$ ) prediction error for choice  $c$  within subgroup  $g$ , the out-group is notated as  $g^{-1}$ , and  $\phi_i D_i$  is a vector of effects from a vector of demographic controls.

### C.1 Risk Error Ratios

The specification results in Table C.1 shows us an interesting trend where there is a measurable change in the size of the in/out group prediction error ratios. The region of the distribution of choices where the risk associated with taking Lottery B is higher also shows a strong trend of smaller error ratios for female predictors. This means that the women in phase 2 were better at making in group prediction over the high risk portion of choices than men were. For example, the coefficient on *female* in the second risk choice can be interpreted as estimating a 0.806 reduction in the error ratio for risk choice 2 when the predictor is a female making predictions about other females versus the baseline of a male making predictions about other males with a high level of significance ( $p < 0.01$ ). We see similar reductions in the error ratio.

The coefficients for risk choices 2, 3, and 4 all estimate an on average reduction of the error ratios to being slightly below 1, such as  $ER_2 = 1.727 - 0.806 = .921$  for the average woman making a prediction about risk preferences about other women:men. This suggests that women not only have an advantage over men in reducing their in-group error ratio but that they are slightly more accurate about their in-group predictions compared their

Table C.1. In:Out Group Risk Prediction Error Ratio by Gender

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$	(7) $ER_7$	(8) $ER_8$	(9) $ER_9$	(10) $ER_{10}$
female	-1.067* (0.544)	-0.806*** (0.216)	-0.495*** (0.114)	-1.085*** (0.391)	-0.138 (0.708)	-0.513 (0.968)	0.578 (0.467)	-1.023 (1.295)	1.259 (1.551)	-1.048 (1.448)
liberal	0.0345 (0.597)	-0.147 (0.235)	-0.181 (0.124)	-0.0178 (0.441)	1.052 (0.793)	0.00287 (1.075)	-0.609 (0.513)	1.014 (1.432)	-0.0385 (1.742)	1.264 (1.557)
conservative	0.755 (0.618)	0.205 (0.246)	0.00184 (0.129)	0.391 (0.444)	0.267 (0.804)	-0.269 (1.097)	0.123 (0.530)	2.415 (1.472)	-1.650 (1.732)	1.206 (1.695)
Constant	1.608*** (0.522)	1.727*** (0.205)	1.468*** (0.107)	1.957*** (0.380)	-0.118 (0.663)	1.076 (0.912)	0.354 (0.445)	1.081 (1.215)	-0.112 (1.463)	0.354 (1.406)
Observations	144	148	151	139	148	153	148	152	148	115
R-squared	0.047	0.125	0.162	0.074	0.013	0.002	0.020	0.023	0.015	0.010

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

out-group predictions with regards to risk preferences. There is no significant or consistent direction of results for the controls on an individual's political alignment at any point along the distribution of risk preference choices so in order, the significant prediction error ratios for women in risk choices 1-4 are 0.541, 0.921, 0.973, and 0.872. These ratios can be thought of as the size of the in-group prediction error as a percentage of the out-group prediction error so that a resulting error ratio of 0.541 translates to the in-group error being 54.1% of the size of the out-group error or 45.9% smaller. The rest of the error comparisons would be 7.9%, 2.7%, and 12.8% smaller error for the in-group predictions than the out-group, respectively.



## D Appendix D: Robustness Check

Table D.1 considers the analysis if all double switchers ( $n=1$ ) from the DMPL elicitation were dropped from the phase 1 deciders, recalculating the prediction error terms without the double switching observation(s). The only double switcher was a political moderate, so the results from Table 8 would not be affected. Additionally, there was no double switching in the time preference elicitation so none of those portions of the analysis were affected either.

Table D.1. Risk Prediction Error Ratio ( $ER_c$ ) by Gender

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$	(7) $ER_7$	(8) $ER_8$	(9) $ER_9$	(10) $ER_{10}$
female	0.315 (0.266)	0.260 (0.161)	-0.0108 (0.270)	0.854 (0.645)	0.0804 (0.359)	-1.334** (0.647)	-0.779 (0.935)	0.380 (0.680)	-0.200 (0.835)	-2.022 (2.259)
$Guess_f$	-0.255*** (0.0950)	-0.288*** (0.0686)	-0.303*** (0.0758)	-0.440* (0.262)	-0.0844 (0.204)	-0.322 (0.729)	0.346 (0.887)	-0.250 (0.390)	-0.563 (0.693)	1.932 (1.666)
female $\times$ $Guess_f$	-0.303 (0.234)	-0.0946 (0.144)	-0.0276 (0.113)	0.167 (0.439)	-0.102 (0.312)	0.325 (0.746)	-0.588 (0.952)	0.226 (0.721)	0.811 (0.851)	-0.967 (1.975)
Constant	1.068*** (0.146)	1.060*** (0.0913)	1.357*** (0.244)	1.587*** (0.229)	1.476*** (0.277)	2.731*** (0.623)	3.242*** (0.823)	2.059*** (0.411)	2.496*** (0.676)	5.023*** (2.023)
Observations	287	303	307	295	308	304	299	304	308	235
R-squared	0.027	0.045	0.016	0.010	0.002	0.036	0.011	0.003	0.002	0.014

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## E Appendix E: Robustness Check for Zero Denominators

To account for missing observations in the main analysis of this paper, I implement a simple adjustment to rectify the observations missing due to undefined ratios where the denominator was equal to zero. Specifically, since the ratio in the models was defined as

$$\varepsilon_{c,g}^r = \hat{P}_{c,g}^r - P_{c,g}^r \quad (7)$$

$$ER_c = \left| \frac{\varepsilon_{c,g}}{\varepsilon_{c,0}} \right| \quad (8)$$

any time that a predictor makes a perfect prediction such that  $\hat{P}_{c,g}^r = P_{c,g}^r$ , then the error ratio becomes  $ER_c = \left| \frac{\varepsilon_{c,g}}{0} \right|$ , which is undefined and thus is processed as a missing observation by the estimation. A simple workaround implemented to check how much impact these missing observations has is used in the following form

$$ER_c = \left| \frac{\varepsilon_{c,g} + .001}{.001} \right| \quad (9)$$

whenever  $\varepsilon_{c,0} = 0$ . This is not a perfect workaround since it does artificially inflate the respective coefficients, since it is only a slight adjustment from being undefined. The purpose of this test though is a check on significance.

**Table E.1.** Risk Prediction Error Ratio by Gender with Zero Adjustment

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$ER_1$	$ER_2$	$ER_3$	$ER_4$	$ER_5$	$ER_6$	$ER_7$	$ER_8$	$ER_9$	$ER_{10}$
female	-0.0999 (0.772)	1,010 (743.6)	-0.0108 (0.270)	289.0 (323.9)	93.61 (164.1)	-1,018 (1,018)	42.38 (167.7)	-490.6 (489.8)	125.7 (177.7)	-2.054 (2.313)
$Guess_f$	39.85 (40.73)	-0.288*** (0.0686)	-0.303*** (0.0758)	57.33 (57.62)	-56.74 (56.53)	-528.7 (758.8)	-115.8 (97.53)	-358.7 (358.0)	18.16 (50.27)	1.869 (1.693)
female $\times$ $Guess_f$	-40.72 (40.73)	-117.8 (204.5)	-0.0276 (0.113)	77.13 (248.6)	-81.93 (116.5)	616.7 (765.2)	84.17 (135.4)	358.3 (358.0)	187.0 (264.6)	-0.935 (1.995)
Constant	1.968*** (0.677)	1.060*** (0.0912)	1.357*** (0.244)	97.48 (96.00)	96.03 (94.18)	1,078 (1,017)	138.5 (102.6)	493.2 (489.8)	190.9* (114.2)	5.086** (2.070)
Observations	295	307	307	297	308	310	307	310	310	235
R-squared	0.017	0.006	0.016	0.002	0.006	0.012	0.003	0.014	0.003	0.014

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table E.2.** Time Prediction Error Ratio by Gender with Zero Adjustment

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$
female	635.5* (353.8)	0.479 (0.457)	0.764*** (0.252)	0.503 (0.317)	-565.1 (565.2)	0.688 (0.776)
$Guess_f$	-0.216 (0.135)	0.101 (0.174)	0.0511 (0.128)	0.254 (0.235)	-340.1 (339.2)	-0.953* (0.555)
female $\times$ $Guess_f$	-100.2 (101.2)	-0.320 (0.404)	-0.309 (0.189)	-0.501* (0.298)	338.7 (339.2)	-0.528 (0.708)
Constant	1.259*** (0.133)	1.185*** (0.223)	1.078*** (0.0758)	1.218*** (0.155)	567.7 (565.2)	2.076*** (0.595)
Observations	294	301	305	304	307	307
R-squared	0.011	0.004	0.026	0.006	0.013	0.037

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table E.3.** Risk Prediction Error Ratio by Politics with Zero Adjustment

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$ER_1$	$ER_2$	$ER_3$	$ER_4$	$ER_5$	$ER_6$	$ER_7$	$ER_8$	$ER_9$	$ER_{10}$
liberal	-0.136 (0.803)	-2,386 (1,679)	0.361 (0.352)	800.7 (802.7)	297.1 (246.8)	230.6 (388.5)	909.7 (639.6)	-641.9 (645.3)	787.8 (719.3)	-0.179 (3.334)
$Guess_t$	-0.447 (0.558)	-340.9 (1,661)	0.143 (0.153)	0.917** (0.396)	-0.475* (0.274)	321.3 (321.6)	-0.377 (0.558)	-357.2 (356.0)	111.7 (135.9)	-1.236 (1.111)
liberal $\times$ $Guess_t$	-0.180 (0.790)	340.4 (1,661)	-0.510 (0.354)	162.5 (164.2)	-148.3 (236.4)	5.156 (459.7)	-391.4 (363.9)	354.4 (356.0)	-367.3 (253.0)	-0.292 (1.893)
Constant	1.947*** (0.546)	2,387 (1,679)	1.052*** (0.0978)	1.388*** (0.415)	1.750*** (0.284)	135.6 (133.5)	3.200*** (0.559)	647.8 (645.3)	269.2 (189.3)	5.296*** (2.243)
Observations	197	198	198	200	197	199	197	200	200	155
R-squared	0.010	0.018	0.011	0.008	0.013	0.004	0.017	0.014	0.008	0.003

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table E.4.** Time Prediction Error Ratio by Politics with Zero Adjustment

VARIABLES	(1) $ER_1$	(2) $ER_2$	(3) $ER_3$	(4) $ER_4$	(5) $ER_5$	(6) $ER_6$
liberal	-34.78 (266.4)	-0.00186 (0.267)	-0.944 (0.906)	-0.0931 (0.222)	-1,689 (1,691)	0.333 (0.572)
$Guess_t$	4.968 (6.744)	-0.0447 (0.215)	-0.474 (0.780)	0.818** (0.330)	-599.5 (600.8)	2.365* (1.267)
liberal $\times$ $Guess_t$	-45.18 (40.95)	0.279 (0.264)	0.402 (0.785)	-0.773** (0.353)	599.2 (600.8)	-2.903** (1.325)
Constant	205.6 (204.8)	1.132*** (0.195)	1.998** (0.892)	1.204*** (0.175)	1,690 (1,691)	1.495*** (0.315)
Observations	194	195	198	196	200	200
R-squared	0.001	0.006	0.018	0.057	0.013	0.043

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## **F Appendix F: Instructions**

### **G Phase II - Predictions**

#### **G.1 General Instructions**

Thank you for your participation today. Just for participating in this study, you will receive \$5 toward your Take-Home Pay. In order to receive your Take-Home Pay, you must complete the entire survey today.

In this study, you are a "Predictor." Your task today will be to make predictions about the behavior of other participants in this study. The more accurate your predictions are, the better the chances that you will earn more money.

We recruited students at the University of Arkansas to be "Deciders." The Deciders made a series of private choices and entered them confidentially into a computer. The Deciders knew that their choices would never be individually observed by anyone other than the researchers. One choice made by each Decider was randomly selected to determine the respective decider's payment for their participation.

The Deciders made 16 decisions in total. The 16 decisions that Deciders made were broken up into two different types. 6 questions were about when the decider preferred to receive payment and 10 questions were about how much risk the decider preferred to take on to earn more money.

Every question had 2 options. For the questions about payment timing: Receive less money now or more money in 2 weeks. For the questions about choosing risks: A "safer" option with lower potential payout, or a "riskier" option with a higher potential payout.

Your job is to predict what percentage of deciders (P) chose each option for each question. One of your predictions will be randomly chosen to determine your payout for today.

Specifically, you will be predicting what percentage of Deciders: Chose to wait 2 weeks for the higher payout Chose the "riskier" option with a higher potential payout

You will be making predictions about all 16 decisions for 5 groups of deciders, one will



be the entire group, the other 4 will be subgroups of the whole group.

One of your predictions will be randomly chosen as the prediction to count for your payment. You should do your best to make accurate predictions because the closer your predicted  $P$  is to the actual  $P$ , the higher your potential payout will be. Since any one of your predictions can be chosen to count, you should treat every prediction as if it could determine your payout.

Deciders were given their choices expressed as "EUs" (Experimental Units). At the end of the survey, their earnings were converted from EUs to US Dollars using the exchange rate  $3 \text{ EUs} = 1 \text{ USD}$ . This means for every 3 experimental units the Deciders earned, \$1 was added to their payment.

### **G.1.1 Comprehension Questions**

1. What is your role in today's study?
  - Make decisions about my preferences
  - Guess what decisions the Deciders made
  - Help the Deciders make decisions
2. Did the Deciders decisions have consequences
  - Yes, their choices mattered item No, their choices did not matter
3. How many of your predictions will count?
  - All of them collectively
  - One prediction chosen at random
  - None of them

## G.2 Payment Instructions

You will have a chance to earn an additional \$5 added to your payment in a lottery based on the accuracy of your prediction that counts.

The computer will randomly draw two numbers from 0-100. We will call these numbers  $X$  and  $Y$ . These numbers are whole integers and all numbers from 0-100 have an equal chance of being chosen.

If your predicted  $P$  is less than the true  $P$ : You will receive the additional \$5 if and only if your predicted  $P$  is greater than or equal to either  $X$  or  $Y$ .

If your predicted  $P$  is greater than the true  $P$  You will receive the additional \$5 if and only if your predicted  $P$  is less than either  $X$  or  $Y$ .

The most important thing is to submit your true belief about  $P$ . Trying to game the system by guessing high all the time that will not help. It will only increase the probability that you miss out on the bonus if you draw a low  $X$  and  $Y$ .

Conversely, trying to guess low will increase the chance of missing out if you draw a high  $X$  and  $Y$ .

### G.2.1 Comprehension Questions

1. If your guess about  $P$  is that  $P = 35$  and the true  $P$  is 40, how will your bonus be determined?
  - I will not get the bonus
  - I will get the bonus if 35 is greater than or equal to either the randomly drawn  $X$  or  $Y$
  - I will get the bonus if 35 is less than either the randomly drawn  $X$  or  $Y$

### G.3 General Time Preference Prediction Instructions

For this section, you will be making predictions about the time preferences of the Decider group as a whole.

For each of the following decisions, move the slider to indicate your predicted  $P$ , which is what percentage of people you believe chose to wait 2 weeks for 20 EUs vs taking the indicated payment on that day.

### G.4 General Risk Preference Prediction Instructions

For this section, you will be making predictions about the risk preferences of the Decider group as a whole.

For each of the following decisions, move the slider to indicate your predicted  $P$ , which is what percentage of people you believe chose the second risk option.

### G.5 Group Level Time Preference Predictions Instructions

For this section, you will be making predictions about the time preference of a group of **female** Deciders.<sup>10</sup>

For each of the following decisions, move the slider to indicate your predicted  $P$ , which is what percentage of **female** Deciders you believe chose to wait 2 weeks for 20 EUs vs taking the indicated payment on that day.

*As a reminder, your prediction for the Decider group as a whole was: \*insert prediction from corresponding general prediction\* of Deciders would wait 2 weeks for 20 EUs.*

---

<sup>10</sup>Instructions were consistent across all subgroups, just replacing "female" with the appropriate subgroup title (male, liberal, conservative).

## G.6 Group Level Risk Preference Predictions Instructions

For this section, you will be making predictions about the risk preference of a group of **female** Deciders.<sup>11</sup>

For each of the following decisions, move the slider to indicate your predicted  $P$ , which is what percentage of **female** Deciders you believe chose the second risk option.

*As a reminder, your prediction for the Decider group as a whole was: \*insert prediction from corresponding general prediction\* of Deciders chose the second risk*

---

<sup>11</sup>Instructions were consistent across all subgroups, just replacing "female" with the appropriate subgroup title (male, liberal, conservative).

## Conclusion

The study of identity as an aspect of behavior is barely 2 decades old in economics, leaving a lot of innovation and discovery to still be achieved at the forefront of the field. Using experimental methods to isolate and understand the nuanced impacts of different aspects of identity and behavior will continue to be an important part in the advancement of the field. The three studies in this dissertation all explore some aspect of behavior and decision making that has a core rooting in identity. While not all decisions may have a conscious identity element in the way they are derived, they at least have an important aspect of self-identifying that has to be done by a person in a way to signal information to themselves and others. Sometimes this identity signaling can be an intentional way to make yourself belong in society or a subgroup thereof. Sometimes the signaling is non-mutable such as with race or gender. The important consideration is that there is information in the way we choose to identify ourselves and the way that people perceive our identities.

The applications of this work mainly speak to areas that need further research and understanding for policy application. Incentivized decision making is an important aspect of understanding how bias plays a role in the way we interact with each other in constrained environments. While the other social sciences have written extensively about the sociological implications of identity and bias, economics needs to continue to speak of the actual cost and value of identity and bias. It is evident that at times our perceptions of identities and how they play into stereotypes often causes us to be less effective in understanding each other and making predictions about the nature of our preferences and decisions. We also have seen that the reality of social signaling is very effective at masking the true desires of a person, even when others are aware of the signaling, due to the inability to extract the truth from bias. Often these biases can become systemic to the point where it is hard to understand the initial preference in the first place or the actual disposition.

The main takeaway from this work is that identity and bias are important to people, even

when they are aware that it may come at a personal cost. There is a level of incentive that can reduce the weighting that individuals put on these but their value is intrinsic to the individual and cannot be well defined to make a sweeping policy suggestion in an applied setting. The reality is that broad spectrum policy cannot correct for the idiosyncrasies apparent in the intersection of bias, identity, and experience resulting in spaces where generalized guidelines need to be flexible to allow reasonable adjustments. This work has added to the literature by attempting to test explicitly the values of identity and the perceptions of preferences and social desirability while leaving room for further exploration beyond the scope of just gender and politics, such as the implications of race, ethnicity, socioeconomic status, and their intersection. The intersection is most important in applied areas of labor and education and offers great opportunities for behavioral and experimental research in economics to make further connections to policy applications.