Spring 5-27-2023

Essays On Political Economy And Behavior

Seth Emblem
Southern Methodist University, semblem@smu.edu

ESSAYS ON POLITICAL ECONOMY AND BEHAVIOR

Approved by:

Dr. Tim Salmon
Professor of Economics

Dr. Bo Chen
Associate Professor of Economics

Dr. Nate Pattison
Assistant Professor of Economics

Dr. Kyle Hyndman
Professor of Economics

ESSAYS ON POLITICAL ECONOMY AND BEHAVIOR

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Parital Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Economics

by

Seth Emblem

B.S., Economics and Philosophy, Texas Christian University
M.A., Economics, Southern Methodist University

May 27, 2023

ACKNOWLEDGEMENTS

First I want to thank my advisor, Dr. Tim Salmon. The path towards completing my doctoral degree had unexpected twists and turns, so I must thank Dr. Salmon for helping to guide me to this point. His suggestions and critiques helped to strengthen my research projects and helped guide me towards avenues of research I did not initially know that I would be interested in. Dr. Salmon helped me to push forwards during some of the more difficult times of being a graduate student, so I cannot help but express my gratitude for guiding me to this point of my academic career.

I also wish to thank Dr. Bo Chen and Dr. Nate Pattison. Their contributions to my research go beyond serving on my dissertation committee. Dr. Chen's first year microeconomic theory course and second year mathematical economics course helped me to discover my broader interest in microeconomics, both in research and teaching, and a specific interest in game theory and strategic behavior. Dr. Pattison's second year labor economics course and the research paper I wrote for the course proved to be an important starting point for the direction of my dissertation and its general theme of political economy.

I want to express my appreciation to some of the graduate students I have met along my way at SMU. First, my cohort members, Parker Fleming, Kristina Bishop, and Hang Chen. In Parker, a fellow TCU grad, I have someone to vent with through the "struggles" of being a TCU sports fan. Kristina is one of the kindest people I have ever met, and it was a joy to share class and study time with such a pleasant person. Hang is probably the funniest and most brilliant person in our cohort, and it was always a pleasure hanging out with him, both in the experiment lab and in general on campus. I could not have asked for a better group of people to go through the struggles of graduate school. I would also like to thank

the fellow Economics Ph.D. students who participated in pilot versions of my experiments and surveys to help me design the final versions of my data collection instruments. On that note, I would like to specifically thank Hang Chen, Kristina Bishop, and William Cunningham for volunteering to judge the responses to the Creative Uses Task in Chapter 2 of my dissertation. Finally, I would like to thank fellow Economics Ph.D. Ding Liu for sparking my interest in following club soccer (European in particular). Thanks to Ding I discovered a new hobby that I could go to during my time as a graduate student.

I would be remiss if I did not acknowledge the professors from undergraduate that have had a major influence on my academic career. From economics, I would like to acknowledge Dr. Harvey, Dr. Garnett, and Dr. Nikaj. Even though my research and teaching interests tend to fall more on the microeconomics side, it was Dr. Harvey's Intermediate Macroeconomics course that ultimately pushed me to major in economics as an undergrad. His pluralistic perspective has been an important influence in how I try to approach my research and teaching. To Dr. Garnett, I must relay my appreciation for providing me my first real exposure to Adam Smith in his History of Economic Thought course. My exposure to the writings of Adam Smith has been a major influence on me and my interest in behavioral economics. Dr. Nikaj helped to peak my interest in graduate study of economics with her Intermediate Microeconomics and Introductory Econometrics courses and by offering me experience as an undergraduate research assistant. From mathematics, I would like to acknowledge Dr. Staples. Dr. Staples is the head of the actuarial program at TCU and was my math advisor. Not only is Dr. Staples one of the kindest professors I have ever had the pleasure to meet, but I believe I must thank the actuarial program for being perhaps the main reason I started my journey to becoming an economist. If it were not for the actuarial program's requirement to take intro and intermediate microeconomics and macroeconomics, I am not sure I would have ever discovered my passion for the subject. From philosophy, I believe I have no choice but to acknowledge the entire department. Fortunately, the department's small size allows me to mention each person by name: Dr. Galvin, Dr. Hestir, Dr.

Roche, Dr. Harris, Dr. McCormick, Dr. Cray, and administrative assistant Beth Philp. The Department of Philosophy at TCU provided me with an "intellectual home" that I could go to discuss ideas and to take in the broader learning experience of being an undergraduate student. The critical thinking skills and habits I picked up as a philosophy student at TCU proved to be invaluable to me in my graduate studies, especially when it came to studying for the first year microeconomic theory and macroeconomic theory qualifying exams.

I would like to acknowledge the following people I met as students at TCU. First, the people involved in the improv and stand-up comedy scene at TCU. In the interest of space, and out of fear that I will inadvertently leave someone out, I will not list out names but rather collectively acknowledge this fine group of people I had the pleasure of socializing with and expressing myself outside of an explicit academic context. With that said, I would like to list a few fellow students from undergrad that were influential to me directly in the academic setting: Timothy, Nicholas, Chrichelle, Stephen, Rachel, Allegra, Rachel, and Miranda. I have limited this list to a select few in my graduating class, but I would like to broadly relay my appreciation to anybody in the departments of Economics, Philosophy, and Mathematics or the John V. Roach Honors College that I had the pleasure of interacting with in my four years at TCU.

There are more people that I should acknowledge than I have space and time, so I will conclude these acknowledgements with a few more people. From my time at China Spring High School, I would like to acknowledge the following teachers and professors: Mr. Dyer, Mr. Notley, Mrs. Davis, Mrs. Smith, Dr. Mitchell, and Dr. Emblem. I also want to acknowledge the communities of Grace Baptist Church and Oak Grove Baptist Church for giving me a place to go to through some of the difficult times I faced in high school and beyond. I would also like to thank my family in Texas, not just in Waco but beyond, for giving me a loving group of people that I could be with. Not everyone has the luxury of being part of such a loving family, so I am grateful to be part of this one. Love and appreciate you all.

Emblem, Seth B.S., Economics and Philosophy, Texas Christian University
M.A., Economics, Southern Methodist University

Essays on Political Economy and Behavior

Advisor: Dr. Tim Salmon

Doctor of Philosophy degree conferred May 27, 2023

Dissertation completed May 25, 2023

This dissertation consists of three essays which focus on the intersection of political economy and behavioral economics. I am interested in understanding how people behave in politically-adjacent areas so that we can find the policy interventions that can best address the issues that arise from these areas. To do so, I utilize incentivized survey instruments and laboratory experiments to study how people behave in these settings. In Chapter 1, I examine the question of how people respond to standard political surveys and whether these surveys reflect people's genuine beliefs regarding factual political questions. Responses to standard political surveys suggest that Republicans and Democrats do not agree on the correct answer to factual political questions, which taken at face value might suggest that implementing policies that allocate time and money to information campaigns would be a good idea. However, if responses to standard political surveys do not reflect people's genuine beliefs, then such a policy would end up wasting time and money. In Chapter 2, I examine whether conspiracy theory endorsement is associated with behavioral differences relevant to economic contexts such as the workplace or broader economic growth. Contemporary discussions about conspiracy theory endorsement in popular media suggest that it is problematic and that policy interventions that would curtail endorsement would be worthwhile, so we want to understand in which behavioral areas conspiracy theory endorsement could lead to problems. In Chapter 3, I look at how in localized, small group settings, social norms in favor of a minority preference can form. The standard conception of social norms is that they

represent majority preferences, so this third chapter provides an initial investigation into how social norms that represent minority preferences can form. Understanding the circumstances under which social norms representing relatively unpopular actions and behaviors can form is relevant for knowing when policy interventions that try to push individuals to take a different action are more (or less) likely to succeed.

The existing literature on political surveys has shown that providing monetary incentives for people to give the correct answer to factual political questions significantly reduces the gap in answers provided by Republicans and Democrats. However, recent behavioral differences between Republicans and Democrats (e.g. involving COVID) casts some doubt about the persisting strength of those prior findings. In the first essay, I investigate whether providing monetary incentives still reduces partisan gaps by asking a set of replication questions as well as some new questions that address new issues that either have or have not been subject to concerted misinformation campaigns. I also explore other aspects pertinent to designing political surveys that might induce responses that better reflect people's genuine beliefs. My results show that where partisan gaps exist, providing monetary incentives significantly reduces the partisan gap.

In the second essay, I investigate whether conspiracy theory endorsers are different from others in the behavioral areas that prior academic studies suggest. I implement an incentivized online survey, with a borrowed measure for conspiracist ideation, to be able to observe individuals' behavior in each of these areas and to test whether conspiracy theory endorsers are different from others in these areas. I find that while conspiracy theory endorsers provide more extreme responses to a set of factual political questions, as we might expect, they did not generally behave differently from others in the incentivized tasks.

In the third essay, I examine how in localized, small group settings, social norms representing minority preferences can form. To do this, I provide a simple theoretical model that shows under which circumstances a social norm representing a minority preference is (and is not) expected to form. I also conduct a laboratory experiment to see whether individuals

behave in this setting in the manner that the theoretical model predicts. I find that while social norms that represent a minority preference do occur, they do not occur quite in the same manner as the theoretical model predicts. The findings from this study provide suggestions for how individual's behavioral considerations and strategy choices can impact the formation of these types of social norms.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This is dedicated to my dear, departed sister Kelsey, the strongest person I have every known.

CHAPTER 1

WHERE CAN FINANCIAL INCENTIVES ELIMINATE PARTISAN ANSWERS TO FACTUAL QUESTIONS

1.1 Introduction

Partisanship has been a part of the American political environment since the founding of the country, to the dismay of the first president George Washington[1]. Political partisanship in the United States continues to thrive in the present-day, where people form an identity with a given political party that goes beyond which candidates they vote for on election day (e.g. [Iyengar et al., 2012], [Iyengar and Westwood, 2015]). One concern with political partisanship is that people who tend to identify with one political party might not live in the same "world" as the people who tend to identify with the other political party. In the American context, the divergence in how those who identify as Republican and Democrat respond to political polls and surveys about facts would suggest that partisans have significantly differing perceptions of reality (e.g. [Bartels, 2002], [Achen and Bartels, 2006]).

However, prior studies in the literature on political partisanship suggest that partisans' answers to these political polls do not necessarily reflect their genuine beliefs regarding facts about the real world. Studies such as [Bullock et al., 2015] and [Prior et al., 2015] find modest partisan differences between republicans and democrats in their responses to factual political questions, but that the gap significantly decreases when subjects are provided small monetary incentives to provide the correct answer. In a slightly different context, [Robbett and Matthews, 2018] also provides evidence that providing monetary incentives reduces the partisan gap between republicans and democrats in their answers to factual political questions. [Bullock and Lenz, 2019] provides a survey of the literature which concludes that,

[1] [Rasmussen, 2021]

while it is difficult to precisely pinpoint the avenue of motivated reasoning, a significant proportion of responses to standard political surveys do not reflect genuine beliefs. These results give a strong suggestion that the genuine beliefs of those who identify as republicans and democrats are more closely aligned than what reports based on standard unincentivized political polls claim.

That said, the COVID-19 pandemic, for instance, has provided examples which seem to suggest that partisans have significant differences in their genuine beliefs. One example is the strong relationship between political affiliation and reported vaccination rates against COVID-19. One study from the Kaiser Family Foundation between January 2021 and January 2022[2] found that the difference in reported vaccination rates between counties that voted for Biden versus counties that voted for Trump in 2020 actually increased across time. Unlike standard political poll questions, the choice of whether to get vaccinated is highly incentivized. Given the potentially fatal consequences of contracting the virus, one should get vaccinated against the virus if she genuinely believes that the vaccine has high efficacy.

How do we reconcile the findings of papers such as [Bullock et al., 2015] and [Prior et al., 2015] with the divergence in behavior between democrats and republicans regarding the COVID-19 pandemic? Perhaps political bias has intensified since the publishing of these papers and partisans are simply more willing to behave in a partisan manner than they were before. The rise of networks such as Newsmax and OANN, as well as an increase in partisanship of more established networks like MSNBC and Fox News, could have a role. Similarly, the increase in social media usage among Americans could be guiding people into information bubbles, which could exacerbate the divergence in behavior between republicans and democrats. Whatever the mechanism might be, the fundamental question is how much a common understanding of facts between republicans and democrats has eroded.

Regarding the fundamental research question, there are three possible cases we could unveil. The first is that providing incentives to political polls no longer shrinks partisan

[2] [Kates et al., 2022]

gaps between republicans and democrats, in which case the findings of the above literature would not replicate. This case could be realized if there has been a universal shift between the accepted "facts" of republicans and democrats. The second is that the findings of the earlier studies replicate, but providing incentives for questions about newer, highly politicized issues (e.g. COVID) does not shrink partisan gaps between republicans and democrats. This case could be realized if, for instance, the increased partisan nature of news organizations has only influenced people's accepted "facts" about newer, more novel issues, since those received more deliberate attention. The third is that the findings of the earlier studies replicate and providing incentives for questions about newer issues shrinks partisan gaps between republicans and democrats. In this case, the observed behavioral differences between republicans and democrats could be due to differences other than their set of accepted "facts" (e.g. differences in vaccine hesitancy between republicans and democrats).

Highlighting these possible cases helps to point us towards our approach to addressing the fundamental question. Given the nature of the question, we constructed our own survey instrument. The first element of our survey was a replication of the survey questions from [Bullock et al., 2015] and [Prior et al., 2015], updating the questions to reflect the most recent numbers at the time of our survey. Second, we added new questions on issues that were subject to partisan misinformation campaigns (e.g. vaccine efficacy). We also added new questions on issues that were not subject to such concerted misinformation campaigns (e.g. farm subsidies). Finally, we implemented random treatments regarding the degree to which subjects received monetary incentives for providing correct answers.

On top of our broader investigation into these three sets of questions, we also wanted to investigate other questions about what might explain the partisan gaps found in standard political surveys and how to design surveys to eliminate these gaps. To do this, we included a few noteworthy elements to our survey design. One of these was to include an "I don't know" option for all questions as a way to limit subjects from providing indifferent partisan responses (e.g. cases where a subject might provide a seemingly partisan response because

he does not have a good estimate for what the correct answer might be). Another element was to include a set of cognitive reflection test (CRT) questions to see if providing partisan responses was associated with inattentiveness. Finally, the one other element was to allow subjects to revise a randomly selected portion of their answers as a test for whether partisan signaling was a conscious choice of subjects.

From our survey data, we found that partisan gaps were smaller than we expected, but that monetary incentives tended to diminish partisan gaps where they existed. Partisan gaps were largest for the Misinformation set of questions, and monetary incentives decreased partisan gaps most for the Misinformation set. Regarding our other research curiosities, responses to our CRT questions was associated with responses to the factual political questions, few subjects chose to revise their answers, and subjects answered "I don't know" more often for the Misinformation and No Misinformation set of questions (compared to the Replication set of questions). These results help to guide further research into the common understanding of facts between republicans and democrats and how to better design political surveys such that responses to these surveys better represent the genuine beliefs of participants. The remainder of this paper is structured as follows. Section 2 goes through the main hypotheses of our study. Section 3 describes our survey design. Section 4 goes through our results in more detail, and Section 5 concludes.

1.2 Hypotheses

The focus of this study is to investigate how much a common understanding of facts has eroded between republicans and democrats. The first part of that investigation to try to replicate the findings of [Bullock et al., 2015] and [Prior et al., 2015]. The apparent disparate behavior between republicans and democrats with regards to the COVID-19 pandemic is consistent with the possibility that the common understanding of facts has eroded to such a degree that providing monetary incentives might not significantly decrease differences in responses between republicans and democrats. Although the stylized fact regarding the

COVID-19 pandemic is consistent with such a shift, we do not have an explicit theoretical reason to expect that republicans and democrats have significantly shifted their genuine beliefs, so our hypothesis for the replication questions mirrors that of the prior papers.

Hypothesis 1 *(Replication Questions) Providing monetary incentives will decrease the partisan gap between republicans and democrats for questions in our replication set.*

The second part of our investigation is to see whether republicans and democrats have significantly disparate genuine beliefs regarding contemporary issues that have been subject to concerted misinformation campaigns. The stylized fact concerning behavior during the COVID-19 pandemic suggests that partisans do in fact significantly disparate genuine beliefs about these issues. If both republicans and democrats have similar genuine beliefs about the efficacy and safety of COVID-19 vaccines or about the risks involved with a coronavirus infection, then we would likely not observe our stylized fact. As such, our hypothesis for this second set of factual questions is that providing monetary incentives will not significantly decrease the gap in responses between republicans and democrats.

Hypothesis 2 *(Misinformation Questions) Providing monetary incentives will not decrease the partisan gap between republicans and democrats for questions in our misinformation set.*

The third part of our investigation is to see whether republicans and democrats have significantly disparate genuine beliefs regarding contemporary issues that *have not* been subject to concerted misinformation campaigns. Issues in this set of questions are similar to the issues in the replication set: these are potentially partisan issues that we might expect republicans and democrats to give different answers to in an unincentivized survey, but they have not been subject to misinformation campaigns to the degree of the issues in the misinformation set. As such, our hypothesis for issues in this set is that providing monetary incentives to subjects will decrease the partisan gap between republicans and democrats.

Hypothesis 3 *(No Misinformation Questions) Providing monetary incentives will decrease the partisan gap between republicans and democrats in our no misinformation set.*

1.3 Survey Design

As an initial investigation of our research question, we conducted a pilot study in 2021. In our study, we randomly[3] selected one-half of our subject pool to receive our incentivized treatment, while the other half received the unincentivized treatment. The instruction text for the incentivized treatment is provided below:

> After the survey is concluded we will randomly select 5 of your answers to compare to the true answer. For each question, if you are close enough to the correct answer, we will pay you $1. This means that you can earn up to $5 for answering these questions accurately. When we show you the correct answers in the end, we will also note the range that we considered close enough to earn payment. For any question considered for payoff, if you have entered "I don't know" you will earn $0.25.

Subjects in the control (unincentivized) group only received the general instructions. In both the control and treatment group, subjects were given the option to answer "I don't know" to avoid subjects providing indifferent partisan responses.

On top of the main research objective to investigate the degree to which a common understanding of facts has eroded between republicans and democrats, there were other curiosities we wanted to be able to investigate with our design. One was whether partisan signaling was a conscious choice of subjects, or whether some other mechanism might be driving partisan gaps between republicans and democrats. If partisan signaling is a conscious choice, then subjects might be willing to change their responses if they are provided greater accuracy incentives (i.e. larger monetary incentives to provide the correct answer). To be able to investigate this, we gave subjects in both the control and treatment group the

[3]Using Qualtrics' "randomizer" function

opportunity to revise a randomly selected set of three questions[4]. The text of the revision option for the control group is as follows:

> Thank you for answering these questions. We have chosen three of your answers at random to possibly generate additional earnings for you. We will compare your answers to the true answers and pay you $1 for each of your answers that is close enough to the true answer. For any question you answered with "I don't know", you will earn $0.25 if it is chosen for payment. You can choose to submit the answers you have already chosen or you can choose to edit your answers.
>
> Would you like to edit your answers or submit ones already entered?

Since the treatment group already received monetary incentives to provide correct answers, the text and payoffs were slightly different. The text of the revision option for the treatment group is as follows:

> Thank you for answering these questions. We have chosen three of your answers at random to possibly generate additional earnings for you. On top of the incentives we have already provided for accurate answers we will pick three more questions and double the potential earnings to $2 for a correct answer. We will compare your answers to the true answers and pay you $2 for each of your answers that is close enough to the true answer. For any question you answered with "I don't know", you will earn $0.50 if it is chosen for payment. You can choose to submit the answers you have already chosen or you can choose to edit your answers.
>
> Would you like to edit your answers or submit ones already entered?

Another curiosity was whether partisan signaling was consistent with inattention. To be able to test whether there was a link between the two, we included three cognitive reflection

[4]We used Qualtrics' "randomizer" function such that each question was chosen for revision an equal number of times

test (CRT) questions to check whether subjects were paying close attention to the prompts they were given. The text for the CRT questions for the treatment group is as follows:

> For the final 3 questions, we will give you three problems to solve. Please give us your best guess at the answer to each. You will have 45 seconds to answer each question.
>
> We will pick one of these questions at random and if your answer to that question is correct you will earn $1.

The CRT prompt for the control group was the same except that they were not given monetary incentives for these questions.

For this study, we distributed our survey via the online recruitment site *Prolific Academic*. The sample included only subjects residing in the United States, and was nationally representative on the demographics of sex, age, and ethnicity[5]. Our study was conducted on October 26 and 27, 2021 and we had a sample of 301 subjects.

The survey that subjects completed included 23 factual questions about issues that ranged from economic conditions to the efficacy of vaccines. In the replication set of questions, subjects were asked to provide their best estimate of the unemployment rate, the size of the public debt in January 2021, the percentage of Americans living in poverty in 2020, the national average price of gasoline, the number of U.S. soldier causalities from 2017 to August 2021, the percentage of the U.S. population born in a foreign country, the percentage of votes that went to Joe Biden in the 2020 general election, and the percentage of public spending that went to the Department of Defense in 2020. In the misinformation set of questions, subjects were asked to provide their best estimate of how much the U.S. spent on border patrol in 2020, the fraction of medically treated firearm injuries that were due to accidents, the number of documented cases of non-citizens registering to vote in the U.S. from 1982 to 2021, the frequency of severe allergic reactions to the COVID-19 vaccine and to

[5]These are the three demographics in which Prolific allows for collection of a representative sample.

the Measles, Mumps, and Rubella (MMR) vaccine, the efficacy of the two mRNA vaccines at preventing symptomatic illness, the percentage of unvaccinated people hospitalized due to the coronavirus, the percentage of vaccinated people hospitalized due to the coronavirus, the average annual growth rate for the end of the Obama presidency, and the average annual growth rate for the beginning of the Trump presidency. In the no misinformation set of questions, subjects were asked to provide their best estimate of the percentage of public spending that went to Social Security and to interest payments on national debt in 2020, how much the U.S. spent on food and nutrition assistance and to farm income stabilization in 2020, and how much wearing a seat belt reduces the risk of fatal injury.

To prevent subjects from seeking help or trying to look up answers, we put a time limit on how long they could take to answer each question. Subjects were given either 45 or 60 seconds to answer all questions on the page. Pages that either included longer text that subjects needed to read or multiple individual questions had a time limit of 60 seconds, while every other page had a time limit of 45 seconds. The survey was programmed in Qualtrics so that the subject would be automatically moved on to the next page whenever they reached the time limit, regardless of whether the subject had answered the question(s) on that page.

1.4 Results

For an initial look at our data we have provided Table 1 which provides the median of answers given to all of our questions in the survey for Republicans and Democrats (answers by Independents are not included here) separated out into the treatments involving incentives or not with each question identified as being in the replication set, misinformation set or non-misinformation set. The relevant comparisons here are first between the answers given by Republicans and Democrats in the Non Incentivized treatment. We expect that for each question we should observe a gap between the answers given by the R's and the D's when incentivizes are not present. In this table we have provided the results of significant tests of the differences between Republican and Democrat answers for each question pulled from

individual level question regressions. The stars representing significance levels are in the column for the Republican data. An important point to note is that we find partisan gaps in the No Incentives treatment for only 8 of the questions. In the treatment involving incentives, we see significant differences that remain for two of those eight and then there is a significant difference for one other question as well.

All of the hypotheses above involve looking at this initial gap and then testing whether the gap found in the incentivized treatments is smaller or not. Our results are heavily foreshadowed by seeing that while we get partisan gaps in eight of the questions without incentives, only two of those gaps remain with incentives, but there is some complexity to pooling these questions into more complete regressions. The units of the responses to questions differ as can the direction of the gap in terms of which group has the bias to be higher / lower than the other. In order to systematize the analysis, we follow the approach of [Bullock et al., 2015] by recoding all of the answers for each question into a common range $[0, 1]$ where a value of 0 corresponds to the most extreme answer given by a respondent in our data in whichever direction corresponds to Republican bias on that question and a value of 1 corresponds to the most extreme answer given in the data corresponding to the Democrat bias on that question.[6] All other responses are then placed along this range accordingly. We will call this measure G_{ij} representing the rescaled answer for an individual i on question j and the interpretation of these scores is that high values indicate Democrat friendly responses and low values indicate Republican friendly responses.

After using this transformation of our data we can again take a summary look in how responses differed according to incentives pooling across multiple questions. Figure 1 shows the difference in the average values of G_{ij} between Republicans and Democrats with each set of questions pooled together. As a benchmark comparison we have also provided Figure 2 which shows the average values of G_{ij} for each set of questions just for those individuals self

[6]To be clear, these most extreme answers need not be given by an actual Republican or Democrat. These two endpoints are simply let us identify the max and min in the data.

identifying as Independents. Examining Figure 1 provides suggestive evidence that the partisan differences may be larger in the No incentives treatment than the incentives treatment at least for the replication and Misinformation set. The No Misinformation set suggests no partisan differences in either treatment. Examining Figure 2 provides a reasonable benchmark comparison for how incentives might alter the behavior of those without a clear interest in signaling any partisan support. In that case we are providing the raw average values for their answers and it seems quite clear that there is no difference in how the independents answered the questions with and without incentives though there do seem to be some shifts in how partisans answer questions with and without incentives.

Testing hypotheses directly off of these summary statistics is problematic due to the fact that we have repeated observations for individuals and that we have additional control variables we would like to include as potential explanators of the responses. We therefore provide Table 2 to show how answers to these questions change based upon the Incentives treatment taking these issues into account. Each column provides a simple OLS regression with the standard errors clustered on the individuals where the columns differ according to which set of questions are included. The dependent variable is the G_{ij} for each respondent on each question. Our independent variables include a treatment variable for the incentive treatment, indicator variables for an individual being a Republican or Independent with these variables interacted with the Incentives treatment variable. We also include the CRT score of the individual and a variable for whether the question allowed a response using a free response text box. The base comparison category here are Democrats. The coefficient on the Republican variable indicates whether a partisan gap exists in the no incentives treatment. This coefficient is always negative which is expected as the most Republican friendly answer is coded to be 0. It is reasonable that on average Republicans trend in that direction. This coefficient is not robustly significant though as we see it is only significant in the Misinformation set of questions. This is an important point as our hypotheses concern where a partisan gap can be reduced and if one does not exist, there is nothing to test. We

can test this in the Misinformation set and there are two ways to do the test. The first is to see if the gap between the two groups narrows. This is determined by the coefficient on the interaction term between Republican and Incentives. It is not significant. We can also test for the existence of a gap overall by examining the linear combination of the Republican variable and the interaction with the incentive treatment. The value of this combined coefficient is -0.05 and the $p-$value from a significance test is 0.018. So that suggests an overall gap remains.

The question by question analysis shown in Table 1 suggests that the first three columns of Table 2 are actually mis-specified for testing our hypotheses. We should not be testing for a gap declining for questions where no gap exists. This fact motivates column 4 which takes all of the questions across all three sets shown in Table 1 to have initial partisan gaps and conducts the regression on them to get at an overall test of whether or not incentives are useful in shrinking partisan gaps. Here we find, as required by construction, that there is in fact a significant partisan gap without incentives and then the Republican interaction is positive and significant which indicates that the gap narrows with incentives. If we test for the existence of the gap in the incentive treatment, the size of the gap is -0.04 and the $p-$value of the significance test is 0.033. So the gap over all of this set of questions declines but does not quite disappear.

To provide a more targeted test of our hypotheses we can redo the regressions for our three sets of questions limiting ourselves to only including the questions which contain a partisan gap in the unincentivized treatment which might be decreased in the incentivized treatment. Table 3 provides the results for that restricted case. Here obviously the Republican variable is always negative and significant, by construction. The only set of questions for which the interaction between that variable and the incentives treatment is significant is the Misinformation set. As this set has the largest initial gap, it seems reasonable to be the one that has the most significant decrease. We can also conduct the tests to see whether a partisan gap exists in the Incentivized treatment using the same linear combination as

above. For the Replication set, the linear combination of those two coefficients is -0.086 and that has a $p-$value of 0.009, for Misinformation -0.03 with a $p-$value of 0.207 and for no Misinformation 0.015 and a $p-$value of 0.815. Consequently we find that there is a partisan gap in the replication set which does not go away but in the other two sets incentives effectively eliminate the partisan gap in responses. In the case of the No Misinformation set, this finding is largely due to the relative lack of a partisan gap even without incentives while for the Misinformation sets we fund robust gaps without incentives which generally shrink with incentives. In the Replication set, incentives do work to decrease the gap in two out of the three questions with a gap, but not on the one question regarding Defense Spending and this leads to the overall finding of a lack of a drop in the gap.

In addition to the questions regarding whether incentives would reduce the partisan gaps, we also had an interest in investigating other questions about what might explain the gaps and how surveys might be designed to eliminate them. An initial question was whether or not an individual's responses were a function of how much attention they were paying to the survey. We tested this by asking a standard battery of CRT questions which are intended to measure how deeply an individual is thinking in how they respond to questions. In each of the prior tables, we have included the CRT score of the individual as an explanatory variable and we do find that answers to the questions are correlated with this measure. In general we find that higher CRT scores lead to answers that have a higher G_{ij} value or one that is closer to the Democrat leaning side. This is not very surprising given that many of the questions that were designed to generate the most extreme bias were ones in which that bias was likely to come from Republicans. We do not interpret this score as suggesting that paying attention leads to more Democratic leaning answers but rather that those who are paying attention are more likely to provide an answer that is less extreme.

A related item of curiosity concerns when and why an individual responds to a question with the answer "I don't know". This option was included as an option which would eliminate people engaging in some form of partisan signaling because they don't know the answer to

a question. In fact, we expect this may account for our lack of finding partisan gaps in many questions even without incentives. In the incentivized treatments most individuals would likely be financially better offer providing this response to many questions, as for many of them getting the right answer is quite difficult. Given that possibility, we need to know if people responded with this answer more often in the incentivized case to see if simply removing some of those answers accounts for the impact of the incentives. Table 4 provides a linear probability regression examining the correlates with whether someone answers a question with "I don't know." Note that this regression includes the full sample of respondents while the previous regressions do not. The reason the full sample is not in the previous sets is in large part due to the fact that we excluded answers of "I don't know" from the partisan gap analysis for obvious reasons. Interestingly we do not find that people are any more likely to use this option in the Incentives treatment rather than the unincentivized treatment. In fact the only robust finding seems to be that individuals are more likely to use this response category in the No Misinformation and Misinformation sets of questions. Many of these questions are difficult to answer and that is likely a reasonable finding.

Our final research question concerns whether individuals would be interested in revising their answers to questions when provided upgraded incentives. In the no incentive treatment, the revise option is essentially a within subjects test of whether or not incentives shift responses and potentially diminish partisan gaps. That is not the best interpretation of the test though. We included this as a test for whether in the No Incentives treatment, people were knowingly and intentionally giving partisan responses when they had no self-interest in submitting a correct answer but would then backtrack once they knew there were incentives at stake. We see this as a test for whether or not the impact we find coming from the incentive treatment is coming from people consciously and intentionally partisan signaling. In the incentives treatment, we offered higher incentives which can be considered a within subject test of asking if higher incentives can get people to reconsider. Table 5 provides a regression looking at what might drive a decision to revise. We include a treatment variable

for incentives as well as variables for Republican or Independent. We include two variables to account for people who give answers that yield extreme high or extreme low values of G_{ij} on average.[7] One would expect that these are the people in these groups who would benefit most from revising. In the end we find no significant correlates. This is likely due to the fact that we find only 36 people choosing to revise their answers. If we take this at face value it means that in the incentivized and non-incentivized treatments that individuals believe that they have provided the answers they truly believe are most likely to be correct. Given that answers do shift between the unincentivized an incentivized treatments, this suggests that people are not consciously choosing to signal due to the lack of incentives. That is, the signaling that occurs in the unincentivized treatment may not be a strategic choice and people may not even be aware that they are doing it. Of course another possible explanation is that individuals did not want to go back and re-answer questions after a relatively long questionnaire. This issue is one that should be pursued in future tests.

1.5 Conclusion

The potential for political partisans to be operating under different factual understandings of the world around us is an important possibility to consider. There are definitely many accounts in the popular press and many behaviors we observe which might suggest that this is the case. Our findings suggest that while certainly not all partisans agree, any disagreements are not as large and as fundamental as often reported on in the popular press. What is accounting for this discrepancy is the fact that when responding to standard surveys and polls, people do not seem to actually provide answers based on what they know to be true. Instead they may be providing answers that flatter their political perspective. When there is no clear reason that they should provide accurate answers, why should we expect anything else?

[7]The high cutoff is for values of G on average above .75 while the low cutoff is for values on average less than 0.4.

It is important to note that there are other potential explanations of our findings. A commonly suggested one is that in the unincentivized cases people provide the answer they think is true. In the incentivized case people provide the answer they think that University based researchers will think is true. That is a possibility but one taken into account in our survey design. We have questions where the "correct" answer is one that might seem like an exaggerated one. We also have questions, such as about debt amounts and growth rates, where we are not aware of any substantial disagreements about how these are calculated. Of course we also have questions regarding issues that might be more subjective such as our question about the number of voter fraud cases. To help with this case, we identified in the question that the source of the "correct" answer was "according to a database maintained by the conservative leaning Heritage Foundation" to signal that we were not basing the answer on the findings of an organization often thought to be tied to Democrats. Despite these attempts to work against this concern, it is still one that must be considered. While it certainly could happen that when faced with incentives that people give the answers they think the researchers will find to be correct, this does not in any way suggest that in the case without incentives that they provide the answers they think are correct. In that case, it is unclear what motives might drive answers and the political signaling motive seems quite plausible.

Our study was intended to try to replicate findings from early papers and see if those findings held for issues that were the subject of strong political disagreements when our survey was conducted. Our perhaps surprising finding is that it was on exactly these current topics of substantial disagreement that the incentive effect was strongest. This is in part because these questions generated the largest partisan gaps in the unincentivzed cases and so there was the most room for movement. In the other cases, any partisan disagreement there may not be as focused and so the incentives for partisan signaling may have declined leading to our findings of inconsistent gaps without always substantial movements to close them.

We also had goals of testing some methodologies for furthering the design of surveys. This included the within subject test of incentives by asking people if they wanted to revise their answers once incentives were provided. Our findings here were that very few people chose to revise. This point warrants further investigation. In the future we intend to conduct a follow-up survey in which we ask a smaller number of questions where we expect strong partisan signalling and then ask about revision. If individuals are knowingly skewing their initial responses, they might be more willing to revise from a shorter survey. On this one, there were many questions without much signaling which diluted the impetus to revise.

In the end we see these results as a positive take on the current climate of polarization as they indicate that partisans may largely possess similar facts about important issues. On the other hand, these results show the clear danger of the common use of polls in the popular press. These surveys may well be worsening polarization by convincing people that the two sides are more different than they are. This is a problem which must be faced and as academics we need to provide journalists better tools for them to be able to provide truthful representations of the thoughts of voters. We see the present study as one of potentially many steps along the path to developing those tools.

	Question	No Incentives		Incentives		Correct
		Rep	Dem	Rep	Dem	
Replication	Unemployed	6.7***	5.4	5.8	5.8	4.8
Set	Trump Debt	22***	24.7	24	25	$27.5
	% in Poverty	12	10	11	12	11.4%
	Gas Price	3.27	3.24	3.31	3.25	$3.28
	Troops Afg	500	1750	2500	1750	70-80
	Foreign born	8	10	7.5	6.5	13.7%
	Biden vote %	52	52.5	52	53	51.3%
	Def Spending	20.35*	24.9	18.85*	20.2	11%
Misinformation	Border Pat	2.5	8	8	10	$4.07
Set	Gun Acc	3	5	4	4	2
	Voter Fraud	65	27	387	20	41
	Covid react	80	7	31	5	2-5
	MMR React	5	12	100	5	3.5-10
	Vaccine Eff	82.5***	94	90	92	95%
	Unvax in hosp	7.3	8.2	7.5	7	1.6%
	Vax in hosp	2.1***	1	1.4	1	.002%
	Obama Grow	2.6**	4	5	3.8	2.2%
	Trump Grow	5**	2.9	5.25**	3.3	2.5%
No MisInformtion	Social Security	16.9*	15.3	10.8	14.25	17%
Set	Debt Service	14.3	15.25	20.8**	11.55	5%
	Food stamps	11	8	10	5	$115
	Seat belts	80	80	83	80	45%

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

Table 1: Median responses to all questions by those self identifying as Republicans vs Democrats for the incentived and non-incentivized treatments. The stars represent a test of whether the Republicans and Democrats gave differences which were on average different using questions by question regressions of the form used for Table 2.

	(1) Replication	(2) Misinformation	(3) No Misinf	(4) Partisan Gap
Incentives	-0.008	0.002	0.009	-0.006
	(0.015)	(0.015)	(0.022)	(0.014)
Republican	-0.033	-0.087***	0.005	-0.122***
	(0.021)	(0.023)	(0.024)	(0.018)
Independent	-0.016	-0.065***	0.018	-0.059***
	(0.017)	(0.022)	(0.023)	(0.021)
Incent x Rep	0.009	0.034	-0.049	0.079***
	(0.029)	(0.032)	(0.033)	(0.027)
Incent x Ind	0.013	0.031	-0.008	0.035
	(0.023)	(0.027)	(0.031)	(0.025)
CRT	0.007	0.014**	-0.001	0.017***
	(0.006)	(0.006)	(0.008)	(0.005)
Text Box	-0.084***	0.217***	-0.258***	0.267***
	(0.010)	(0.010)	(0.015)	(0.014)
Constant	0.480***	0.597***	0.680***	0.584***
	(0.013)	(0.015)	(0.024)	(0.013)
Obs (Clusters)	1,800 (299)	2,007 (284)	910 (253)	1,785 (295)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2: OLS regressions with standard errors clustered on the individual respondents.

	(1) Replication	(2) Misinformation	(3) No Misinf
Incentives	-0.004	-0.014	0.027
	(0.022)	(0.019)	(0.041)
Republican	-0.141***	-0.173***	0.101*
	(0.029)	(0.030)	(0.054)
Independent	-0.055**	-0.098***	0.078*
	(0.028)	(0.033)	(0.045)
Incent x Rep	0.055	0.141***	-0.086
	(0.043)	(0.040)	(0.085)
Incent x Ind	0.031	0.050	-0.036
	(0.034)	(0.038)	(0.061)
CRT	0.016**	0.023***	0.003
	(0.008)	(0.007)	(0.014)
Text Box		0.212***	
		(0.015)	
Constant	0.506***	0.654***	0.643***
	(0.020)	(0.017)	(0.038)
Obs (Clusters)	717 (286)	850 (279)	218

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3: OLS regressions with standard errors clustered on the individual respondents except for the No Misinformation case as it contains only a single question. Here we limit to only include questions which have a significant partisan gap in the unincentivzed treatment.

	(1) I Don't Know
Incentives	-0.071
	(0.049)
Republican	-0.060
	(0.065)
Independent	0.001
	(0.055)
Incentives X Rep	-0.002
	(0.099)
Incentives X Ind	0.018
	(0.074)
CRT	-0.013
	(0.018)
Misinformation	0.095***
	(0.012)
No Misinfo	0.134***
	(0.014)
Textbox	0.096***
	(0.012)
Constant	0.254***
	(0.040)
Obs (Clusters)	6,405 (305)

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: OLS regressions with standard errors clustered on the individual respondents on whether a subject responded to a question with "I don't know".

	(1) Revise
Incentive	-0.014
	(0.038)
Republican	0.043
	(0.058)
Independent	-0.006
	(0.042)
High Avg G	-0.105
	(0.107)
Low Avg G	0.035
	(0.095)
CRT	-0.020
	(0.020)
Constant	0.148***
	(0.041)
Observations	301

Standard errors in parentheses

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

Table 5: OLS regression on whether a subject chose to reivse their answers for updraged incentives.

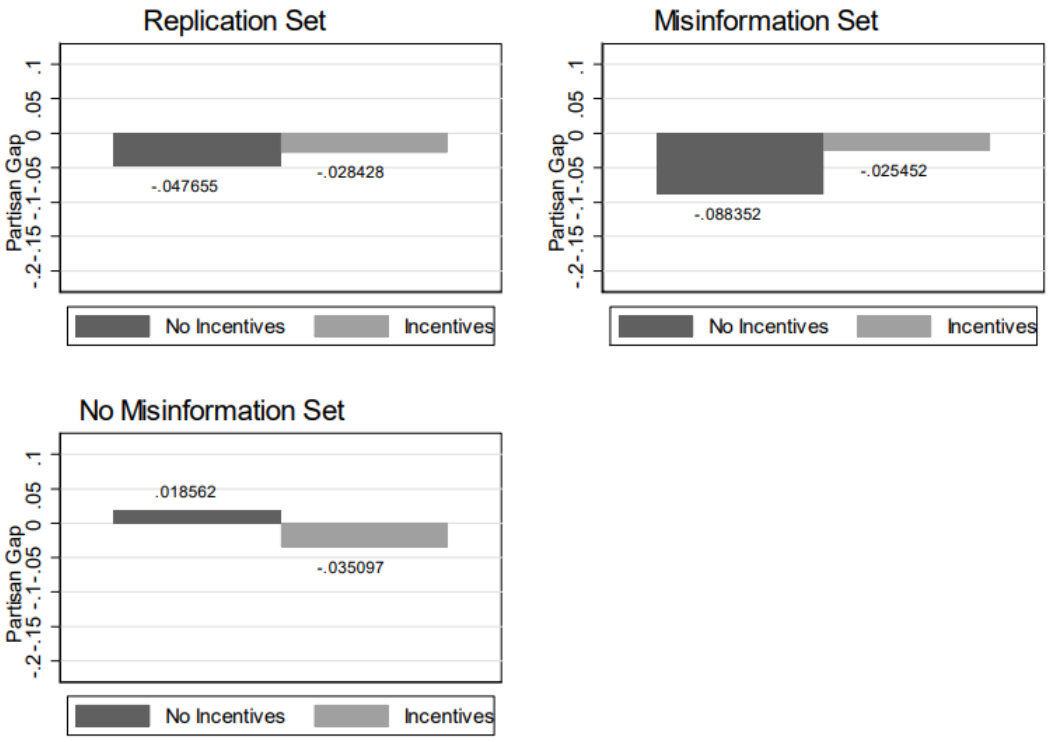

Figure 1: Average difference in Gij values between Republicans and Democrats for each set of questions in incentivized and non-incentivized treatments.

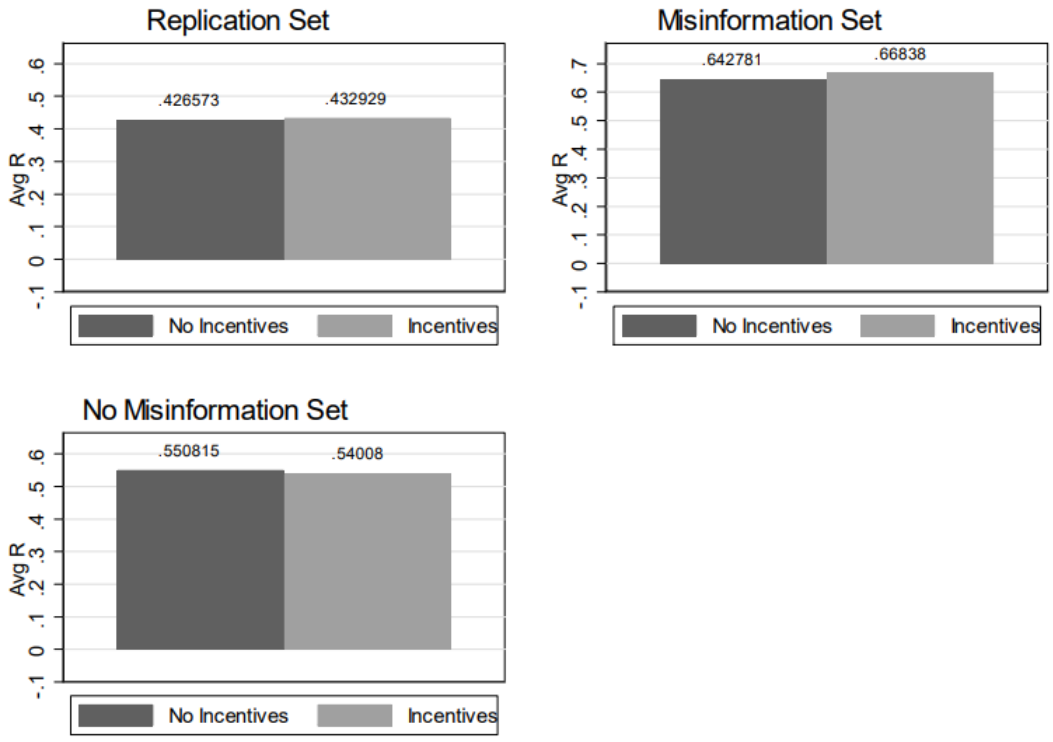

Figure 2: Average Gij values for Independents for each set of questions in incentivized and non-incentivized treatments.

CHAPTER 2

CONSPIRACIST IDEATION AND BEHAVIOR: EVIDENCE FROM AN INCENTIVIZED SURVEY

2.1 Introduction

While conspiracy theories have been an element of the American political sphere since the foundation of the country in the late 18th century, media coverage on issues relating to conspiracy theories has appeared to increase in recent years. An indication of this rise in media coverage can be found in data from the *Media Cloud* database which shows that the percentage of online news stories in the United States containing the phrase "conspiracy theory" has risen ten-fold from 2011 to 2021. While this statistic does not rigorously prove that belief in conspiracy theories is on the rise in the United States, it does provide evidence that there is a growing concern about the potential consequences of conspiracy theory belief. There does not exist clear time series data measuring belief in conspiracy theories and so proving that conspiracy theory belief is on the rise in the U.S. would be difficult, but we can address another important question regarding conspiracy theory belief: are there any real consequences of conspiracy theory belief? Putting the issue of tracking conspiracy theory belief across time aside, if conspiracy theory belief has real consequences, then we should be able to observe behavioral differences between those who are more likely to endorse conspiracy theories and those who are less likely to endorse conspiracy theories. The purpose of the present paper is to investigate whether there are associations between observed behavior and conspiracy theory belief.

For there to be real consequences of conspiracy theory belief, we need two things. First, there must be behavioral differences between those who are more likely to endorse conspiracy theories and those less likely to endorse conspiracy theories. Second, those behavioral differences must be relevant to contexts such as in the workplace. Regarding potential behavioral differences between those more likely to endorse conspiracy theories and others, there are reasonable claims in prior literature that those more likely to endorse conspiracy theories are less trusting, more entitled, more dishonest, more likely to rely on their own judgements about issues, and more creative. The rationale behind these claims is that these personal tendencies are more consistent with being a person who endorses conspiracy theories than not. In the academic literature, we can find studies backing the claims that those who endorse conspiracy theories are less trusting (e.g. [Goertzel, 1994], [Miller et al., 2016]), more narcissistic (e.g. [Cichocka et al., 2016], [Enders et al., 2021]), a character trait intertwined with entitlement (see [Krizan and Herlache, 2018]), more dishonest (e.g. [Jolley et al., 2019]), more likely to rely on their own judgements about issues (e.g. [Goertzel, 1994], [Harambam and Aupers, 2017]), and more creative (e.g. [Harambam and Aupers, 2017], [Bonetto and Arciszewski, 2021]). These academic studies are suggestive that conspiracy theory endorsers are different from others in the aforementioned ways, but we should point out that none of these studies observe incentivized behavior. This is an important methodological point, as there is a long-standing literature showing that responses to hypothetical scenarios do not always translate to incentivized behavior (e.g. [Bohm, 1972], [Neill et al., 1994], [Cummings et al., 1995], [Holt and Laury, 2002], [Holt and Laury, 2005]). More recently, [Esarey et al., 2012] showed how questions relating to a person's self-image (e.g. questions about how much you trust others) are specifically subject to this hypothetical bias. Therefore, if we are going to investigate whether there are real consequences of conspiracy theory belief, we need to be able to observe incentivized behavior in some fashion. The academic studies mentioned above point us towards a set of behavioral areas we need to observe to investigate the issue of whether there are real consequences of conspiracy theory belief. We will now highlight

how these suggested behavioral differences can have real consequences.

As an example of how these suggested behavioral differences can have real consequences, consider how employee behavior can influence the environment and productivity of the workplace. Employees who do not trust others are not likely to cultivate the relationships necessary to promote a positive and productive work environment (e.g. [Brown et al., 2015]). Trust has also been shown to be an important factor for broader economic growth (e.g. [Algan and Cahuc, 2010]). Furthermore, entitled employees can create a toxic work environment (e.g. [Harvey and Harris, 2010], [Harvey et al., 2014]) and dishonest employees are more likely to steal from the workplace (e.g. [Graham et al., 2002], [Coffin, 2003]). On a more positive note, employees who rely on their own judgements can help avoid group think and lead workplaces to adopting better practices than currently in place (e.g. [Sinaiko and Hirth, 2011], [Bhargava et al., 2017]). Furthermore, creative employees can help induce innovation in the workplace and consequently increase the productivity of their fellow employees (e.g. [Gong et al., 2013], [Zhou and Hoever, 2014]). If conspiracy theory endorsers are less trusting, more entitled, and more dishonest than others, then a surge of conspiracy theory belief among employees could be concerning for the quality of workplace environments we have in the United States. Of course, if conspiracy theory endorsers are also more likely to rely on their own judgements and are more creative, then the net impact of a rise in conspiracy theory belief on U.S. workplaces is ambiguous. In any case, if conspiracy theory endorsers are different from others in a majority of these behavioral areas, then a rise in conspiracy theory belief could have real consequences, positive and/or negative. What we need to know is whether any of these potential behavioral differences between conspiracy theory endorsers and others exist in reality. To answer this question, we need a methodology that can first give us a measure of who is a conspiracy theory endorser, and second allows us to observe the behavior of participants. A methodology with both of these primary elements will enable us to test whether conspiracy theory endorsers are behaviorally different from others in these aforementioned ways.

One approach to address this issue would be to obtain an instrument for conspiracy theory belief among employees in U.S. workplaces and field data measuring behavior for each of our relevant areas (e.g. rates of petty theft as a measure for dishonesty). However, while one can find field data measuring amount of petty theft in the workplace, it is more difficult to find field data which provides an incentivized measure of trust or employees' propensity to rely on their own judgements. Given this difficulty, another approach we can take is to design our own instrument that is able to both give us a measure for conspiracy theory belief and provide incentivized measures for each of our relevant behavioral areas.

To test for the potential behavioral differences between those more likely to endorse conspiracy theories and others, we constructed and implemented an online survey instrument. In this survey, subjects make decisions within a set of behavioral measures. The behavioral measures used are borrowed from previous research studies where possible, and are of our novel design where necessary. The key of the survey instrument is that subjects receive payment for the decisions they make. In this sense, we can observe subjects' behavior rather than just their responses to hypothetical situations where nothing is explicitly at stake. Along with these behavioral measures, we ask subjects a set of factual political questions as a way to measure for potential non-behavioral differences between conspiracy theory endorsers and others. Another important element of the survey instrument is a measure of subjects' belief in conspiracy theories. Measuring conspiracy theory belief for each subject is not as simple as asking subjects whether they think a given conspiracy theory is true (e.g. conspiracy theories about the assassination of JFK), as responses to such questions could depend on factors such as age rather than the subject's general tendency to endorse conspiracy theories. We need a measure of conspiracy theory belief that can more properly measure subjects' propensity to endorse conspiracy theories. Fortunately, there already exists a field validated mechanism for measuring subjects' general tendancy to endorse conspiracy theories: the generic conspiracist beliefs (GCB) scale from [Brotherton et al., 2013]. The set of questions we borrow from [Brotherton et al., 2013] are meant to identify "conspiracist

ideation" in subjects. We can define conspiracist ideation as "individual differences in the general tendency to engage with conspiracist explanations for events."[8] This concept is based on research (e.g. [Goertzel, 1994], [Wood et al., 2012], [Enders et al., 2021]) suggesting that individuals who endorse one conspiracy theory are more likely to endorse others, even if those other conspiracy theories are unrelated or directly contradictory to the original conspiracy theory. The questions from [Brotherton et al., 2013] do not ask about specific conspiracy theories, but rather ask about the general notions involved with most conspiracy theories (e.g. the government is involved in the murder of innocent citizens and/or well-known public figures, and keeps this a secret). With this measure for conspiracy theory belief, we can test whether those more likely to endorse conspiracy theories behave differently than those who do not generally endorse conspiracy theories.

Reference to contemporary dialogue and the existing academic literature lead us to the hypothesis that conspiracy theory endorsers behave differently from others in the areas regarding trust, dishonesty, entitlement, relying on one's own judgement, and creativity. Our main findings of interest are that, while conspiracy theory endorsers provide different answers to the factual political questions than others, they do not, in general, behave significantly differently from others in our incentivized tasks. In the following sections, we will go through the design and outcomes of our online survey instrument. In Section 2, we will discuss the methodology of the present study. In this section, we will provide details about each behavioral measure, including cases where we designed a novel measure, and state the hypotheses for each measure. In Section 3, we will briefly discuss the data collection process and provide some summary statistics of the data. In Section 4, we will present the results of the study. In Section 5, we will discuss the implications and interpretations of our results and provide concluding remarks.

[8] [Brotherton et al., 2013]

2.2 Methodology

To test whether conspiracy theory endorsers are different from others in the aforementioned behavioral areas, we need our survey instrument to include three main parts: first, a set of standard demographic questions to act as control variables for our regression analysis, second, a set of questions which allows us to appropriately identify who is a conspiracy theory endorser, and third, a set of behavioral measures which allow us to effectively observe subjects' behavior in the relevant areas.

As mentioned in the introduction, we borrowed our measure for conspiracy theory belief from [Brotherton et al., 2013]. To limit the length of the survey instrument, and to reduce potential experimenter effects, we selected a subset of eight questions from their original 15-item generic conspiracist beliefs (GCB) scale. We used questions that were most closely related to politics and political current events. For instance, we included questions that asked subjects about whether small, secret groups controlled world events/politics, but we did not include questions about extraterrestrials. This way, our conspiracy ideation questions can be framed as a continuation of a section asking subjects about political questions instead of a separate section blatantly asking subjects whether they are conspiracy theorists. The third part of our survey instrument is the incentivized tasks, which we will go through in the following paragraphs.

2.2.1 Interpersonal Trust

The first behavioral measure deals with interpersonal trust. To measure trust, we borrowed a version of the standard investment game from [Bohnet and Zeckhauser, 2004].[9] As we can see from the game tree in Figure 3, incentives are for Second Mover to play "End," unless he wants to reward First Mover for trusting him, in which case he should play "Continue." By backward induction, First Mover thus has incentive to play "End," unless she

[9] [Cox, 2004] points out that other-regarding preferences could also motivate First Mover to choose Continue. Ultimately, while other-regarding preferences might play a role, his results provide strong support to the conclusion that the standard investment game measures trusting behavior. Given its simplicity, we therefore utilize this version of the standard investment game as our measure of trusting behavior.

believes Second Mover is trustworthy and will reward her for trusting him, in which case she should play "Continue." In this sense, we can view a first mover choosing "Continue" as "trusting," and we can view a second mover choosing "Continue" as "trustworthy."

Whether an individual chooses "End" or "Continue" in the first mover position depends both on her preferences and her subjective beliefs about what the second mover will do. If we find any differences in behavior in the trust game between conspiracy theory endorsers and others, it would be useful to know how much subjects' subjective beliefs might be driving those differences. The design of [Andreoni and Sanchez, 2020] provides us an incentive compatible elicitation of subjects' beliefs in our trust game. After subjects submit their choice for the trust game, they receive the prompt for the belief elicitation task. For this task, there are two options: Option 1 and Option 2. For each option, subjects could earn either \$0.75 or \$0.25. In Option 1, the probabilities of each payment were predetermined, ranging from a 100% chance of \$0.75 and a 0% chance of \$0.25 in row 1, to a 0% chance of \$0.75 and a 100% chance of \$0.25 in row 11. In Option 2, the probabilities of each payment were determined by the results of the trust game. For the first movers, the probability of \$0.75 was the proportion of second movers who chose "Continue," and the probability of \$0.25 was the proportion of second movers who chose "End." For the second movers, the probability of \$0.75 was the proportion of first movers who chose "Continue," and the probability of \$0.25 was the proportion of first movers who chose "End."[10] For this task, subjects needed to choose when they would switch from preferring Option 1 to preferring Option 2. Figure 4 shows the table first movers saw when making their choice. Since the subjects did not know the results of the trust game, this task provides an incentive compatible elicitation of their beliefs. As an example, if a first mover thought 35% of second movers chose "Continue," then he would switch from preferring Option 1 to preferring Option 2 in row 8. For each subject, we randomly selected one of the rows to use to determine their payoff. For the row

[10]Before subjects are given the task, we ask them a comprehension question as a check to see if they understood the directions. We give them the correct answer to the comprehension question after they submitted their answer.

we selected, we used the option the subject indicated they preferred for that row. If in the example above we randomly selected row 6, then we used Option 1 to determine their payoff. In this case, the subject would have a 50% chance of receiving \$0.75, and a 50% chance of receiving \$0.25.

If conspiracy theory endorsers choose the trusting action (Continue) less often than others, then we can check the responses to the belief elicitation task to see if their subjective beliefs are a significant factor driving this difference. Related to subjective beliefs, [Douglas and Sutton, 2011] suggests that conspiracy theory endorsers engage in projection: they think others might engage in conspiracies because they themselves would engage in conspiracies if given the opportunity. This concept of projection suggests that if conspiracy theory endorsers choose "Continue" less often than others, then we might expect that conspiracy theory endorsers will switch from preferring Option 1 to Option 2 in our belief elicitation task later than others (i.e. conspiracy theory endorsers have more pessimistic beliefs about the trustworthiness of others).

2.2.2 Entitlement

The next behavioral measure deals with entitlement. We can define (psychological) entitlement as "a stable and pervasive sense that one deserves more and is entitled to more than others."[11] With this definition in mind, one way to measure entitlement is to utilize a task where a subject can choose to take some sort of prize even though the "rules" state that someone else should receive the prize. Such a task requires two important elements. First, the task needs a prize assignment rule that an entitled person would be willing to ignore. Second, this task needs to be able to distinguish between entitlement and standard selfish preferences. As an example of this distinction, consider a setting where a group of people are deciding on a rule to determine who gets the last slice of pizza from a luncheon. A selfish person would of course want to get the last slice, but as long as she knows that

[11] [Campbell et al., 2004]

the agreed upon rule is fair, she will accept the outcome of the rule. An entitled person, on the other hand, will only accept the outcome of the rule if he gets the last slice (he would take any outcome where he did not get the last slice as evidence that the rule was unfair). To the best of our knowledge, there is not a preexisting game which distinguishes between entitlement and standard selfish preferences. We therefore designed a novel game that makes this distinction for our measure of entitlement.

For the prize assignment rule of our game, subjects are first asked to tell us which of two paintings they prefer: Painting 1 (Kandinsky) and Painting 2 (Klee). We chose the Kandinsky and Klee painting task as our prize assignment rule because its original intent in [Chen and Li, 2009] was to create a assignment device that was random[12], but also felt nonrandom to subjects because they were making a choice based on their preferences. We paid subjects \$0.50 for stating their painting preference. Once a subject submitted their painting preference, we informed them that an additional \$1.00 prize was available to take. Subjects who chose Painting 1 were paired with subjects who chose Painting 2.[13] After subjects submitted their painting preference, we informed subjects who chose Painting 2 that the additional \$1.00 prize was intended for them to take, and we informed subjects who chose Painting 1 that the additional \$1.00 prize was not intended for them to take. Regardless of which painting they chose, each subject had the option to try to take the prize, or to defer it to the subject they were paired with. Figure 5 shows the game tree from the perspective of a subject who chose Painting 2. Subjects who chose Painting 2 (i.e. "winning" subjects) effectively go first in this game, so they make a single choice between trying to take the prize and deferring the prize to the other subject. Subjects who chose Painting 1 (i.e. "losing" subjects) effectively go second in this game. Because we are using

[12]In our sample of 303 subjects, 49% chose Painting 1 (148) and 51% chose Painting 2 (155).

[13]There were seven more subjects who chose Painting 2 (155) compared to Painting 1 (148). So, we had 148 pairs, each with one subject who chose Painting 1 and one subject who chose Painting 2. The seven leftover subjects who chose Painting 2 were each matched with a subject who chose Painting 1. For these seven, the matched actions only affected their own payoff. For the 148 pairs, the paired actions affected both subjects' payoffs. The instructions given to subjects were constructed to allow for this multiple pairing.

an asynchronous online survey to observe subjects' behavior, we cannot simply ask losing subjects whether they want to take the prize or defer it back to the winning subject. If we only asked losing subjects this single question, then their subjective beliefs about what the winning subject will do could be relevant to their decisions. This would make it difficult for us to distinguish between standard selfish preferences and entitlement. To avoid this issue, we use the strategy-method where losing subjects have two choices to make: whether to take or defer when the other subject chose to take, and whether to take or defer when the other subject chose to defer. The other main difference between the choices of a winning subject and a losing subject is that a losing subject must give up the \$0.50 we paid them to state their painting preference if they want to take the prize. A winning subject does not have to give up anything to take the prize.

As we can see in the game tree, a losing subject who only cares about expected payoffs will only choose to take if the winning subject he is paired with chooses to defer. If the winning subject chooses to take, the losing subject only makes \$0.33 in expectation if he also chooses to take, so he is better off deferring and earning \$0.50. For an entitled losing subject, his utility from winning the \$1 prize may be greater than the consumption value of \$1, so the expected utility of taking might be greater than the expected utility of deferring. Furthermore, an entitled losing subject could also gain utility from changing the rules in his favor. Instead of painting preferences choosing who should get the prize, which a losing subject might have qualms with, a "truly" random process gives each subject equal chance of (not) getting the prize. Such motivations would not appeal to a selfish subject, he will try to earn as much as he can, but such motivations might appeal to someone with the trait of (psychological) entitlement.

2.2.3 Dishonesty

The next behavioral measures deal with dishonest behavior. To measure dishonest behavior, we need a set of incentivized tasks that requires subjects to make a choice between being honest and maximizing their monetary payoff. We use two such tasks to measure

dishonesty, one of which was of our own design based on a relatively unknown quirk about United States' area codes. Our first lying task is a novel design where we ask subjects to provide the middle digit of the first area code they remember using in the United States. Subjects were paid \$0.10 plus \$0.10 times the number they enter. A feature of U.S. area codes is that none have "9" as their middle digit. As such, any subject who reports "9" as the area code's middle digit is definitely lying. Meanwhile, since all subjects have at least one area code they remember using, each subject can tell the complete truth if they desire.

We also use a version of the classic coin flip experiment (e.g. [Jacquemet et al., 2021], [Dickinson and McEvoy, 2021]) as a second lying task. We ask subjects to find a fair coin and flip it five (5) times. We then ask subjects to report the number of times a flip came up as heads instead of tails. Subjects were paid \$0.10 for each time they report a coin flip came up heads. We obviously cannot verify that each subject took the time to do all five coin flips, nor can we verify the outcomes of the coin flips. As such, subjects are free to lie and say that they flipped five heads. However, although most subjects are probably lying in the sense that they did not take the time to flip a coin five times, some subjects might not want to come off as a selfish liar, so they might only report flipping heads two or three times. So, even though most subjects are likely lying by default, this task gives us another measure about the extent to which subjects are willing to lie.

We utilize the standard coin flip task because it has been verified as a measure for dishonesty in previous studies, but there are issues that arise when using it in an online survey. Because some of the subjects might not have a fair coin readily available when taking our survey, these subjects will be lying by default. Given this possibility, we included our novel area code question to make sure that we have a task that allows subjects to be completely honest if they so desire. Also, having two tasks for measuring dishonesty allows us to have provide subjects a higher guaranteed base payment for taking the time to complete our survey.

2.2.4 Propensity to Rely on Own Judgement

The next task deals with one's propensity to rely on their own judgement. In the words of a conspiracy thinker, we can conceptualize this propensity as an individual's commitment "to look at things from multiple perspectives, to consult multiple sources, but mostly to think for yourself and be able to adjust previously held convictions."[14] So, when we refer to a propensity to rely on one's own judgement, we mean one's general propensity to test and explore things for herself rather than simply taking someone's word or advice. To measure this, we need a task where subjects are given incentive to make the "correct" choice among a set of options, about which they can choose to either take the advice of someone else, or to gather information about the options himself or herself. Although there are preexisting tasks that address the issue of advice giving, to the best of our knowledge, there are not any preexisting tasks designed specifically to address this issue of relying on one's own judgement. We therefore designed a novel task appropriate for an asynchronous online survey which identifies a general propensity to explore things for oneself.

The novel task we designed is based on a modified version of a two-armed bandit game. Subjects have the choice between two income generating options, labeled "Option A" and "Option B," respectively. Subjects know that the possible payoffs are \$0.25, \$0.50, \$0.75, and \$1.00, but they do not know the distribution of payoffs for each option nor the expected payoff of each option. Option B has an expected payoff \$0.25 higher than Option A. Subjects are informed that one of the options has an expected payoff \$0.25 higher than the other, but they are not told which one. Subjects in the main sample were given two opportunities to help them pick between the two income earning options. For the first opportunity, subjects could conduct 10 trial rounds where they could choose either Option A or Option B and observe the outcome of each of the draws. For the second opportunity, subjects could instead accept the advice of a random previous participant who had already conducted 10 trial rounds in a pilot survey. Subjects who opted to conduct the 10 trial rounds themselves had to pay a

[14] [Harambam and Aupers, 2017]

monetary price for the opportunity, while subjects who opted to accept a previous subject's advice did not have to pay any monetary price. Subjects were randomly given one of three treatments: a *low cost* treatment where the price to conduct 10 trial rounds was \$0.01, a *medium cost* treatment where the price to conduct 10 trial rounds was \$0.05, and a *high cost* treatment where the price to conduct 10 trial rounds was \$0.10. A rational, risk-neutral, money-maximizing agent will choose to pay the testing cost if the following inequality holds:

$$p_i - p_a > 4c$$

where p_i is the agent's subjective belief that she will choose the correct option (Option B), p_a is her subjective belief that the advice is correct, and c is the cost to test the options. At the low, medium, and high cost conditions, she must believe that she is at least 4%, 20%, and 40%, respectively, more likely to choose the correct option than the average pilot subject to decide to test the options herself. If a subject had the most pessimistic belief about the accuracy of advice (i.e. 50% chance of correct advice), then she must believe that she is at least 54%, 70%, and 90% likely to choose the correct option to be willing to pay the low, medium, and high cost, respectively. There is also a time cost involved in testing the options for oneself. Subjects can finish the survey sooner and spent their time doing other activities if they choose to simply take the advice of someone else. Given the high belief requirements and additional time costs, a subject would only test the options for herself if she is both highly confident in her ability to choose the higher paying option and thinks that going through the 10 testing rounds is worth her time.

In the pilot survey, $\frac{2}{3}$ of subjects recommended Option B and $\frac{1}{3}$ recommended Option A. If a subject knew that the true probability of correct advice was $\frac{2}{3}$, then she could not justify testing the options herself at the high cost, unless she placed high importance on testing things for herself. Therefore, our design not only allows us to test whether conspiracy endorsers have a higher propensity to rely on their own judgement on average, but also the strength of their propensity if it exists.

2.2.5 Creativity

To measure creativity, we want an incentivized task that can elicit responses from subjects such that we can judge their responses based on how novel and useful they are. Fortunately, there already exists such a measure specifically designed for this purpose, namely the Creative Uses Task. We use the standard Creative Uses Task which dates back to [Torrance, 1966]. We used the instructions and procedures of [Dutcher and Rodet, 2018] as the template for our design. On the first screen, we told the subjects that they would have three minutes to come up with alternative uses of an object to be displayed on a following screen. Using a tin can as an example, we gave the subjects potential uses that would and would not receive credit. We also gave the subjects instructions for how they should enter their responses in Qualtrics.[15] On the second screen, we explained to them how they would be paid for this task. We paid subjects \$0.15 for each valid use they provided. A valid use is any plausible use different from the object's primary intended use. Furthermore, we paid subjects an additional \$0.15 for each valid use that was unique among a set of 10 subjects.[16] We also told subjects that their responses would be judged by two independent judges (not the experimenters) to determine whether they satisfy the requirements. On the task screen, subjects were told that their object was an extension cord.[17] Subjects entered their responses into a text box, and they could either click the "next" button to continue when they were finished or be automatically moved to the next task when the three-minute timer hit zero.

2.2.6 Factual Political Questions

The last part of our survey design to discuss is our set of factual political questions. These factual political questions act both as a segue between the demographic questions

[15]We asked the subjects a comprehension question about these directions as a check to see if they were reading carefully. We gave subjects the correct answer to the comprehension question on the following screen.

[16]Since there were 303 subjects, three of the groups had 11 subjects.

[17]Of the objects used in [Dutcher and Rodet, 2018], extension cord was the only one where a Google search for creative uses did not immediately provide a list of valid uses. Note that [Dutcher and Rodet, 2018] was conducted in a laboratory setting, so they did not need to worry about subjects trying to do a Google search for creative uses.

and our conspiracist ideation questions, and as an additional behavioral measure for differences between conspiracy theory endorsers and others. These four political questions ask subjects about facts, such as how Joe Biden's 100-day approval rating compared with the 100-day approval ratings of George W. Bush and Donald Trump. Three of these factual political questions were either directly borrowed from or based on questions from [Robbett and Matthews, 2018]. We constructed the fourth factual political question in the survey instrument ourselves. We normalize responses to these questions such that providing the most democratic-friendly answer is scored as "0," so a low sum across the four questions reflects a bias towards democrats and a high sum reflects a bias towards republicans. As an example, we asked subjects how much of the promised border wall was built during the four years of the Trump administration. Subjects were given five options: less than 150 miles, between 150 and 300 miles, between 300 and 450 miles, between 450 and 600 miles, and more than 600 miles.[18] So, scores for this question could range from "0" if a subject answered with "less than 150 miles," to "5" if a subject answered with "more than 600 miles."

As mentioned above, these four factual political questions not only act as a segue between our demographic questions and our conspiracist ideation questions, but they also provide us an additional, non-behavioral, test for differences between conspiracy thinkers and others. The political mindset of a conspiracy endorser may be different from someone who does not endorse conspiracy theories. For instance, a conspiracy endorser might believe that the mainstream consensus on a relevant issue is incorrect, while others would generally have opinions more in line with the mainstream consensus. As such, we might expect that conspiracy endorses will provide answers that are more extreme than those who generally do not endorse conspiracy theories. For our measure of this, we take the sum of each subject's responses to the four factual political questions. So for the purposes of our survey instrument, our expectation is that the total answers for conspiracy endorsers will be significantly higher (or lower) than the total answers for others.

[18]The correct answer was "between 450 and 600 miles."

2.2.7 Hypotheses

Before we discuss our data and results, we will now explicitly state the hypotheses our survey instrument was designed to test. These hypotheses reflect the suggestions from the prior literature on how conspiracy thinkers might behave.

Hypothesis 1 *(Trust) Conspiracy theory endorsers in the position of "First Mover" will choose "Continue" less often than others.*

This hypothesis is based on the expectation that conspiracy theory endorsers are less trusting than others. Since a first mover choosing "Continue" in our trust game indicates trusting behavior, we therefore expect conspiracy theory endorsers to choose "Continue" less often than others.

Hypothesis 2 *(Trustworthiness) Conspiracy theory endorsers in the position of "Second Mover" will choose "Continue" less often than others.*

Hypothesis 3 *(Subjective Beliefs) Conspiracy theory endorsers' elicited beliefs about the percentage of second movers choosing "Continue" will be lower than the elicited beliefs of others.*

These two hypotheses are based on the expectation that conspiracy theory endorsers will engage in projection. Assuming that conspiracy theory endorsers are less trusting than others, projection first requires that conspiracy theory endorsers are less trustworthy than others, and second that conspiracy theory endorsers are more pessimistic about the trustworthiness of others than others are. Since a second mover choosing "Continue" in our trust game indicates trustworthy behavior, we therefore expect conspiracy theory endorsers in the position of "Second Mover" will choose "Continue" less often than others, and furthermore that conspiracy theory endorsers' elicited beliefs about the percentage of second movers choosing "Continue" will be lower than the elicited beliefs of others.

Hypothesis 4 *(Entitlement) Conspiracy theory endorsers who picked Painting 1 (i.e. a "losing subject") will be more likely to take the* $1 *prize than others.*

This hypothesis is based on the expectation that conspiracy theory endorsers are more entitled than others. Since a losing subject choosing to always take the $1 prize in our entitlement task indicates entitlement, we therefore expect conspiracy theory endorsers who picked Painting 1 will choose to always take the $1 prize more often than others.

Hypothesis 5 *(Dishonesty 1) Conspiracy theory endorsers will report "9" as their area code's middle digit more often than others.*[19]

Hypothesis 6 *(Dishonesty 2) Conspiracy theory endorsers will report a higher average number of heads than others.*

These hypotheses is based on the expectation that conspiracy theory endorsers are more likely to engage in dishonest behavior than others. Since reporting a high number of heads or reporting "9" as one's area code middle digit is an indication of dishonesty, we therefore expect conspiracy theory endorsers will report a higher average number of heads than others and conspiracy theory endorsers will report "9" as their area code's middle digit more often than others.

Hypothesis 7 *(Relying on Own Judgement) Conspiracy theory endorsers will choose to test the options more often than others.*

This hypothesis is based on the expectation that conspiracy theory endorsers are more likely to rely on their own judgements than others are. Since choosing to test the options in our modified two-armed bandit game is an indication of relying on one's own judgement, we therefore expect conspiracy theory endorsers will choose to test the options more often than others.

[19]We could also test for the average area code response. The issue with this test, however, is that conspiracy theory endorsers could have randomly lived in locations with higher/lower area code middle digits, so the average area code response might not be an informative measure of dishonesty.

Hypothesis 8 *(Creativity) On average, conspiracy theory endorsers will provide a greater number of valid/unique uses than others.*

This hypothesis is based on the expectation that conspiracy theory endorsers are more creative than others. Since providing a high number of valid/unique uses in the Creative Uses Task is an indication of creativity, we therefore expect that conspiracy theory endorsers will provide a greater number of valid/unique uses than others.

Hypothesis 9 *(Political Bias) The average answer conspiracy theory endorsers give to the factual political questions will be higher (or lower) than the average answers of others.*

This hypothesis is based on the expectation that conspiracy theory endorsers will provide more extreme answers to the factual political questions than others. Since we expect that conspiracy theory endorsers are of a different political mindset than others, we expect to see conspiracy theory endorsers provide answers further away from the "mainstream consensus" (i.e. more extreme answers).

2.3 Data

303 subjects participated in our online survey in September 2021. Subjects were recruited via the online recruiting service, *Prolific Academic*. A breakdown of the summary statistics for our sample can be found in Table 6. For our specification, we group subjects into one of three categories based on their score on the conspiracy ideation questions: low score (0-9), medium score (10-17), and high score (18-32). The score cutoffs were selected so that each category had as close to one-third of the subject sample as possible, without having any overlap in scores. Looking at Table 6, we see that the main demographics that vary some across categories are political affiliation, education, and age. Such variation is in line with prior literature on the demographics of conspiracy thinkers. Regarding political affiliation, as we have previously discussed, the existing literature (e.g. [Douglas et al., 2019], [Smallpage et al., 2020]) suggests that identifying as an independent tends to be positively correlated with being a conspiracy thinker. Furthermore, this literature suggests that those on the

losing end of the political process (e.g. in elections) tend to be more likely to endorse conspiracy theories. As such, we are not surprised by the negative correlation between the conspiracy ideation score and identifying as a democrat, nor with the positive correlation between the conspiracy ideation score and identifying as a republican or independent. The aforementioned literature also suggests that there is a negative correlation between conspiracy ideation scores and education, which we see in our data. Finally, [Galliford and Furnham, 2017] suggests that conspiracy thinkers tend to be younger, which is reflected in the median age of those in each of our categories.

2.4 Results

Recall that our fundamental question is whether there are general behavioral differences between conspiracy thinkers and others. To address this question, we want to first summarize the outcomes of our behavioral measures across conspiracist ideation scores. This will give us the broad picture of how conspiracy thinkers behave compared to others. In Table 7, we summarize the outcomes of the factual political questions and the incentivized tasks. Given demographic differences by conspiracist ideation score, we will ultimately need to conduct regression analysis to test for differences, controlling for demographic variables. That said, the information in Table 7 gives us an indication about the structure of our data. In the first row, we report the average total response in the factual political questions for the three conspiracy score categories. As we can see, those with low scores on the conspiracy ideation questions tended to give answers that were more congenial towards democrats, while those with high scores tended to give answers that were more congenial towards republicans. On average, those with medium scores did not tend to give answers that favored either democrats or republicans. Looking at the outcomes for the incentivized tasks, we notice that for the most part, there is not much of a difference between the average outcomes of those with low scores and those with high scores. There are a couple of cases, specifically the outcomes for the Creative Uses Task and the responses to the area code question, where there could

possibly be a significant difference, but overall the outcomes look similar. So, the summary of our measures suggests that if there are any general behavioral differences between conspiracy theory endorsers and others, these differences are small and seemingly not economically significant, contrary to the majority of our hypotheses. Now that we have the broad picture that conspiracy theory endorsers do not seem to have significant behavioral differences from others, we will run regressions with control variables to make sure this suggestion from our summary holds up to more careful analysis.

In our regression analysis, we want to run regressions for each of our stated hypotheses. The structure for each of our regressions is the same, the only thing changing across regressions is the dependent variable (and as a consequence, whether we utilize an OLS regression or a logit/probit regression). For these regressions, we utilize the following set of explanatory variables. Our main variables of interest are indicator variables for having a medium conspiracy score and having a high conspiracy score. We therefore use those with a low conspiracy score as the reference group. The control variables we include are indicator variables for identifying as a republican and identifying as an independent, an indicator variable for having at least a Bachelor's degree, the subject's self-reported risk preferences, and the subject's age. We chose these as our set of control variables because these were the demographics that differed most across conspiracy scores. Therefore, if the coefficients on the conspiracy score indicator variables are significantly different from zero, we would have indication of a pure effect of being a conspiracy theory endorser rather than the effect of a (negatively) correlated demographic like education or age. We will now go through the result statements for each of our hypotheses.

Result 1 *(Trust) Conspiracy theory endorsers in the position of "First Mover" do not choose "Continue" less often than others.*

In column (2) of Table 8, we have the results of the probit regression for the behavior of first movers in our trust game. As we can see, the coefficients on our conspiracy variables are

not statistically significant, suggesting that conspiracy theory endorsers were not significantly less trusting than others, which is contrary to our hypothesis.

Result 2 *(Trustworthiness) Conspiracy theory endorsers in the position of "Second Mover" do not choose "Continue" less often than others.*

In column (3) of Table 8, we have the results of the probit regression for the behavior of second movers in our trust game. As we can see, the coefficients on our conspiracy variables are not statistically significant, suggesting that conspiracy theory endorsers were not significantly less trustworthy than others, which is contrary to our hypothesis.

Result 3 *(Subjective Beliefs) Conspiracy theory endorsers' elicited beliefs about the percentage of second movers choosing "Continue" is not lower than the elicited beliefs of others.*

In columns (4) and (5) of Table 8, we have the results of the OLS regressions for the belief elicitation task for our trust game. As we can see, the coefficients on our conspiracy variables are not statistically significant, which suggests that conspiracy theory endorsers did not have significantly more pessimistic beliefs than others, contrary to our hypothesis.

Result 4 *(Entitlement) Conspiracy theory endorsers who picked Painting 1 (i.e. "losing subjects") are not more likely take the* \$1 *prize than others.*

In column (6) of Table 8, we have the results of the probit regression for the behavior of losing subjects in our entitlement game. As we can see, the coefficients on our conspiracy variables are not statistically significant, suggesting that conspiracy theory endorsers were not significantly more entitled than others, which is contrary to our hypothesis.

Result 5 *(Dishonesty 1) Conspiracy theory endorsers do not report "9" as their area code's middle digit more often than others.*

Result 6 *(Dishonesty 2) Conspiracy theory endorsers do not report a higher average number of heads than others.*

In column (1) of Table 9, we have the results of the probit regression for the behavior of subjects in our area code task. In column (2) of Table 9, we have the results of the OLS regression for the behavior of subjects in our coin flip task. As we can see for both tasks, the coefficients on our conspiracy variables are not statistically significant, suggesting that conspiracy theory endorsers were not significantly more dishonest than others, which is contrary to our hypothesis.

Result 7 *(Relying on Own Judgement) Conspiracy theory endorsers do not choose to test the options more often than others.*

In column (3) of Table 9, we have the results of the probit regression for the behavior of subjects in our modified two-armed bandit game. As we can see, the coefficients on our conspiracy variables are not statistically significant, suggesting that conspiracy theory endorsers were not significantly more likely to rely on their own judgements compared to others, which is contrary to our hypothesis. In light of this null result, we further test whether, for those who did test the options for themselves, conspiracy theory endorsers were more likely to choose the correct option (Option B). The results of this probit regression are in column (4) of Table 9. As we can see, conspiracy theory endorsers who tested the options for themselves were not more likely to choose the correct option.

Result 8 *(Creativity) On average, conspiracy theory endorsers provide fewer valid/unique uses than others.*

In columns (5) and (6) of Table 9, we have the results of the OLS regressions for the behavior of subjects in the Creative Uses Task. As we can see, the coefficients on our conspiracy theory endorsers variables are in fact statistically significant, but they are negative, suggesting that conspiracy theory endorsers were *less* creative than others, which is contrary to our hypothesis that conspiracy theory endorsers would be *more* creative than others.

Result 9 *(Political Bias) The average answer conspiracy theory endorsers give to the factual political questions is higher than the average answers of others.*

In column (1) of Table 8, we have the results of the OLS regression for our factual political questions. As we can see, even when controlling for political affiliation, conspiracy theory endorsers provide answers to the factual political that are a higher than others, which matches with our hypothesis.

2.5 Conclusion

Given the concern present in media pieces about conspiracy theory belief in the United States, we need to better understand whether conspiracy theory endorsers are behaviorally different from others as prior authors have claimed, as such behavioral differences could have real consequences such as in the workplace. In this study, we implemented an incentivized online survey designed to test for behavioral differences between conspiracy theory endorsers and others in the areas of interpersonal trust, (psychological) entitlement, dishonesty, propensity to rely on one's own judgement, and creativity. To be able to test for behavioral differences in some of these areas, namely the areas of entitlement, dishonesty, and the propensity to rely on one's own judgement, we needed to construct novel measures of our own design. To measure for entitlement, we constructed a dynamic game where a "losing" subject could choose to take a prize that subjects were told was intended for a "winning" subject. A crucial feature of this game is that it can distinguish between standard selfish preferences and entitlement. To measure for dishonest behavior, on top of a standard coin-flip task, we constructed a second task based on a little known feature of U.S. area codes, namely that no U.S. area codes have "9" as the middle digit. Since subjects were likely lying on the coin-flip task by default, having this second task allowed subjects to be completely honest if they desired. To measure for a propensity to rely on one's own judgement, we constructed a modified version of the two-armed bandit game where subjects could either accept the advice of someone else, or pay a small fee to test the options for themselves. As a way to test for the strength of a propensity to rely on one's own judgement, *if* one exists, we randomized the size of the fee subjects had to pay.

From our data, we found that while conspiracy theory endorsers tended to give more extreme answers to the factual political questions, as we hypothesized, they did not generally behave differently from others in our incentivzed tasks. Table 10 provides a summary of how our hypotheses compared to our results. For each of our behavioral measures, we hypothesized that conspiracy theory endorsers would be significantly different from others. For instance, we hypothesized that conspiracy theory endorsers would be less (-) trusting than others, and that conspiracy theory endorsers would be more (+) dishonest than others. As we can see from Table 10, for the majority of our behavioral measures, we instead found a null result (0). These results are striking considering that our expectations, which were backed by the existing literature regarding the behavior of conspiracy theory endorsers, were that conspiracy theory endorsers would significantly differ in each of our behavioral measures. The primary implication of our findings is that, when it comes to behaviors relevant to the workplace and broader economic growth, if there are any general behavioral differences between conspiracy theory endorsers and others, these differences are small and seemingly not economically significant. As such, conspiracy theory belief might not have much direct impact on outcomes in the workplace or broader economic growth. To be clear, we are not claiming that a growth in conspiracy theory belief would be inconsequential to American society, only that conspiracy theory endorsers do not seem to have large behavioral difference from others that would be directly pertinent to the workplace or broader economic growth. With this implication in mind, our results point us towards a couple of avenues for further research into the behavior of conspiracy theory endorsers.

One possible avenue for further investigation relates to the creativity of conspiracy theory endorsers. Although we found a null result for the majority of our behavioral measures, one case where the coefficients were statistically significant was the Creative Uses Task. The first thing to point out here, as we can see in Table 10, is that our hypothesis was that conspiracy theory endorsers would be *more* (+) creative than others. However, the coefficients on the conspiracy scores for this task were negative. So, relative to our expectation, we fail to reject

the null hypothesis that conspiracy theory endorsers are not *more* creative than others. This is similar to the conclusions we make from the results for the other behavioral measures. However, since the coefficients are statistically significant, there is potentially a stronger interpretation that conspiracy theory endorsers are *less* creative than others. Under the assumption that this stronger interpretation is correct, we can first point out how the creative uses task is slightly different from the other behavioral measures. While for the other behavioral measures we are measuring whether a subject takes a particular action (e.g. whether a subject chooses "End" or "Continue" in the trust game), for the Creative Uses Task, we are measuring a subject's ability. Related to the ability of conspiracy thinkers, [Brotherton and French, 2014] suggests that conspiracy theory endorsers are more likely to commit the conjunction fallacy than others. So, there is an existing literature suggesting that the ability of conspiracy theory endorsers might be different from others. Having said that, we should not overemphasize this single set of significant coefficients. Perhaps these significant coefficients are meaningful, but the appropriate conclusion from our results is that the general behavior of conspiracy theory endorsers does not tend to widely differ from the general behavior of others in these behavioral areas relevant to the workplace and broader economic growth. To be able to say more about the creativity of conspiracy theory endorsers, we need to conduct further study directly focused on testing for the creativity of individuals.

Another potential avenue for future research is to investigate whether conspiracy theory endorsers might be different from others in the behavioral areas we studied, focusing on specific contexts rather than general behavior. For instance, our study does not find any large differences in general trust between conspiracy theory endorsers and others, but the notion of a difference is still intuitively appealing. One possibility is that while conspiracy theory endorsers might not be less trusting in general, they might be less trusting of specific types of people. One could conduct a separate study focused on testing for differences in trust between conspiracy theory endorsers and others in specific contexts. We also do not find that conspiracy theory endorsers have a higher general propensity to rely on their own

judgement, but perhaps conspiracy theory endorsers are more likely to rely on their own judgement in specific contexts (e.g. politically charged contexts like climate change). One could also conduct a separate study where we place subjects into stylized contexts to test whether conspiracy theory endorsers might have a higher propensity to rely on their own judgements in specific contexts. What our study contributes to this literature on the behavior of conspiracy theory endorsers is to narrow the possibility of economically significant behavioral differences into smaller categories. Our study suggests that if there are any general differences between conspiracy theory endorsers and others in these behavioral areas, they are small and not economically significant, so we can approach the next step of testing for behavioral differences in specific cases if we have good reason to think conspiracy theory endorsers will differ from others in those specific contexts.

	Low Score (0-9)	Medium Score (10-17)	High Score (18-32)
Democrat	68	50	45
Republican	13	16	23
Independent	22	29	37
Bachelor's (or more)	65	48	49
Income ($100,000+)	35	22	27
Employed	62	57	66
Male	50	46	52
White	77	61	69
Urban	77	70	75
Median Age	54	41	35
N	103	95	105

Table 6: Demographic Summary Statistics by Conspiracy Score

Conspiracist Ideation Score	Low (0-9)	Medium (10-17)	High (18-32)
Mean Total Political Question Response	5.57	7.00	8.01
% FM Choosing Continue (Trust)	56.60%	73.47%	65.38%
% SM Choosing Continue (Trust)	72.00%	54.35%	75.47%
FM Median Belief about SM (Trust)	50-60%	60-70%	50-60%
SM Median Belief about FM (Trust)	50-60%	50-60%	50-60%
% Always Take (Entitlement)	16.36%	15.22%	23.40%
% Reporting "9" (Dishonesty)	4.85%	6.32%	10.48%
Mean Reported # of Heads (Dishonesty)	2.80	2.81	2.70
% Choosing to Test (Own Judgement)	23.30%	24.21%	28.57%
% Tester Chose Correct (Own Judgement)	60.87%	70.83%	63.33%
Average # of Valid Uses (Creativity)	6.02	4.34	4.31
Average # of Unique Uses (Creativity)	2.98	1.86	1.92
Average Earnings	$4.71	$4.42	$4.49

Table 7: Behavioral Measures Summary

VARIABLES	(1) Political Bias	(2) Trust	(3) Trustworthy	(4) FM Belief	(5) SM Belief	(6) Entitled
Medium Score	1.161***	0.358	-0.538*	-0.695	-0.131	-0.017
	(0.389)	(0.273)	(0.282)	(0.473)	(0.521)	(0.327)
High Score	1.945***	-0.028	0.090	-0.442	-0.620	0.345
	(0.397)	(0.276)	(0.294)	(0.494)	(0.523)	(0.327)
Republican	3.337***	-0.063	0.432	1.276**	1.430**	-0.103
	(0.436)	(0.292)	(0.346)	(0.512)	(0.602)	(0.331)
Independent	0.944***	-0.029	-0.188	0.433	0.547	-0.732**
	(0.362)	(0.268)	(0.257)	(0.463)	(0.476)	(0.342)
Bachelor's (or more)	-0.520	0.232	0.072	0.275	0.691	0.064
	(0.323)	(0.217)	(0.260)	(0.377)	(0.470)	(0.242)
Risk Preferences	0.068	0.051	-0.111**	-0.001	0.198*	0.063
	(0.071)	(0.047)	(0.057)	(0.083)	(0.103)	(0.060)
Age	0.009	-0.018***	-0.011	-0.019	-0.020	0.005
	(0.010)	(0.007)	(0.008)	(0.012)	(0.014)	(0.008)
Observations	303	154	149	154	149	148

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

Table 8: Political Bias, Interpersonal Trust, and Entitlement

VARIABLES	(1) Lie Area	(2) Lie Coin	(3) Own Judge	(4) Correct Judge	(5) Creative Valid	(6) Creative Unique
Medium Score	0.071	-0.007	0.035	0.149	-1.582***	-1.122***
	(0.313)	(0.145)	(0.205)	(0.432)	(0.524)	(0.314)
High Score	0.251	-0.142	0.175	-0.263	-1.558***	-1.065***
	(0.306)	(0.148)	(0.207)	(0.424)	(0.536)	(0.321)
Republican	0.173	-0.111	0.175	-0.512	0.163	0.005
	(0.320)	(0.163)	(0.218)	(0.399)	(0.588)	(0.352)
Independent	-0.198	-0.240*	0.015	0.624	1.662***	0.893***
	(0.293)	(0.135)	(0.191)	(0.441)	(0.488)	(0.293)
Bachelor's (or more)	0.709**	0.073	-0.092	-0.121	0.086	0.165
	(0.275)	(0.121)	(0.168)	(0.341)	(0.436)	(0.261)
Risk Preferences	0.032	0.073***	0.057	0.131	0.011	0.059
	(0.057)	(0.026)	(0.037)	(0.083)	(0.096)	(0.057)
Age	-0.020**	-0.002	0.010*	-0.004	0.037***	0.014*
	(0.009)	(0.004)	(0.005)	(0.011)	(0.014)	(0.008)
Observations	303	303	303	77	303	303

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

Table 9: Dishonesty, Relying on Own Judgement, and Creativity

Measure	Hypothesis	Result
Political Bias	+	+
Trust	-	0
Trustworthiness	-	0
Trust Game Belief	-	0
Entitlement	+	0
Dishonesty	+	0
Relying on Own Judgement	+	0
Creativity	+	-

Table 10: Hypothesized Sign Versus Result for Each Measure

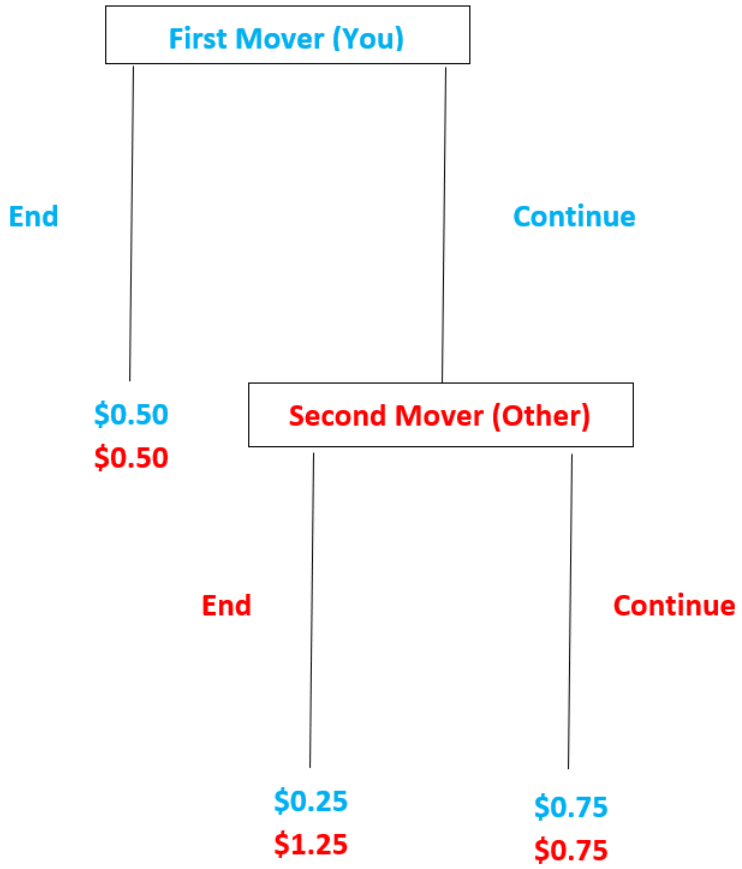

Figure 3: Trust Game Tree

	Option 1:		Or	Option 2:	
	Chance of $0.25	Chance of $0.75		Chance of $0.25	Chance of $0.75
1	**0 in 100**	**100 in 100**	*Or*		
2	**10 in 100**	**90 in 100**	*Or*		
3	**20 in 100**	**80 in 100**	*Or*		
4	**30 in 100**	**70 in 100**	*Or*		
5	**40 in 100**	**60 in 100**	*Or*	% second movers	% second movers
6	**50 in 100**	**50 in 100**	*Or*	chose to end	chose to continue
7	**60 in 100**	**40 in 100**	*Or*		
8	**70 in 100**	**30 in 100**	*Or*		
9	**80 in 100**	**20 in 100**	*Or*		
10	**90 in 100**	**10 in 100**	*Or*		
11	**100 in 100**	**0 in 100**	*Or*		

Figure 4: Belief Elicitation Table

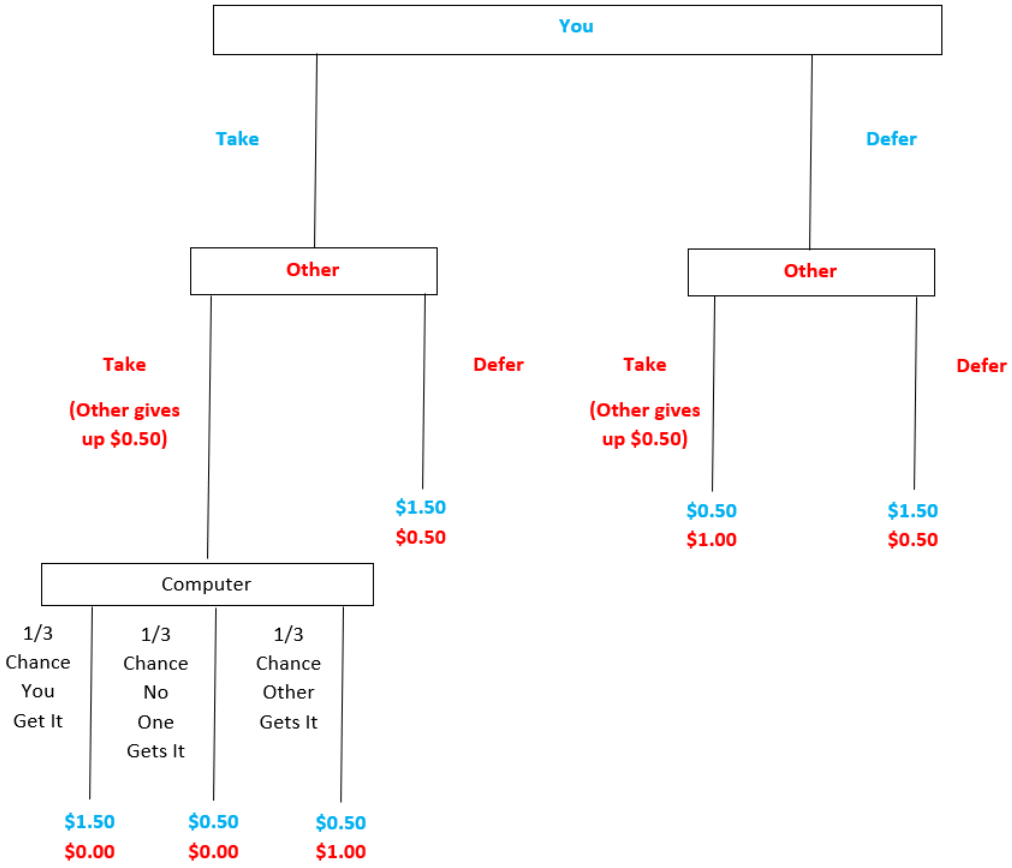

Figure 5: Entitlement Game Tree

CHAPTER 3

NORM FORMATION CASCADES: THEORY AND EXPERIMENTAL EVIDENCE

3.1 Introduction

A social norm is a standard of behavior for people in a given social group. A common conception of social norms is that they represent values shared by the majority of people in that social group. Extending this conception to individuals' preferences, it is natural to expect that a given social norm represents the preferences of the majority of the group (e.g. a norm for fairness might indicate that most people want to act fairly). In which case, one function of a social norm could be to direct others in the group, via social pressure or a related mechanism, to behave in accordance with the majority preference. While it is possible that social norms reflect majority preferences in most cases, there have been cases both in experimental settings and outside the lab which suggest that some social norms could represent minority preferences or values. The purpose of the present study is to investigate how social norms representing minority preferences can form, specifically focusing on small group settings.

The conception of social norms as representing majority values or preferences can be found throughout the economics literature on social norms (e.g. [Fehr and Fischbacher, 2004], [Barr and Serra, 2010], [d'Adda et al., 2020]). Although this conception might reflect the underpinnings of most social norms, there seem to be cases where the social norm does not represent the majority. One example is the 50-50 norm in the dictator game, as studied in [Andreoni and Bernheim, 2009]. In their experiment, subjects participated in a standard dictator game where the dictator must choose how to split $20 with a recipient, except that there is some probability (either 0, 0.25, 0.5, or 0.75) that "nature" makes a forced choice

of either \$0 or \$1 given to the recipient depending on the treatment. They find that when the probability of a forced choice is 0, most subjects offer an equal split of \$10 in accordance with the 50-50 norm. However, as the probability of nature forcing a choice increases, fewer subjects choose an even split, as a plurality of subjects voluntarily choose the "nature" choice of either \$0 or \$1 (a strict majority in the \$0 treatment). The results of their experiment suggest that most subjects do not have a preference for acting fairly, but rather, at most a preference to appear as being a fair person to others. In other words, the 50-50 norm appears to represent a preference held by a minority of subjects.

While the 50-50 norm example might provide evidence of a social norm that does not represent majority preferences, it does not give us any indication about how a social norm representing minority preferences can form. Even if the 50-50 norm does not currently represent a majority preference, it is possible that the norm represented majority preferences at the time that it formed. We want to demonstrate that an action preferred by the minority of a group at that time could nonetheless take over and become the social norm for the group. As an example of how such a process can take place, consider the following choice (from the COVID-19 pandemic) of whether to face mask in a small group setting. Consider a scenario where six employees, one-at-a-time, walk into a seminar room for a business meeting. Each person can independently choose whether to wear a face mask. Suppose that the first person in the room chooses to not wear a face mask. This action sets an initial precedent against mask wearing for the group. The second person arriving to the meeting may individually prefer to wear a mask, but she may also wish to avoid an awkward situation where she is the only group member wearing a mask, so she also chooses to not wear a face mask. At this point, if the remaining group members also wish to avoid an awkward situation where each might be the only group member wearing a mask, then the remaining group might also choose to not wear a face mask. Such an outcome could occur in a case where each of the last five members of the group individually prefer to wear a face mask, all thanks to the first person choosing not to wear a mask. We will refer to the phenomenon described in this

example as a "norm formation cascade," since after a certain point, each person ignores his individual preference and simply chooses the action most common among the group.

Our concept of "norm formation cascades" is closely related, in both name and setting, to the concept of information cascades (see [Anderson and Holt, 1997]). In both settings, cascades can form when people rationally ignore either their private information (information cascade) or their private consumption preference (norm formation cascade) to simply follow what a majority of people have already done. The reason why people care about the choices of others, however, is different in the two settings. In the information cascade setting, your payoff only depends on choosing the correct action for the unknown underlying state of the world, of which the only information you initially have is a private signal. The choices of others thus provides you with further information about what the true underlying state might be (i.e. others' choices only have instrumental value). In the norm formation cascade setting, your payoff not only depends on your own choice, but also on the choices of others. Since others' choices factor directly into your payoffs in the norm formation cascade setting, the solution concept of the two concepts are different as well. In the information cascade setting, you only care about others' actions insofar as they provide information about the underlying state, so you can ignore the potential choices of those who follow you in the sequence. Therefore, in this setting you are just solving a Bayesian updating problem. In the norm formation cascade setting, you have to take into account the actions of all people in your sequence, both those who precede you and those who follow you. As such, the solution concept for the norm formation cascade setting is subgame perfect equilibrium via backward induction.

The mechanism driving the potential occurrence of these norm formation cascades is closely related to the concept of "rational irrationality" (see [Caplan, 2001]). The concept of rational irrationality has been put forth as an explanation for why people may hold beliefs that are distant from reality (e.g. insisting that the Earth is about 6,000 years old instead of billions of years old). As long as the personal cost for maintaining an incorrect belief is

not too high, an individual may want to keep hold of that incorrect belief. For instance, having an incorrect belief about the unemployment rate during a given president's administration probably has little consequence for most people in their daily lives. As such, if you personally dislike a given president and enjoy expressing that dislike with friends, you will probably hold onto that incorrect belief because you have no incentive not to (hence the term "rational irrationality"). Our setting of norm formation cascades works in a similar fashion. As long as the personal cost to ignoring your private consumption preference is not too high, you may be willing to ignore it for the sake of showing a stronger consensus about what action people in your small group should take. Going back to the face mask example, you may have a preference towards wearing a face mask, but unless you are worried that being infected will lead to severe consequences, you may be willing to ignore your preference if no one else in your small group is wearing a face mask.

Regarding the potential divergence between a given social norm and majority preferences, the present paper is not the first to suggest that there can be a disconnect. For instance, in an earlier study into the topic of social norms, [Elster, 1989] points out the existence of social norms that do not appear to either benefit the individual or the social group (e.g. norms against achieving and displaying success beyond the success of others in the group). More recently, [Bursztyn et al., 2020] conducted an experimental study with married men in Saudi Arabia where they show that most subjects privately support women working outside the home, which suggests that the perceived social norm in Saudi Arabia does not represent the majority preference. While these studies provide a precedent for the idea that some social norms might not represent majority preferences, they do not address our central question of how social norms in favor of a minority preference can form. In the case of social norms for women's labor opportunities in Saudi Arabia, one possibility is that at the time of its formation, the social norm did in fact represent the majority preference and only through time has that preference become the minority (with the social norm withstanding possibly due to "pluralistic ignorance"). There could be other explanations, but the point is that our

present paper looks to investigate how the formation of social norms representing minority preferences is possible, an issue that prior studies have remained agnostic about.

To address our fundamental question about how the formation of social norms representing minority preferences might be possible, we present a simple theoretical model showing under which conditions a social norm representing the minority preference will (or will not) form and conduct a laboratory experiment testing our theoretical predictions. Our theoretical model predicts that, depending on the prior probability of the minority preference, a norm formation cascade will occur when the first person in the sequence has the less likely private preference (if the prior probability of the minority preference is too low, then a norm formation cascade will not occur). The results of our laboratory experiment aligned with our theoretical predictions in some regards, but deviated from the theoretical predictions in others. While norm formation cascades did occur in the scenario predicted by our theoretical model, they did not occur as often as predicted. That said, there were also scenarios in which norms in favor of the minority preference formed where the theoretical model predicted that the norm should represent the majority preference. Ultimately, these deviations from the theoretical predictions seem to be explained by two main things. First, some subjects in the second position of the sequence were more insistent about choosing according to their private consumption preference than the theoretical model predicted. Second, to a lesser extent, some subjects in the first position did not choose according to their private preference when they had the less likely preference. The findings of our experimental study can help to inform further investigation into how and when the formation of social norms in favor of minority preferences is possible. The remainder of the paper is structured as follows. Section 2 presents our theoretical model for norm formation cascades. Section 3 presents our experiment design as well as our hypotheses for our experiment, which are based on the predictions from our theoretical model and other theoretical considerations. Section 4 presents the results of our laboratory experiment and Section 5 concludes.

3.2 Theoretical Model

The purpose of our study, and thus the purpose of our theoretical model, is to show how and under which circumstances norm formation cascades are possible. To do this, we will build up the theoretical model in the following manner. First, we will present a simple utility function that will specify the payoffs that are present in our setting. From these payoffs, we can identify a particular class of agents that make norm formation cascades more likely to occur and we will focus our analysis on this class of agents. Once we have specified the payoffs and agent type, we will go through the backward induction analysis and provide the subgame perfect equilibrium for two cases: one where our model predicts that norm formation cascades can occur and one where our model predicts that norm formation cascades will not occur. Our theoretical model not only provides us predictions about when norm formation cascades will occur, but also the type of behavior required in each position to have norm formation cascades. The model's predictions regarding these two main issues of interest will form the basis of our hypotheses for our laboratory experiment.

In this setting, agent i must choose which value of $x = \{0, 1\}$ to take (e.g. $x = 0$ is not wearing a mask, $x = 1$ is wearing a mask). Agents are in groups of N, and agents make their choices sequentially until all N agents have made a choice. At the beginning of this process, there are no social norms for or against either of the two options. However, a social norm can form if (say) at least $\frac{3}{4}$ of agents pick a given value of x (e.g. $\frac{3}{4}$ of agents pick $x = 0$). Agent i has both a preference for which value of x she would want to choose, and a preference for complying with the social norm if one exists. We can write i's utility function as follows:

$$u_i(x_i) = h(x_i) + \alpha_i g(x_i, x_{-i})$$

x_i is agent i's choice, x_{-i} are the choices of all agents other than i, and $\alpha_i \geq 0$ represents the degree to which agent i cares about complying with social norms.

$h(x_i)$ represents the utility agent i gets from consuming her preferred value of x. She gets $y > 0$ if she picks her preferred value of x, and 0 otherwise. $\alpha_i g(x_i, x_{-i})$ represents the utility (disutility) agent i gets from complying with (defying) the social norm if one exists. She gets $\alpha_i z$ ($z > 0$) if she complies with the social norm, $-\alpha_i z$ if she defies the social norm, and 0 if there is no social norm. As such, agent i could receive one of six possible payoffs:

	Comply	Defy	No Norm
Consume	y+α_iz	y-α_iz	y
Don't Consume	α_iz	-α_iz	0

With these payoffs, consider the decision problem for the last agent in the sequence. What this agent chooses comes down to his value of α_i. If $\alpha_i < \frac{y}{2z}$, then he always chooses his privately preferred value regardless of the social norm. We can refer to these agents as **norm defiers**. These agents will only follow a social norm if it happens to align with their private preferences. If $\frac{y}{2z} \leq \alpha_i \leq \frac{y}{z}$, then he complies with the social norm if one exists, otherwise he chooses his privately preferred value. We can refer to these agents as **norm compliers**. If $\alpha_i > \frac{y}{z}$, then he will choose his privately preferred value *unless* he can form a norm in favor of the other value (in which case he will form the norm). We can refer to these agents as **norm creators**.

Since the purpose of this study is to show how norm formation cascades can occur, we will focus on the case where all agents are "norm creators." To be clear, we are not assuming that all people are norm creators, but rather we are focusing on a case where norm formation cascades can occur depending on the probability of each preference. One of the conditions that would allow norm formation cascades to occur is that all agents are norm creators as a social norm will always form in this case (though, of course, not always in favor of the minority preference). Of course, including norm compliers and (especially) norm defiers would necessarily decrease the likelihood of norm formation cascades occurring, but the effect of their presence on the occurrence of norm formation cascades is an avenue for future research once we have investigated how likely norm formation cascades are to occur in this baseline case.

Using $\frac{3}{4}$ as the threshold for a social norm forming, the cases with 2 or 3 agents are trivial. If Agents 1 and 2 pick the same value, Agent 3 will pick that value to form a social norm and receive the norm compliance payoff (given our assumption about α, we know Agent 3 prefers complying with a norm to consuming his preferred value in the absence of a social norm, as do Agents 1 and 2). Agent 2, knowing this, will simply pick what Agent 1 chooses. Therefore Agent 1, knowing this, will pick the value she prefers to consume, and Agents 2 and 3 will follow along since they want the payoff from complying with a social norm. This is the only equilibrium in this setting when there are 2 or 3 agents. This equilibrium allows for the possibility that Agent 1 has a preference that has a very low prior probability (e.g. a prior probability of $p = 0.01$), but in terms of realizations, Agent 1 is in no lower than a simple minority (at the lowest a 2-to-1 minority). We want to show that a social norm cascade can form when preference for the formed social norm is in a significant minority both in terms of prior probability and realizations. To show this, we need to consider cases with at least 4 agents.

We are in the same setting as above, except there are now four agents. Starting with Agent 4, we know that he will always choose the value which has been chosen more often by the first three agents (regardless of which value gives him the better consumption payoff). To see this, first consider the case where the first three agents all picked the same value. In this case, a social norm has already formed, so Agent 4 will of course comply with the norm to receive the compliance payoff. If the first three agents' choices are split 2-to-1, Agent 4 will still always pick the value chosen 2 times. This is because in this case, Agent 4 has the power to create the social norm, and since he places more importance on the compliance payoff relative to the consumption payoff, he will make sure there is a social norm he can comply with. Given that Agent 3 knows that this is how Agent 4 will play, her choice is also straightforward. If Agents 1 and 2 pick the same value, Agent 3 will pick that same value (regardless of which value gives her the better consumption payoff). If Agents 1 and 2 pick different values, then Agent 3 will simply pick the value that gives her the better

consumption payoff. Agent 3 knows that she can do this because she knows Agent 4 will form a social norm in favor of 3's preferred value, as Agent 4 will always choose the value that was chosen more often by the first three agents. The fact that Agent 3, unlike Agent 4, will sometimes choose a value based on her consumption preference, makes Agent 2's choice a little more complex. Now, if Agent 1 chooses the value that matches with Agent 2's consumption preference, then Agent 2 will of course choose that value, since Agent 2 knows that Agent 3 will then form a social norm in favor of the value (giving Agent 2 his best possible payoff). If Agent 1 chooses the other value, then Agent 2's choice will come down to his subjective beliefs about Agent 3's preferences. Agent 2 knows that in the case of split between Agent 1 and 2, Agent 3 will choose her preferred value. So, if Agent 2 thinks it is sufficiently likely that Agent 3 has the same preference as himself, he chooses his preferred value. Otherwise, Agent 2 will choose the same as Agent 1, which guarantees Agent 2 his norm compliance payoff (as the best response for Agent 3 is to then form the norm and Agent 4 to comply with the norm). Agent 2's threshold probability calculation is therefore the following:

$$p(y + \alpha z) + (1 - p)(y - \alpha z) = \alpha z$$

where p is the probability Agent 2 and Agent 3 have the same consumption preference (i.e. p is Agent 2's subjective belief about the prior probability of his consumption preference). After some algebra, we find that the cutoff probability is

$$p = \frac{2\alpha z - y}{2\alpha z}$$

The fact that Agent 2 might make a decision based on his subjective beliefs adds an additional complexity to Agent 1's decision. Namely, Agent 1 will have to form subjective beliefs about Agent 2's subjective beliefs. Agent 1 can potentially have all the information necessary to solve Agent 2's threshold calculation from above, but just because she knows Agent 2's threshold does not tell her what Agent 2's subjective belief is (e.g. if Agent 2 thinks the actual probability of his preference is 0.99, then he will probably choose according to his

consumption preference). For simplicity, assume that Agent 1 holds two beliefs, $r_1 \in (0,1)$ and $r_2 \in (0,1)$. If Agent 1 prefers a different action from Agent 2, r_1 is Agent 1's belief that Agent 2 would reciprocate her choice if she chooses her preferred action. If Agent 1 prefers the same action as Agent 2, r_2 is Agent 1's belief that Agent 2 would reciprocate her choice if she chooses the other action. In this case, Agent 1's threshold probability calculation is the following:

$$\begin{aligned} &p(y+\alpha z) + (1-p)(r_1(y+\alpha z) + (1-r_1)(p(y+\alpha z) + (1-p)(y-\alpha z)) \\ &= (1-p)(\alpha z) + p(r_2(\alpha z) + (1-r_2)((1-p)(\alpha z) + p(-\alpha z)) \end{aligned}$$

The left-hand side is Agent 1's expected payoff if she chooses the value aligned with her consumption preference. The first term shows that if Agent 2 has the same consumption preference, Agent 2 will pick the same value and then Agents 3 and 4 (as we have shown) will also choose that value, giving Agent 1 a payoff of $y+\alpha z$. The second left-hand side term shows what happens if Agent 2 has a different consumption preference from Agent 1. Agent 2 might choose the same as Agent 1 anyways (e.g. if he thinks the probability of his preference isn't high enough), or he might choose to go with the value aligned with his consumption preference (e.g. if he thinks the probability of his preference is high enough). If Agent 2 chooses the opposite of Agent 1, then the direction of the norm formation cascade will hinge on Agent 3's preference. The right-hand side goes through the exact same cases, just for when Agent 1 chooses the opposite of her consumption preference. After some algebra, we find that the threshold probability is

$$p = \frac{-4\alpha z + 4\alpha z r_1 \pm \sqrt{16\alpha^2 z^2 - 16\alpha^2 z^2 r_1 - 16\alpha^2 z^2 r_2 + 16\alpha^2 z^2 r_1 r_2 - 8y\alpha z r_1 + 8y\alpha z r_2}}{2(2\alpha z r_1 - 2\alpha z r_2)}$$

Recalling that the purpose of the present study is to show how norm cascades in favor of minority preferences can form, it will behoove us to give values to our parameters to give us settings with clean theoretical predictions for which we can test. For each of our settings, let $y = 2$, $z = 2$, and $\alpha = 2$. Furthermore, for our first setting, let the probability an agent

prefers $x = 0$, $Pr(x = 0) = 0.3$, the probability an agent prefers $x = 1$, $Pr(x = 1) = 0.7$, and assume that these probabilities are common knowledge amongst all agents. Recall that the strategies of Agents 3 and 4 did not depend on the exact values of our parameters, so we can move to the strategy of Agent 2 in this setting. Plugging the values of these parameters into Agent 2's threshold calculation solved for above, we get that Agent 2's threshold probability in this setting is $p = 0.75$. Notice that the probability of neither preference meets this minimum threshold. Therefore, in this setting, Agent 2 will simply choose whichever action Agent 1 chooses, regardless of which action Agent 2 prefers.

Moving to the strategy of Agent 1 in this setting, since Agent 1 knows that Agent 2 will always reciprocate her decision since neither preference meets the threshold, we know that Agent 1's beliefs are $r_1 = 1$ and $r_2 = 1$. If we plug all of our parameter values into Agent 1's threshold calculation, we see that the value is undefined. This is because there is not a "threshold" probability in this setting; Agent 1 should always choose according to her preference because Agent 2 will always reciprocate her choice. As we can see, a norm formation cascade can occur in this setting whenever Agent 1 prefers $x = 0$ (and no one else does), as we have shown that the remaining agents in the sequence will also choose $x = 0$ (thus forming a norm for $x = 0$).

For our second setting, let the probability an agent prefers $x = 0$, $Pr(x = 0) = 0.1$, the probability an agent prefers $x = 1$, $Pr(x = 1) = 0.9$, and assume that these probabilities are common knowledge amongst all agents. Keeping the values of our other parameters the same, Agent 2's threshold probability in this setting is still $p = 0.75$. So, in the case where Agent 2 prefers $x = 1$, our theoretical model predicts that he will choose $x = 1$ regardless of what Agent 1 chooses.

Since when Agent 2 prefers $x = 1$ he will choose $x = 1$ regardless of Agent 1's choice, Agent 1 might not choose her preference if she prefers $x = 0$. If Agent 1 prefers $x = 1$, then she will of course choose her preference since Agent 2 would always reciprocate her choice in that case. When Agent 1 prefers $x = 0$, we know from our above analysis that

$r_1 = 0$ and $r_2 = 1$ since Agent 2 will always choose $x = 1$. With the values of r_1 and r_2 known, the threshold probability for Agent 1 (when she prefers $x = 0$) is $p = 0.5$. Since $Pr(x = 0) = 0.1$, our theoretical model predicts that Agent 1 will always choose $x = 1$ regardless of her preference. Therefore, in this second setting, our model predicts that norm formation cascades for the minority preference are not possible.

Ultimately, regardless of the specific parameter values, if all agents are "norm creators," then there are only two general classes of equilibria. The first general class is where Agents 1 and 2 choose the same value, in which case Agents 3 and 4 will also choose that value. The second general class is where Agents 1 and 2 choose different values, in which case Agent 3 chooses a value based on her consumption preference and Agent 4 simply chooses the value that Agent 3 chose. With the specific parameters used above, we can see how this process can, though not always, generate a cascade in favor of a minority consumption preference.

We have shown how norm formation cascades can occur in favor of a minority preference with four agents, but as it turns out, the process is almost identical for any number of agents from four to seven. Consider a case with six agents, which is what we will ultimately utilize in our experiment design. Agent 6 will pick with the majority if the choice arrives to him at "5-0" or "4-1" (either a norm is already established, or he can be the one to form the norm). He will only choose based on his consumption preference if the choice arrives to him at "3-2". Agent 5, knowing this, will pick with the majority if the choice arrives to her at "4-0" or "3-1" (she will only choose based on her consumption preference if it arrives to her at "2-2"). Agent 4, knowing this, will always choose the value picked more often by the first three agents (does not matter if it's "3-0" or "2-1"). Agent 3, knowing this, will pick whatever Agents 1 and 2 picked if the first two agreed. If Agents 1 and 2 picked different values, then Agent 3 will pick the value that matches with her consumption preference. Notice that Agent 3's strategy in this case with six agents is exactly the same as her strategy in the case with four agents. Therefore, Agent 2's threshold calculation is exactly the same as the four agent case, and by extension Agent 1's threshold calculation is also exactly the same as the

four agent case. We thus have the same result regarding equilibria outcomes as with the four agent case. In the first general class of equilibria, Agents 1 and 2 choose the same value, and then Agents 3 through 6 also choose that value. In the second general class, Agents 1 and 2 choose different values, Agent 3 chooses based on her consumption preference, and Agents 4 through 6 choose whichever value Agent 3 chose. Of course, with the parameter values we specified in our two settings, and by assuming that the probabilities of each preference are common knowledge, our theoretical model provides a unique equilibrium prediction for each setting.

3.3 Experiment Design

The norm formation cascade experiment consists of three active phases (Phases 1-3). In the Phases 1 and 2 of the experiment, subjects made choices in the six-agent groups we discussed in the theoretical model. We kept subjects in an abstract choice setting to make sure subjects' decision making was not influenced by any preexisting social norms for behavior. The purpose of our study is to investigate how social norms in favor of minority preferences can form, so we needed to make sure subjects did not have any preconceptions about how they should behave. In each group, the first three positions in the sequence were occupied by the subjects, and the last three positions were occupied by robots. Each subject and robot in each group had to make a choice between two actions: Action A and Action B. The robots were programmed to choose whichever action was chosen by the majority of subjects in their group. Recall from the theoretical model that we are focusing on cases where all agents are "norm creators". We therefore programmed our robots to behave in this manner because, as we showed in the theoretical model, a rational "norm creator" in positions four through six will always choose whichever action was chosen most often. One reason[20] we used robots in these positions and not subjects is because, for the purposes of this study, we are not

[20]Other reasons are that it is simpler to create illustrative examples to introduce subjects to the strategic setting when the choices of the last three positions are automated and that it is easier to conduct experiment sessions for multiples of 3 subjects instead of multiples of 6.

interested in testing whether subjects make mistakes in these last three positions. The focus of our study is to investigate whether norm formation cascades can occur and what types of behavior make these cascades more-or-less likely to occur. If people make mistakes these last three positions, we know that norm formation cascades are less likely to occur, as such mistakes lower the probability of a social norm forming. We want to better understand how people make choices in the first three positions, as the choices of those in the first three positions dictate which action could potentially become the social norm for the group. By automating the choices of the last three positions, we make it easier to pin down the decisions of the subjects in the first three positions. For instance, if a subject's behavior in Position 2 deviates from our theoretical predictions, we now know that the deviation is not because the subject was worried about someone in the last three positions making a mistake since these choices are now automated. The decision making of people in the last three positions can be useful to investigate in the future, but first we want to better understand whether the people in the first three positions process the strategic environment in the manner that our theoretical model predicts.

Before subjects began Phase 1 of the experiment, we provided subjects a few illustrative examples to help them understand how the payoffs work. In these examples, as well as in Phase 1, the payoffs were as follows. Choosing one's privately preferred action provides a payoff of \$0.30, while choosing the other action provides a payoff of \$0.10. On top of that, choosing the same action as the majority of agents in one's group provides an additional payoff of \$0.80. So, subjects can earn either \$0.10, \$0.30, \$0.90, or \$1.10 depending on their choice and the choices of others in their group. Going back to the theoretical model, these payoffs correspond to $y = 0.2$, $z = 0.2$, and $\alpha = 2$. Because we are focusing on settings where all agents are "norm creators", a norm, in one direction or the other, will always form. Therefore, the payoffs where no norm forms are never realized in our experiment. Compared with the payoffs presented in the theoretical model, we increased payoff levels by \$0.50 (e.g. the payoff of $-\alpha z$ is now \$0.10 instead of $-\$0.40$). While increasing payoffs levels does not

change the strategic setting, we wanted all payoff outcomes to be positive so that subjects did not make choices simply to avoid possible negative or zero payoffs[21].

Phase 1 (Small Payoff)

In Phase 1 of the experiment, subjects play through 10 rounds of the interaction to help familiarize themselves with making choices in each position. The purpose of this phase of the experiment is to help subjects become familiar with the strategic setting and to give them an opportunity to learn how to make choices in various circumstances. We wanted subjects to be thoroughly acclimated to our setting so that their behavior in the next phase of the experiment reflected their genuine decision making rather than a lack of familiarity with the choice setting.

At the beginning of each round, subjects were randomly assigned to a group of three humans and three robots. They were each also randomly assigned to one of three positions in their group's sequence: Position 1, Position 2, or Position 3. For every round, we did a random draw for every subject to determine which action will pay them better (i.e. their private consumption preference)[22].

When it was each subject's turn to choose an action, we showed them the choices of any previous positions, as well as who in their group had yet to make a choice. Figure 6 provides a screenshot of this choice screen from the perspective of a subject in Position 2. A group's round was completed once every subject in the group made their choice of action, after which all subjects in the group were presented with a results screen summarizing the outcome and subject's payoff[23] for that round. After all groups completed a round, subjects

[21]Another benefit of including this increase in payoff levels was to give subjects a guaranteed baseline payment for participating in our experiment, on top of the $10 show up fee

[22]The probability of preferring each action depends on the treatment. In the "Baseline" treatment, the probability of preferring Action B is 30%, while in the "Low Probability" treatment, the probability of preferring Action B is 10%.

[23]As previously mentioned, the purpose of these 10 rounds was to help subjects familiarize themselves with making choices in each position (i.e. to learn how to behave in this setting). [Merlo and Schotter, 1999] provides an experimental study showing that experimenters should think carefully about how to utilize payoffs in "learning" rounds of an experiment. As they point out, some experiments may be better suited for a "Learn-Before-You-Earn" payoff structure, while for others a "Learn-While-You-Earn" payoff structure

were randomly assigned to a new group of three humans and three robots. We also randomly reassigned subjects to potentially new positions in their new group sequence, as well as conducting a new random draw for each subject's private consumption preference.

Phase 2 (Strategy Method)

In Phase 2 of the experiment, subjects made choices for each position, in every scenario which they could potentially face (i.e. subjects made choices via the strategy method). The purpose of this phase of the experiment was to obtain each subject's fully specified strategy such that we could study their decision making and how it did (not) align with our theoretical predictions.

The choices subjects made in this phase of the experiment were the same as they made in Phase 1, except for this phase payoffs were multiplied by 10. As such, subjects could earn either \$1.00, \$3.00, \$9.00, or \$11.00 depending on which action they preferred, their choice of action, and the choices of the other subjects in their group. We showed subjects the choice screens they would be using before the start of Phase 2 to make sure they understood how they would be making their choices before they had to do so for payment. Figure 7 provides a screenshot of the subjects' choice screen for Position 2. The choices made in this phase in the experiment will allow us to fully characterize each subject's strategy in our small group norm formation setting and to determine how often norm formation cascades occur in each possible sequence of subjects and preference arrangements.

After all subjects had made their choices for each position, we randomly assigned them into groups, randomly assigned them a position in their group, and conducted a random draw to determine each of their private consumption preferences. By pairing each subject's random draw of position and private consumption preference with the choices they made via the strategy method, we were able to determine each subject's payoff for this round of the experiment. Just as in the 10 rounds of Phase 1, we presented subjects with a results screen

would be appropriate. Ultimately, given the clear feedback subjects receive after the conclusion of each round, and the simplicity of the task itself, we determined that a "Learn-While-You-Earn" payoff structure would be appropriate for our experiment.

summarizing the outcome and subject's payoff.

Phase 3 (Behavioral Measure)

The last phase of the experiment had subjects make choices that were different in nature from the choices in Phases 1 and 2. In Phase 3, subjects were now choosing between two options: Option 1 and Option 2. Just as in the first two phases of the experiment, subjects would be in groups with two other humans. There were nine rows in which subjects had to choose between Option 1 and Option 2. Figure 8 provides a screenshot of the choice screen for Phase 3 of the experiment. The maximum payoff a subject could earn in this phase of the experiment was \$4.00. Option 1 would always pay a subject the best possible payoff of \$4.00, but the other subjects in their group would generally earn less. Namely, starting from \$4.00 in Row 1, the payoff for the other two subjects would decrease by \$0.50 for each row, ending with \$0.00 in Row 9. Option 2 would always pay the subject \$2.00 in each row, but would pay the other two subjects \$4.00 in each row. Subjects had to choose which option they preferred for each row. We programmed the choice screen so that once a subject selected Option 2 for a given row, Option 2 would be automatically selected for all rows below. Similarly, we also programmed the choice screen so that if a subject selected Option 1 for a given row, Option 1 would be automatically selected for all rows above[24]. Once all subjects submitted their choices, we randomly assigned subjects into groups of three. We then randomly picked the choices from one subject in each group to determine the payoffs for all three subjects in that group. For each group, we also randomly drew one of the nine rows. If the selected subject from a group chose Option 1 for that group's row, then the three subjects were paid according to Option 1. If the selected subject from a group chose Option 2 for that group's row, then the three subjects were paid according to Option 2.

The purpose of Phase 3 was to provide a behavioral measure that might be able to explain a subject's choices in Phase 2. For instance, in the baseline treatment of our experiment,

[24]We programmed the choice screen in this manner to guarantee that all subjects submitted choices that reflected consistent preferences. Subjects were told that choices would be automatically selected in this manner in the experiment instructions (see Appendix).

the theoretical model predicts that a subject who prefers Action B in Position 1 will choose Action B. If a given subject does not choose Action B in this case, there are several possible explanations for this deviation from the theoretical prediction. One possibility is that the subject simply looks at the probabilities of each preference and does not choose Action B because it is the less likely preference. Another possibility is that while the subject is sufficiently sophisticated to understand the strategic setting, she might think that the others in her group are not strategically sophisticated so she chooses Action A since it's more likely that the subject in Position 2 prefers Action A than Action B and she thinks the subject in Position 2 will just choose according to their preference. A third possibility is that the subject does not want to force her preference on the others in her group. If some subjects are not choosing Action B in Position 1 of the baseline treatment because of this third possibility, we would observe a difference in our behavioral measure between those who do and do not choose Action B in Position 1 of the baseline treatment.

Having now described the design of our experiment, we will go through the main hypotheses for the outcomes of the experiment. Our hypotheses are based upon the predictions from our theoretical model. To understand how norms in favor of minority preferences can form, we would first want to see whether these norm formation cascades occur as our theoretical model predicts. Our first hypothesis therefore regards the frequency of norm formation cascades in our data. For any given sequence of subjects, there are eight possible arrangements of preferences (listed in order of the preference for the subject in Position 1, the preference of the subject in Position 2, and the preference of the subject in Position 3): BAA, ABB, BAB, BBA, BBB, AAB, ABA, and AAA. The preference arrangement of greatest interest to us is BAA. Since only one subject prefers B in this sequence, this preference arrangement would definitely constitute a norm formation cascade if Action B becomes the norm for the group. Furthermore, our theoretical model predicts that in the baseline treatment, Action B will become the norm, while in the small probability treatment Action A will become the norm. Three other preference arrangements that are of relatively high interest are ABB,

BAB, and BBA. In cases where Action B becomes the norm, these preference arrangements *can* constitute norm formation cascades if the other three agents in the group, which are played by robots in our experiment, all privately prefer Action A. In the case of the preference arrangements BAB and BBA, our theoretical model predicts that norm formation cascades in favor of Action B will occur in the baseline treatment, but not in the small probability treatment. In the case of the preference arrangement ABB, our theoretical model predicts that norm formation cascades in favor of Action B *will not* occur in either treatment. That said, it is possible that subjects in Position 2 are more insistent about choosing according to their private consumption preference than laid out in our theoretical model. In which case, it is possible that we do see norm formation cascades in favor of Action B for the preference arrangement ABB. As these are the four preference arrangements where norm formation cascades in favor of the minority preference (Action B) are theoretically possible without subjects making outright mistakes[25], we will focus on the theoretical predictions for these in our first hypothesis.

Hypothesis 1 *(Norm Formation Cascades) In the baseline treatment, norm formation cascades in favor of Action B will (always) occur for the preference arrangements BAA, BAB, and BBA, but will never occur for the preference arrangement ABB. In the small probability treatment, norm formation cascades in favor of Action B will never occur for any preference arrangement.*

If we find that norm formation cascades do not occur as often in cases where we predict that they should happen, or occur more often in cases where we predict that they should not happen, we can look at the subjects' strategies to test where their behavior deviates from our theoretical predictions. We will focus our hypotheses regarding subjects' strategies to their strategies in Position 1 and Position 2, as these are the two positions for which we can provide a theoretical justification for divergence from the predicted strategy other

[25]The preference arrangement BBB cannot constitute a norm formation cascade in favor of Action B because the preference for Action B will never be in the minority of a six person group. The preference arrangements AAB, ABA, and AAA can only lead to a norm favoring Action B if someone makes a mistake.

than the subject making a mistake. For Position 1, a subject might dislike trying to force a minority preference onto the group as its norm, perhaps because in expectation she is lowering the payoff to the other two group members, so she always chooses Action A. For Position 2, a subject might insist on trying to force their private preference as the norm, because even though doing so might lower their expected payoff, they may feel optimistic that the norm will end up in their favor resulting in their best possible payoff outcome. Given these theoretical possibilities, we have reason to test the hypotheses that subjects will behave as predicted by our model in Position 1 and Position 2.

Hypothesis 2 *(Position 1 Strategy) When the subject in Position 1 prefers Action A, they will choose Action A. When the subject in Position 1 prefers Action B, they will choose Action B in the baseline treatment and choose Action A in the small probability treatment.*

Hypothesis 3 *(Position 2 Strategy) In the baseline treatment, the subject in Position 2 will choose whatever the subject in Position 1 chose. In the small probability treatment, if the subject in Position 2 prefers Action B, they will choose whatever the subject in Position 1 chose. If the the subject in Position 2 prefers Action A, then they will choose Action A.*

If we find that a significant portion of subjects' strategies deviate from the theoretically predicted strategies in our model, our behavioral measure (Phase 3) could potentially provide an explanation for some of the deviation. We can categorize each subject's strategy into different types based on the manner in which the strategy deviates from our predictions. The first type of strategies are those aligned with our predictions ("Theory" strategies). The second type are strategies that, in at least one case, insist on choosing the preferred action when our theoretical model predicts otherwise, but is aligned with our predictions otherwise ("Optimist" strategies). An example of this type of strategy is one, from a subject in the baseline treatment, where Action A is chosen in Position 2 after Position 1 chose Action B, but is aligned with our predictions for every other case. The third type are strategies where the subject always chooses his preferred action, including cases in Position 3 where doing so

lowers the subject's payoff ("Preference" strategies). The fourth type are strategies where the subject will only choose Action B if his choice can guarantee a norm in favor of Action B, otherwise he will choose the action with the higher prior probability, Action A ("Probability" strategies). The fifth type are strategies where the subject always defers to Action A when he is in Position 1, and always defers to the choice of the subject in Position 1 when he is in Position 2 ("Deferential" strategies). As discussed above, one potential reason a subject might take a "deferential" strategy is that she dislikes trying to force the minority preference onto the group as its norm, as it might decrease the group's combined payoff, so she prefers choosing Action A. Another possible reason a subject might take a "deferential" strategy, though, is because she doubts the strategic sophistication of the others in her group, so she chooses Action A to increase the likelihood that she will receive the norm compliance payoff. To see if the former can explain why some subjects choose a "deferential" strategy, we can test whether subjects choosing a "deferential" strategy had a lower score on our behavioral measure than the subjects choosing a different type of strategy.

Hypothesis 4 *(Behavioral Measure) Subjects with a "deferential" strategy type will have a lower behavioral measure score than the other subjects.*

3.4 Results

The first question we want to address is whether norm formation cascades occur as our theoretical model predicts. To address this question, we can generate each possible permutation within our two treatments[26], and from the subjects' provided strategy, we can determine whether each possible permutation of subjects results in a norm in favor of the (ex ante) minority preference. A methodological point to address here is why we are not analyzing actual occurrences of sequences. One issue with trying to analyze actual occurrences is that we would have needed subjects to go through many more rounds of the experiment to be able

[26]As pointed out in [Mullin and Reiley, 2006], we can generate hypothetical combinations (in our case hypothetical permutations) with experimental data so long as subjects provide complete strategies and the setting is consistent between sessions.

to collect sufficient data for each arrangement of preferences. This issue is exacerbated by the fact that each preference arrangement is not equally likely. The preference arrangement AAA is of little interest to our study as norm formation cascades are practically impossible here unless someone makes a mistake, but this arrangement is more likely than any other, expected to occur 34.3% of the time in the baseline treatment and a whopping 72.9% of the time in the small probability treatment. The preference sequences AAB and ABA are similarly uninteresting as norm formation cascades cannot happen for these either unless someone makes a mistake. These three relatively uninteresting preference arrangements nonetheless combine to make up 63.7% of all arrangements in the baseline treatment and 89.1% of all arrangements in the small probability treatment. Fortunately, we do not lose anything in our data by having subjects specify their complete strategies via the strategy method. Our experiment does not involve any form of communication between subjects or any other special form of interaction that the strategy method would fail to pick up. We therefore use strategies obtained from the strategy method to generate hypothetical permutations of subjects from which we can analyze whether norm formation cascades occur.

For each possible permutation of subjects, there are eight possible arrangements of preferences (e.g. all three subjects prefer B, only Position 1 prefers B, etc.). So, for each possible arrangement of preferences, we can calculate how many of the permutations of subjects lead to a norm for the minority preference (Action B). The outcome of these calculations can be found in Table 11. Column (1) in Table 11 shows the preference sequence of the three human positions, and column (2) indicates whether our theoretical model predicts that Action B will become the norm in that preference sequence for both the baseline and small probability treatment. Columns (3) and (4) of Table 11 provide the calculations for the whole sample of the baseline and small probability treatment, respectively. There were 42,840 $(36 * 35 * 34)$ permutations in our full baseline treatment and 120 $(6 * 5 * 4)$ permutations in our full small probability treatment. For each preference sequence, we have indicated the number of (hypothetical) subject permutations that would lead to Action B becoming the norm for the

group. Below each number is the corresponding percentage.

Looking at the results for the full sample, we notice that there are some significant deviations from our theoretical predictions. First, let's look at the outcomes for the baseline treatment. For the preference sequences BAB and BBA, we can see that Action B becomes the group norm the vast majority of times, 75.89% and 73.84% of the time, respectively, but not quite as often as predicted. Perhaps more striking is the outcome for the BAA preference sequence. As noted above, this is the sequence that would definitively count as a norm formation cascade whenever Action B becomes the group norm. While these definitive norm formation cascades are occurring in our data, 35.76% of permutations, they are occurring much less often than predicted. That said, if we look at the outcome for the ABB preference sequence, we see that this outcome deviates from the theoretical prediction in the opposite direction. The theoretical prediction for this sequence was that norm formation cascades would never occur, but in our data they would occur a significant amount of the time (34.21% of permutations). These results are suggestive that some subjects are not choosing their preference in Position 1, while others are choosing their preference in Position 2 even after the other action was chosen in Position 1. Looking at the outcomes for the small probability treatment, we notice that Action B becomes the group norm some of the time, even though our theoretical prediction was that Action B would never be the group norm in this treatment. Qualitatively, however, we can see that Action B becomes the group norm less often (in percentage terms) in the small probability treatment than in the baseline treatment, which matches our theoretical predictions.

Before we investigate the subjects' individual strategies, we may wish to recalculate the number of permutations that generate a norm formation cascade in favor of the minority preference under a certain restriction. Namely, that we restrict each of our treatments' samples to include only those subjects who provided the theoretically predicted strategy when in Position 3. We do this to account for potential noisy data points in our full sample. Unlike the choices in positions 1 and 2, there are no "reasonable" justifications for deviating from

the theoretically predicted choices in Position 3. If positions 1 and 2 choose the same action, then you should always choose that action. If positions 1 and 2 choose different actions, then you should always choose the action you prefer. You cannot benefit the payoffs of the other two group members by deviating from this rule; deviating from this rule only serves to harm your own payoff. Understanding that this is the optimal strategy in Position 3 thus serves as a basic indication that the subject has at least some degree of sophistication regarding the strategic environment. We classify a given subject as "sophisticated" if they make the correct decision in all eight possible scenarios they can face in Position 3. A total of 32 subjects in our sample, 27 from the baseline treatment and 5 from the small probability treatment, meet this condition of sophistication. In the sub sample of sophisticated subjects, there are 17,550 $(27 * 26 * 25)$ permutations in our baseline treatment and 60 $(5 * 4 * 3)$ permutations in our small probability treatment. Columns (5) and (6) of Table 11 show the results of our calculations for the sub sample. As we can see, the qualitative results for the sub sample are the same as for the full sample. That said, the outcomes for our sub sample are closer to our theoretical predictions. In the baseline treatment, for sequences BAA, BAB, and BBA, Action B becomes the group norm more often (in percentage terms) than in the full sample. For sequence ABB, Action B becomes the group norm less often than in the full sample. The outcomes for the small probability treatment are also closer to the theoretical prediction in the sub sample.

Given that we observe some significant deviations between our theoretical predictions and results for the frequency of norm formation cascades, we should next dive into the subjects' strategies and how choices within these strategies might deviate from our predictions. First we will look at subjects' choices for Position 1. Unsurprisingly, all 42 subjects choose Action A when they prefer Action A in Position 1, so there are no deviations from theory there. There are, however, deviations from our predictions when subjects prefer Action B. In the baseline treatment, our model predicts that subjects will choose Action B, but only 28 of 36 subjects chose B in this case. A t-test shows this difference to be statistically sig-

nificant ($p < 0.01$). In the small probability treatment, our model predicts that subjects will not choose Action B, but 3 of 6 subjects chose Action B in this case. A t-test shows this difference to be statistically significant ($p < 0.05$). As we did for our permutations of subjects above, we can redo our analysis just for the subset of "sophisticated" subjects. In the baseline treatment, we now have 23 of 27 subjects that chose B in this case. While not as extreme, a t-test shows this difference to be statistically significant ($p < 0.05$). In the small probability treatment, we now have 2 of 5 subjects that chose Action B in this case. A t-test shows this difference to be marginally significant ($p < 0.1$).

Now looking at subjects' choices in Position 2, we can first note that there were no significant deviations between our predictions and observed choices when the subject in Position 2 preferred the action that was chosen in Position 1. Unsurprisingly, in cases where Action A was chosen in Position 1 and the subject in Position 2 preferred Action A, the subject chose Action A. Likewise, in cases where Action B was chosen in Position 1 and the subject in Position 2 preferred Action B, the subject chose Action B. The interesting cases were when the subject preferred a different action from what was chosen in Position 1. In both the baseline and small probability treatments, our model predicts that when Action A was chosen in Position 1 and the subject in Position 2 prefers Action B, the subject will not choose Action B. In the baseline treatment, however, 13 of 36 subjects chose Action B in this case. A t-test shows this difference to be statistically significant ($p < 0.01$). In the small probability treatment, 2 of 6 subjects chose Action B in this case. A t-test shows this difference to be marginally significant ($p < 0.1$). As before, we can redo our analysis just for the subset of "sophisticated" subjects. In the baseline treatment, we now have 8 of 27 subjects that chose Action B in this case. A t-test shows this difference to be statistically significant ($p < 0.01$). In the small probability treatment, we now have 1 of 5 subjects that chose Action B in this case. A t-test shows this difference to *not* be statistically significant.

Now considering the case where Action B was chosen in Position 1 and the subject in Position 2 prefers Action A, in the baseline treatment our model predicts that the subject

will choose Action B, while in the small probability treatment our model predicts that the subject will *not* choose Action B. In the baseline treatment, only 16 of 36 subjects chose Action B in this case. A t-test shows this difference to be statistically significant ($p < 0.01$). In the small probability treatment, 2 of 6 subjects chose Action B in this case. A t-test shows this difference to be marginally significant ($p < 0.1$). As before, we can redo out analysis just for the subset of "sophisticated" subjects. In the baseline treatment, we now have 13 of 27 subjects that chose Action B in this case. A t-test shows this difference to be statistically significant ($p < 0.01$). In the small probability treatment, we now have 2 of 5 subjects that chose Action B in this case. A t-test shows this difference to be marginally significant ($p < 0.1$).

Comparing our results for each subject's choices in Position 1 and Position 2 to our hypotheses, we can see that a significant number of subjects did not play the theoretically predicted strategy. To get a better sense of how subjects' strategies deviated from our predictions, we can classify each subject's strategy into one of five types, as described at the end of Section 3: Theory, Optimist, Preference, Probability, and Deferential. Table 12 shows the number of subjects whose strategies were classified as each type. As we can see, 15 of the 42 subject strategies classified as Theory, meaning that 27 subject strategies deviated from our theoretical predictions in some meaningful, and usually consistent, manner. Of these 27 strategies, 13 classified as Optimist, 4 classified as Preference, 4 classified as Probability, 4 classified as Deferential, and the last 2 did not deviate from theory in a consistent manner and so were grouped as "Other". This frequency count indicates that a significant portion of the deviation from our theoretical predictions comes from subjects taking an Optimist-type strategy and trying to force their preference as the norm in Position 2. The other reason for the deviation from our predictions can come from the subjects who chose Deferential-type and did not choose their private consumption preference when the theoretical model predicted them to. As a check of the rationale for these Deferential strategies, we included in Table 12 a column indicating the average behavioral measure scores for each strategy

type. As we can see, the subjects who chose Deferential strategies had the lowest scores of any strategy type, suggesting that these subjects chose Deferential strategies because they cared about the other group member's payoffs and did not want to try force their minority preference as the group norm. A t-test comparing the average behavioral measure score of Deferential-type strategies to the the average behavioral measure score of all other types (the average score for the other 38 strategies was 7) shows this difference to be statistically significant ($p < 0.01$). Therefore, the results regarding our behavioral measure align with our hypothesis that subjects who choose Deferential-type strategies do so because of this behavioral consideration.

3.5 Conclusion

A common conception of social norms is that they represent the majority preference of the group. However, there seem to be cases, both inside and outside of the lab, where a given social norm does not represent the majority preference. That said, those existing examples do not provide any indication into the process in which social norms in favor of minority preferences form. The purpose of this paper was to investigate how social norms in favor or minority preferences can form, focusing on one potential mechanism we refer to as norm formation cascades. To study this issue, we presented a simple theoretical model which predicted that in certain cases, norm formation cascades in favor of a minority preference would occur. From the predictions of our theoretical model, we formed hypotheses that we could test with a laboratory experiment.

From the results of our experiment, we found that while norm formation cascades in favor of the minority preference occurred, they did not occur as often in the cases that our theoretical model predicted that they should occur. On the other hand, we also found that norm formation cascades in favor of the minority preference occurred in some cases that our theoretical model predicted that they should not occur. In total, we found that norm formation cascades in favor of the minority preference occurred in our experiment, but

they did not occur in quite the same manner as we had hypothesized. By taking a closer look at subjects' individual strategies, we found that these deviations from our theoretical predictions occurred for two main reasons. First, there were some subjects in Position 1 who did not choose according to their preference when our model predicted that they would. Second, there were other subjects who in Position 2 insisted on choosing according to their preference when our model predicted that they would not. From a behavioral measure we included in our experiment, we found that at least some of the reason for the former was that some subjects did not like trying to force their minority preference as the norm for the group.

The results of our study suggest a few avenues for future research into how social norms in favor of minority preferences can form. First, our experimental results suggest that enough of the subjects understood the strategic environment for norm formation cascades to form, but as mentioned in the experiment design we had automated the choices of the last three positions and only had subjects occupy the first three positions. To get a clearer picture about how norm formation cascades can occur outside the lab, future studies could study the effect of putting subjects into the last three positions. Having subjects in those last three positions would potentially complicate the considerations of the subjects in the first three positions as there is now the possibility of mistakes in the later positions, but it is not obvious how much this would change the occurrence of norm formation cascades. Second, we could look deeper into the motivation of subjects who insist on choosing according to their preference in Position 2. One possibility is that these subjects are just simply optimistic that they can change the group's norm in their favor. Another possibility is that these subjects feel entitled to having the norm match their private consumption preference, to the point where they would be willing to significantly lower their expected payoff just to have a chance at forming the group's norm in their favor. A third potentially interesting avenue for future research is to make each subject's position in a group's sequence endogenous. Our theoretical model predicts that, at least in our baseline treatment, the subject in Position 1

has the ability to dictate which direction the group's norm will go. Although norm formation cascades did not occur quite in the same manner has our model predicted, our results showed that the subject in Position 1 had a strong influence over which direction the social norm would go. One interesting question here is how much more eager would a subject with a minority preference be to be in Position 1 compared to a subject with a majority preference (e.g. by comparing subject's willingness to pay or some related measure of eagerness). In summary, the present paper offers initial indications about how social norms in favor of minority preferences can form, and our results offer potential avenues into further research into the issue of how social norms form.

(1) Sequence	(2) Theory	(3) Full Base	(4) Full Small	(5) Soph Base	(6) Soph Small
BAA	Base: Yes	15,320	20	7,200	0
	Small: No	(35.76%)	(16.67%)	(41.03%)	(0.00%)
ABB	Base: No	14,654	40	5,200	12
	Small: No	(34.21%)	(33.33%)	(29.63%)	(20.00%)
BAB	Base: Yes	32,510	60	14,950	24
	Small: No	(75.89%)	(50.00%)	(85.19%)	(40.00%)
BBA	Base: Yes	31,635	48	14,950	18
	Small: No	(73.84%)	(40.00%)	(85.19%)	(30.00%)
BBB	Base: Yes	36,540	84	15,750	33
	Small: No	(85.29%)	(70.00%)	(89.74%)	(55.00%)
AAB	Base: No	1,156	0	0	0
	Small: No	(2.70%)	(0.00%)	(0.00%)	(0.00%)
ABA	Base: No	884	0	0	0
	Small: No	(2.06%)	(0.00%)	(0.00%)	(0.00%)
AAA	Base: No	68	0	0	0
	Small: No	(0.16%)	(0.00%)	(0.00%)	(0.00%)

Table 11: Frequency of Action B Becoming Norm for Each Preference Sequence Compared to Theoretical Prediction

Strategy Type	N	Avg Behavioral Measure
Theory	15	6.6
Optimist	13	6.31
Preference	4	9
Probability	4	8.25
Deferential	4	4.5
Other	2	8
All	42	6.76

Table 12: Frequency and Average Behavioral Measure Score of Strategy Types

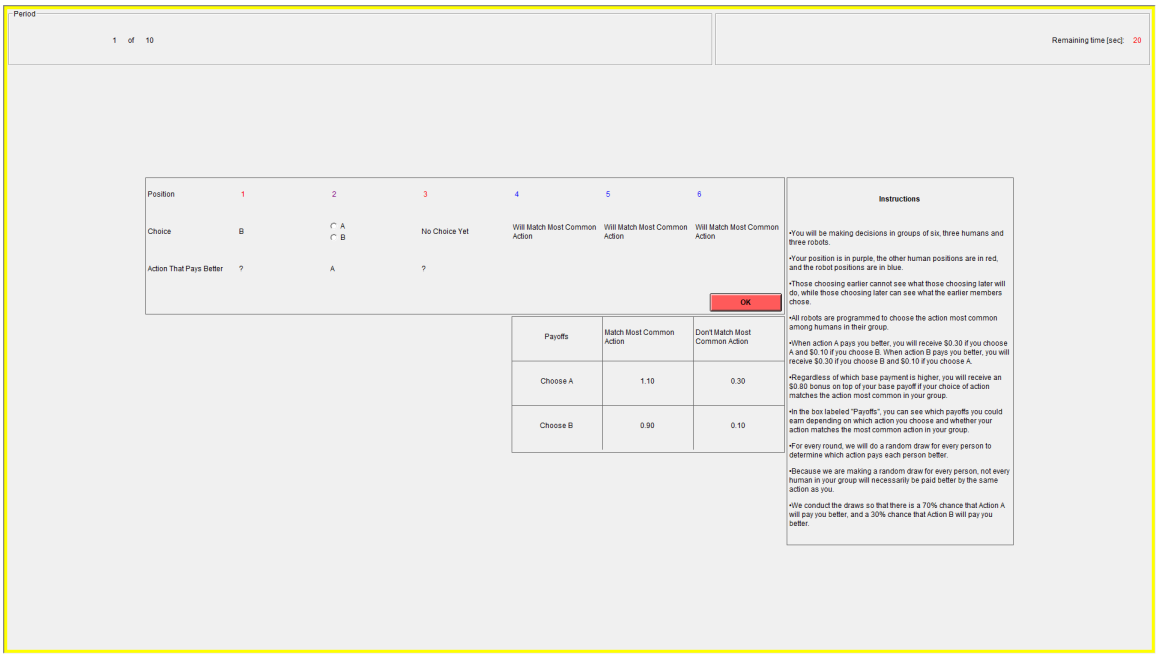

Figure 6: Phase 1 Choice Screen for Position 2

Figure 7: Phase 2 Choice Screen for Position 2

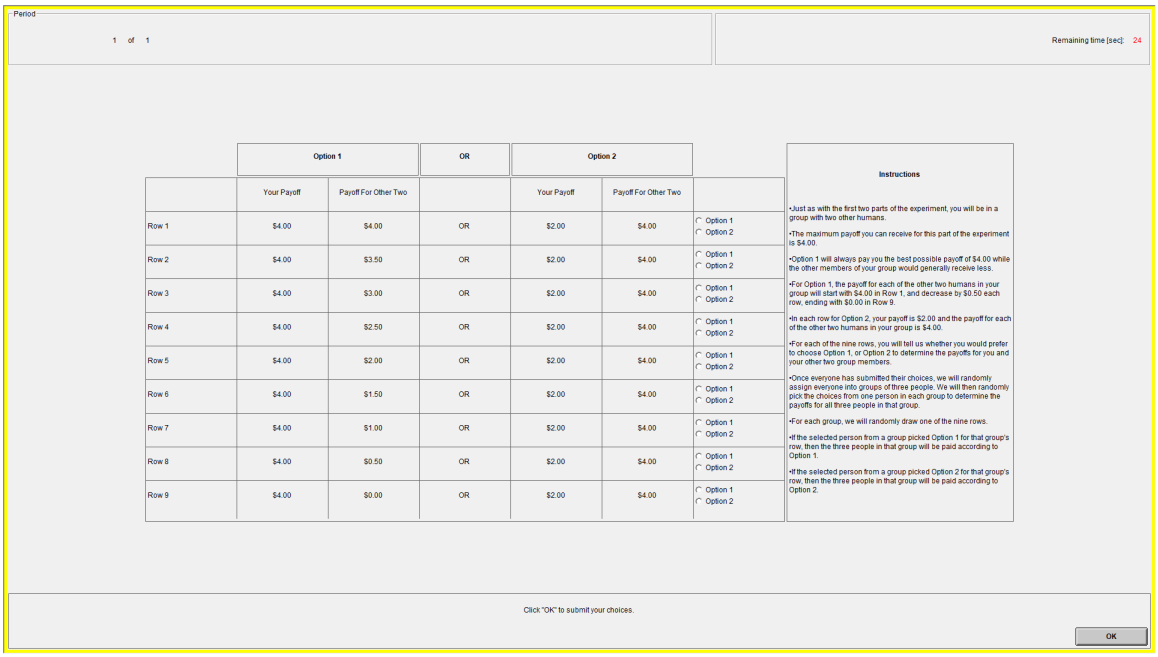

Figure 8: Behavioral Measure Choice Screen

APPENDIX

Survey Instrument for Chapter 1

Informed Consent

We are conducting a research study to learn more about knowledge people possess regarding public policy related issues. Your participation in this study is voluntary. If you agree to take part and then change your mind, you can withdraw for any reason. There are no penalties if you withdraw, or decline to participate in the study. If you agree to participate, you will be asked to respond to a series of questions in which we will ask about a variety of public policy, political and cultural issues. We will also ask a few basic questions about demographics and include a few additional problem solving questions. Your participation should take about 10-20 minutes. There are no expected risks from participating in this survey. The results from this study will help our team of researchers better understand the level of knowledge the public possesses on these issues. You must be at least 18 years old to participate. Would you like to participate in this research study? If so, please click on the button to consent to continue the study.

Demographic Questions

Please enter your Prolific ID:

What is the highest level of education you have achieved?

- Less than High School/GED
- High School/GED
- Some college but no degree

- Associate's degree
- Trade/technical/vocational training
- Bachelor's degree
- Master's degree
- Doctorate or other terminal degree

How much total combined income did all members of your household earn last year?

- $0 to $24,999
- $25,000 to $49,999
- $50,000 to $99,999
- $100,000 to $199,999
- $200,000 and up

Do you consider yourself a Republican, a Democrat, or an Independent?

- Strongly Democratic
- Weakly Democratic
- Independent (lean Democratic)
- Independent (do note lean towards either party)
- Independent (lean Republican)
- Weakly Republican
- Strongly Republican
- Other

What is the zip code where you live for most of the year?

No Incentive Treatment

For the upcoming set of questions, we will be asking questions about public policy, political and cultural issues. We expect that most people may not know the precise answers to many of these questions. We would like for you to give your best guess of the actual answer. You will have a short period of time to answer each one and we request that you not search for the answer but rather provide what you think the answer is. You will also be able to answer to indicate that you do not know. For each of the questions you will enter your answer using either a slider bar or a text entry box. For each question you will have the option to enter your choice as "I don't know." If you do not actively make an entry using the slider bar or text box or click the box for "I don't know" and continue through a question or the timer moves you along, no response will be recorded for that question. We will show you the correct answers at the end of the survey if you wish to check them.

Incentive Treatment

For the upcoming set of questions, we will be asking questions about public policy, political and cultural issues. We expect that most people may not know the precise answers to many of these questions. We would like for you to give your best guess of the actual answer. You will have a short period of time to answer each one and we request that you not search for the answer but rather provide what you think the answer is. You will also be able to answer to indicate that you do not know. For each of the questions you will enter your answer using either a slider bar or a text entry box. For each question you will have the option to enter your choice as "I don't know." If you do not actively make an entry using the slider bar or text box or click the box for "I don't know" and continue through a question or the timer moves you along, no response will be recorded for that question. We will show you the correct answers at the end of the survey if you wish to check them.

After the survey is concluded we will randomly select 5 of your answers to compare to the true answer. For each question, if you are close enough to the correct answer, we will pay you

$1. This means that you can earn up to $5 for answering these questions accurately. When we show you the correct answers in the end, we will also note the range that we considered close enough to earn payment. For any question considered for payoff, if you have entered "I don't know" you will earn $0.25.

Political Fact Questions

The U.S. Bureau of Labor Statistics counts a person as unemployed if the person is not employed at any job and is looking for work. By this definition, 6.3 percent of Americans were unemployed in January 2021, the beginning of President Biden's term in office. What percentage of Americans are currently unemployed? (Slider ranging from 0 to 10)

The outstanding public debt of the United States is the total amount of money owed by the federal government. Every year the government runs a deficit, the size of the public debt grows. Every year the government runs a surplus, the size of the public debt shrinks. In January of 2017, when President Trump took office, the outstanding public debt of the United States was approximately 19.9 trillion dollars. What was the size of the outstanding public debt when President Trump left office in January 2021 in trillions of dollars? (Slider ranging from 10 to 35)

In September 2021, the Census Bureau reported how many Americans live in poverty. The poverty threshold depends on the size of the household. For example, a person under age 65 is considered to live in poverty if his or her 2020 income was below $13,465 and a family of four is considered to live in poverty if its 2020 income was below $26,496. By this definition, what percentage of Americans lived in poverty in 2020? Please type your answer as a percentage (e.g. 6 = 6%).

According to the American Automobile Association (AAA), the national average price for a gallon of regular gasoline one year ago (October 2020) was $2.19. What is the current

national average price for a gallon of regular gasoline? (Slider ranging from 1.5 to 4.5)

About how many U.S. soldiers were killed in Afghanistan between 2017 and the time at which troops were withdrawn in August of 2021?

According to the U.S. Census Bureau, in 2019 what percentage of the total population of the United State was born outside of the United States (this includes both legal and non-legal individuals living in the US at that time)? Please type your answer as a percentage (e.g. 6 = 6%).

In the 2020 Presidential Election, Democrat Joe Biden defeated his Republican opponent Donald Trump. In the nation as a whole, according to the results certified by the Federal Election Commission, what percentage went to Biden? Please type your answer as a percentage (e.g. 6 = 6%).

For every dollar spent in fiscal year 2020, what percentage went to the:

Department of Defense (Slider ranging from 0 to 50)

Social Security (Slider ranging from 0 to 50)

Interest payments on national debt (Slider ranging from 0 to 50)

In billions of dollars, how much did the US spend on the following programs in 2020:

Food and Nutrition Assistance (Food Stamps or SNAP, child nutrition programs, etc ...) Please type your answer as dollars (in billions) (e.g. 6 = $6 billion).

Farm Income Stabilization (e.g. farm subsidies, farm insurance programs etc ...) Please type your answer as dollars (in billions) (e.g. 6 = $6 billion).

Border Patrol Please type your answer as dollars (in billions) (e.g. 6 = $6 billion).

Based on data collected by the Centers for Disease Control and Prevention (CDC), in 2019 out of every 10 medically treated firearm injuries in the US, how many are due to unintentional firearm accidents? (Slider ranging from 0 to 10)

According to a database maintained by the conservative leaning Heritage Foundation documenting all cases of voter fraud they can identify from 1982 up to 2021, how many documented cases exist of non-citizens registering to vote in US elections?

According to data collected by the Centers for Disease Control and Prevention (CDC), for every million people who receive the following types of vaccinations, how many experience severe, and sometimes life-threatening, allergic reactions including Anaphylaxis?

COVID 19

MMR (Measles, Mumps and Rubella)

According to data collected by the CDC, how effective are the two mRNA COVID-19 vaccines at preventing symptomatic illness? Please type your answer as a percentage (e.g. 6 = 6%).

Based on data collected by the CDC:

What percentage of unvaccinated people have been hospitalized due to the coronavirus? (Slider ranging from 0 to 10)

What percentage of fully vaccinated people have been hospitalized due to the coronavirus? (Slider ranging from 0 to 10)

According to the National Highway Traffic Safety Administration, by how much does wearing a seatbelt in the front seat of a passenger car reduce your risk of fatal injury? Please type your answer as a percentage (e.g. 6 = 6%).

According to the data collected by the US Bureau of Economic Analysis, what was the average annual growth rate in GDP for the following 3-year periods:

End of Obama Presidency, 2014-2016 (Slider ranging from 1 to 10)

Beginning of Trump Presidency, 2017-2019 (Slider ranging from 1 to 10)

No Incentive Treatment Revise

Thank you for answering these questions. We have chosen three of your answers at random to possibly generate additional earnings for you. We will compare your answers to the true answers and pay you $1 for each of your answers that is close enough to the true answer. For any question you answered with "I don't know", you will earn $0.25 if it is chosen for payment. You can choose to submit the answers you have already chosen or you can choose to edit your answers.

Would you like to edit your answers or submit ones already entered?

- I would like to edit my answers
- I would like to submit the ones already entered

Incentive Treatment Revise

Thank you for answering these questions. We have chosen three of your answers at random to possibly generate additional earnings for you. On top of the incentives we have already provided for accurate answers we will pick three more questions and double the potential earnings to $2 for a correct answer. We will compare your answers to the true answers and pay you $2 for each of your answers that is close enough to the true answer. For any question you answered with "I don't know", you will earn $0.50 if it is chosen for payment. You can choose to submit the answers you have already chosen or you can choose to edit your answers.

Would you like to edit your answers or submit ones already entered?

- I would like to edit my answers
- I would like to submit the ones already entered

No Incentive Treatment CRT

For the final 3 questions, we will give you three problems to solve. Please give us your best guess at the answer to each. You will have 45 seconds to answer each question.

Incentive Treatment CRT

For the final 3 questions, we will give you three problems to solve. Please give us your best guess at the answer to each. You will have 45 seconds to answer each question.

We will pick one of these questions at random and if your answer to that question is correct you will earn $1.

CRT

If 3 elves can wrap 3 toys in 1 hour, how many elves are needed to wrap 6 toys in 2 hours?

In an athletic team, tall athletes are three times more likely to win a medal than short athletes. This year the team has won 60 medals so far. How many of those medals were won by short athletes?

A bat and a ball together cost 110 cents. The bat costs 100 cents more than the ball. How much does the ball cost? Please type you answer in cents.

Survey Instrument for Chapter 2

Informed Consent

We are conducting a research study to learn more about how behavior in certain incentivized tasks that deal with economically relevant behavior may vary with demographic characteristics. Your participation in this study is voluntary. If you agree to take part and then change your mind, you can withdraw for any reason. There are no penalties if you withdraw or decline to participate. If you agree, you will take the following survey. The survey will ask for your choices in various incentivized tasks as well as a few basic demographic questions which include a few questions regarding political viewpoints. Your participation should take about 20 minutes. You will not be asked any personal questions about sensitive topics, nor will you be asked any questions that would compromise your confidentiality. Potential benefits of this study include a better understanding of how individuals make choices in various settings. Would you like to participate in this research study? If you do, please select "I consent to participating in this study" and continue to the next page. If you do not, please select "I do not consent to participating in this study" and follow the directions for returning your submission on Prolific.

If you choose to participate, you will receive a base payment of $2.50 for completing the survey and answering basic demographic/opinion questions. In addition to the basic demographic/opinion questions, there will be an incentivized portion, for which you can earn an additional bonus payment. Make sure to carefully read all instructions and complete all tasks. You will be asked a couple of comprehension questions throughout the survey to make sure you are reading directions. If all tasks are completed with appropriate responses and you do not fail both comprehension questions, we will approve your submission and pay your base amount of $2.50 within one day. It will take us some time to calculate your bonus payments, so we will pay you your bonus amount shortly after your submission is approved. You will receive payment of bonus amounts within two weeks of your submission.

This survey is part of a research project being conducted by researchers at Southern

Methodist University (SMU).

Demographics

What is your current age?

Please enter your Prolific ID:

What is the highest level of education you have achieved?

- Less than High School/GED
- High School/GED
- Some college but no degree
- Associate's degree
- Trade/technical/vocational training
- Bachelor's degree
- Master's degree
- Doctorate or other terminal degree

How much total combined income did all members of your household earn last year?

- $0 to $24,999
- $25,000 to $49,999
- $50,000 to $99,999
- $100,000 to $199,999
- $200,000 and up

Are you currently employed on a full or part time basis?

- Yes
- No

With which ethnicity do you identify?

- White
- Mixed
- Asian
- Black
- Other

How would you describe your gender?

- Male
- Female
- Other/Prefer not to say

Do you usually think of yourself as

- Republican
- Democrat
- Independent
- Other (please specify)

What is the zip code where you live for most of the year?

Please describe the area you live in

- Urban
- Rural

Are you generally a person who is fully willing to take risks or do you try to avoid taking risks? Please select on the scale below where 0 means "not at all willing to take risks" and 10 means "fully prepared to take risks" (Slider ranging from 0 to 10)

Are you generally an impatient person, or someone who always shows great patience? Please select on the scale below where 0 means "very impatient" and 10 means "very patient" (Slider ranging from 0 to 10)

Factual Political Questions

For the next portion of the survey, we will ask about your opinion/beliefs regarding various political issues and current events.

How does President Joe Biden's approval rating at 100 days in office, as measured by Gallup, compare with the 100-day approval ratings of former President George W. Bush and former President Donald Trump?

- Biden's 100-day approval rating is lower than Bush's and lower than Trump's
- Biden's 100-day approval rating is lower than Bush's and the same (within 1%) as Trump's
- Biden's 100-day approval rating is lower than Bush's and higher than Trump's
- Biden's 100-day approval rating is the same (within 1%) as Bush's and higher than Trump's

- Biden's 100-day approval rating is higher than Bush's and higher than Trump's

How much did the unemployment rate increase or decrease during Obama's presidency (January 2009 - January 2017)?

- increased by more than 5 percentage points
- increased by 2-5 percentage points
- increased by less than 2 percentage points
- decreased by less than 2 percentage points
- decreased by 2-5 percentage points
- decreased by more than 5 percentage points

A major campaign promise Donald Trump made in 2016 was that he would build a wall on the United States' southern border. How much border wall/fence was built during the Trump administration (January 2017 – January 2021)? This includes both completely new structure (where none previously existed) and fencing put up to replace existing structure.

- less than 150 miles
- between 150 and 300 miles
- between 300 and 450 miles
- between 450 and 600 miles
- more than 600 miles

In 2008, the year before former President Obama took office, it was estimated that just under 12 million unauthorized immigrants were living in the United States. How did this number change over the course of Obama's administration (2009-2017)?

- increased substantially (by 2 million or more)

- increased somewhat (by fewer than 2 million)
- decreased somewhat (by fewer than 2 million)
- decreased substantially (by 2 million or more)

Conspiracist Ideation Questions

There is often debate about whether the public is told the whole truth about various important issues. Please indicate the degree to which you believe each statement is likely to be true on the following scale: Definitely not true; Probably not true; Not sure/cannot decide; Probably true; Definitely true.

- The power held by heads of state is second to that of small unknown groups who really control world politics
- A small, secret group of people is responsible for making all major world decisions, such as going to war
- Certain significant events have been the result of the activity of a small group who secretly manipulate world events
- The spread of certain viruses and/or diseases is the result of the deliberate, concealed efforts of some organization
- Experiments involving new drugs or technologies are routinely carried out on the public without their knowledge or consent
- The government uses people as scapegoats to hide its involvement in criminal activity
- The government permits or perpetrates acts of terrorism on its own soil, disguising its involvement
- The government is involved in the murder of innocent citizens and/or well-known public figures, and keeps this a secret

Coin Flip Task

Over the course of this survey, you will be asked various questions where you will be paid based on answers/responses you give. This bonus payment will be on top of the $2.50 you will earn for completing the survey. To make sure all survey participants receive a base amount from the incentivized portion of the survey, we are using a simple mechanism to assign a base random payment. We ask that you find any coin you have nearby and flip it 5 times. Please do so and then report how many times it comes up heads rather than tails. You will be paid $0.10 for each time you report that it came up heads. (Slider ranging from 0 to 5)

Pick a Painting Task

For the next interaction, we need you to first tell us which of two paintings you prefer. You will be paid $0.50 for stating this preference and your choice will have an impact on the subsequent (and only the subsequent) interaction.

Which painting do you prefer?

- Painting 1
- Painting 2

Entitlement Task - Picked Painting 2

Those who chose painting 2, as you did, have been awarded an additional $1 on top of the original $0.50 received for stating a painting preference. Those who chose painting 1 are not entitled to this award but still can receive the $0.50 for stating a painting preference. You have the option of taking your additional prize or deferring it to an individual who stated a preference for painting 1. We will pair this choice up with the choices of an individual who preferred painting 1 to determine your payoff. If you defer the prize to this person and they defer it back to you, you get the $1. If you defer the prize and they claim it, they give up their original $0.50 but they receive the $1 intended for you.

If you choose to take the prize and they defer it to you, you keep the $1 that you were awarded. If you choose to take it and they also try to take it, then the final recipient will be determined probabilistically with you receiving it with 1/3 probability, the other receiving it with 1/3 probability, and no one receiving it with 1/3 probability. If the other person tries to take the award, they give up their original $0.50 to make the attempt to take the award. Regardless of your choices, you will always keep the original $0.50 payment.

This interaction is summarized in the diagram below. Your choices and payments are shown in blue.

Do you wish to defer your award to another or claim the award intended for you?

- Take
- Defer

Entitlement Task - Picked Painting 1

Those who chose painting 2 have been awarded an additional $1 on top of the original $0.50 received for stating a painting preference. You chose painting 1 and so are not entitled to this award but you still can receive the $0.50 for stating a painting preference. There is however a possibility that you could end up with the additional $1 award. We will pair your choices up with those of someone who received the award to determine your payoff. They have the option of taking their additional prize or deferring it to you. If they defer it to you, you can claim the award or defer it back to them. If you take it, you give up your original $0.50 but you get the $1. If you defer it back to the intended recipient, they receive the $1 they were awarded but you keep your original $0.50.

If, however, they choose to directly take the prize they have been awarded, you will still be able to attempt to take the award from them. You can try to take the award the other claimed or you can defer to them. If you defer, they keep the $1 they were awarded but you keep your original $0.50. If you try to take it then the final recipient will be determined

probabilistically with you receiving it with 1/3 probability, the other receiving it with 1/3 probability, and no one receiving it with 1/3 probability. If you try to take the award, you give up your original $0.50 to make the attempt to take the award. Thus, if you choose to try to take the award, you will have a 1/3 chance of receiving the $1 and 2/3 chance of receiving $0.

This interaction is summarized in the diagram below. Your choices and payments are shown in red.

If the recipient of an award defers the $1 to you, do you choose to give up your $0.50 and take it or defer it back to the original recipient?

- Take
- Defer

If the recipient of an award chose to accept the award, do you wish to give up your $0.50 and attempt to take the $1 or defer the award to the original recipient?

- Take
- Defer

Area Code Question

As another way of ensuring respondents receive compensation for the incentivized part of the survey, we are using a second simple mechanism to assign a base random payment. Please enter in the middle digit of the area code of the first phone number you remember using in the United States. You will receive a payment of $0.05 plus $0.05 times the number you enter.

Please enter the middle digit from the area code from the first phone number you remember using in the United States.

- The area code's middle digit is 0 (payment = \$0.05)
- The area code's middle digit is 1 (payment = \$0.10)
- The area code's middle digit is 2 (payment = \$0.15)
- The area code's middle digit is 3 (payment = \$0.20)
- The area code's middle digit is 4 (payment = \$0.25)
- The area code's middle digit is 5 (payment = \$0.30)
- The area code's middle digit is 6 (payment = \$0.35)
- The area code's middle digit is 7 (payment = \$0.40)
- The area code's middle digit is 8 (payment = \$0.45)
- The area code's middle digit is 9 (payment = \$0.50)

Trust Game First Mover

You will now be asked to engage in an interaction with another survey respondent by making a series of choices. Your role in this interaction was randomly determined. You can find a diagram depicting the nature of the interaction below. Your choices and payments are shown in blue. You will first have a choice of ending the interaction immediately or continuing it. If you end it immediately you and the other individual will both receive \$0.50. If you continue, then the other individual gets to choose to end or continue as well. If they choose to end the interaction, then you will only receive \$0.40 while they will receive \$1.10. If they choose to continue the game, then you both receive \$0.75 which is greater than what you would get by ending the interaction immediately.

Do you wish to end the interaction immediately or continue and let the other choose?

- End
- Continue

We now ask you to make several choices between two options for receiving earnings from a lottery. Both options will always give you either a payment of $0.75, or a payment of $0.40. If you choose Option 2, the probability of $0.75 will always be equal to the actual percentage of second movers who chose to continue in the preceding exercise. The probability of $0.40 will then be the percentage of second movers who chose to end. We will not tell you what those probabilities are as we cannot know until the choice data has been collected. If you choose Option 1, the probability of each payment is specified in the table below. For each of the rows the probability of the $0.75 payoff declines with the probability of the $0.40 payoff rising.

Comprehension Question 1

Based on the text above, and table below, what determines the probability of receiving $0.75 from Option 2?

- The percentage of second movers who chose End
- The percentage of second movers who chose Continue

In row 1, Option 1 delivers $0.75 with certainty and so most would prefer to choose Option 1 in row 1. As you move down the rows, the probability of this payoff declines suggesting that at some point Option 2 may be preferred. We will therefore assume that everyone would prefer Option 1 in the first row and simply ask you to indicate at which row, if any, you would like to switch to choose Option 2 and continue choosing Option 2 for all subsequent rows. On the slider below, indicate for which row you would switch from preferring Option 1 to preferring Option 2.

We will randomly pick one of the 11 rows to generate a payoff from this task. After picking a row we will determine a payoff based on taking a draw from the option you chose for that row. (Slider ranging from 1 to 11)

Trust Game Second Mover

You will now be asked to engage in an interaction with another survey respondent by making a series of choices. Your role in this interaction was randomly determined. You can find a diagram depicting the nature of the interaction below. Your payments and choices are shown in red. You will be the second mover in the game. The person you are matched with can choose to end the game immediately leading to both of you receiving payments of \$0.50. If they choose to continue then you will be able to make a choice. You can choose to end or continue as well. If you choose to end the interaction, then the other individual will only receive \$0.40 while you will receive \$1.10. If you choose to continue the game, then you both receive \$0.75.

If the other individual chooses to continue the game and allow you to choose, would you choose to end the interaction, or to continue the game?

- End
- Continue

We now ask you to make several choices between two options for receiving earnings from a lottery. Both options will always give you either a payment of \$0.75, or a payment of \$0.40. If you choose Option 2, the probability of \$0.75 will always be equal to the actual percentage of first movers who chose to continue in the preceding exercise. The probability of \$0.40 will then be the percentage of first movers who chose to end. We will not tell you what those probabilities are as we cannot know until the choice data has been collected. If you choose Option 1, the probability of each payment is specified in the table below. For each of the rows the probability of the \$0.75 payoff declines with the probability of the \$0.40

payoff rising.

Comprehension Question 1

Based on the text above, and table below, what determines the probability of receiving $0.75 from Option 2?

- The percentage of first movers who chose End
- The percentage of first movers who chose Continue

In row 1, Option 1 delivers $0.75 with certainty and so most would prefer to choose Option 1 in row 1. As you move down the rows, the probability of this payoff declines suggesting that at some point Option 2 may be preferred. We will therefore assume that everyone would prefer Option 1 in the first row and simply ask you to indicate at which row, if any, you would like to switch to choose Option 2 and continue choosing Option 2 for all subsequent rows. On the slider below, indicate for which row you would switch from preferring Option 1 to preferring Option 2.

We will randomly pick one of the 11 rows to generate a payoff from this task. After picking a row we will determine a payoff based on taking a draw from the option you chose for that row. (Slider ranging from 1 to 11)

Creative Uses Task

In the next segment you will be given three minutes to come up with alternative uses for an object displayed on your screen. The uses you come up with must be different than the use for which the object was originally designed or intended. As an example, suppose the object that appeared was "a tin can." Some uses which you would get credit for could be a "flower pot", a "holder for my dreams", "attach a string between two cans to make telephones", "cut the can and make a pinwheel", etc. Intended uses which you would not get credit for, would involve things like "storing carrots" or "container for evaporated milk" since these are uses the tin can was intended for. Also, you would not get credit for uses that are the same. In the example above, if you had put "a daisy pot" and "a petunia pot" this

would get counted as one valid use since both are considered flower pots. You will see the remaining time on your computer (in the top left corner). When the time is up, you will not be able to enter new entries. You will enter all of your uses into a single entry box, separated by a comma, so do not click the "next" button until you have typed all of your answers in the text entry box, as this will end the task prematurely. Please answer the question below and then continue to the next page for an explanation of how you will be paid for this task.

Comprehension Question 2

Based on the text above, when should you click the "next" button when entering proposed uses for the object?

- After typing the first use
- Only after typing all uses I can think of, separating each use with a comma

For this task, you will receive $0.15 for each valid use which conforms with the rules. In addition, your answers will be compared to those of 9 other people taking this survey. For any entries you submit which are unique among that set, you will receive an additional $0.15. Your responses will be judged by two independent judges (not the experimenters) to determine whether or not they satisfy the requirements. Please continue to the next page to start the task.

You will have three minutes for this task (your remaining time is in the top left corner). You may take less time if you want to, but you will be automatically moved to the next page when the three minutes are up. Do not click the "next" button until you have typed all of your answers in the text entry box, separating each entry by a comma, as this will end the task prematurely.

Your object is an Extension Cord: an electric cord fitted with a plug at one end and a receptacle at the other used to power electronic devices.

Propensity to Rely on Own Judgement Task

You will now have the chance to earn an additional payment by choosing between one of two options labeled A and B. Regardless of which you choose, you will have the opportunity to earn either \$1.00, \$0.75, \$0.50 or \$0.25. The difference between A and B is that the likelihood of each of those prize amounts differs between the two options. When making your choice you will not know the probability of winning these prizes from either option, so you will not know which option might generate the highest payment on average. We will tell you that one of the options pays \$0.25 more on average than the other option.

To help with your decision, you can gain information about these options through one of two ways. First, we have conducted a prior version of this survey where other participants faced this same choice between these same two alternatives. These individuals were allowed to engage in 10 trials in which they could choose between each option and see the outcome of each trial prior to making a final choice for payment. We then asked these individuals which option they would suggest for those facing this choice in the future. We will offer you the chance to take one of these recommendations picked at random from the set of all suggestions made and have this recommendation be your choice.

Alternatively, you can pay \$0.01 (\$0.05) (\$0.10) of money you have already earned in this survey to engage in 10 test trials of your own before making your final decision. You will be able to make a choice between A and B 10 times and see the outcome from each choice, and then we will ask you to make your choice to determine your payment. You will not be able to see a recommendation from a previous player.

Do you want to accept the advice from a prior participant or test the options yourself 10 times for a price of \$0.01 (\$0.05) (\$0.10)?

- Accept the advice from a prior participant
- Test the options myself for a price of \$0.01 (\$0.05) (\$0.10)

You have chosen to engage in 10 test rounds of your own. After you have completed the 10 test rounds, we will show you a summary of your 10 test rounds before you have to make your choice. Continue to the next page to begin your 10 test rounds.

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

Choose option A or B to take a test draw and see how much you could earn from that option.

- Option A
- Option B

Given the outcomes you have observed from A and B, please make your final choice between them to determine your earnings.

- Option A
- Option B

Experiment Instructions for Chapter 3

Thank you for participating in today's experiment. I will read through a script to explain to you the nature of today's experiment as well as how to navigate the computer interface with which you will be working. The instructions I will read will be duplicated on your computer screen. I will be using this script to make sure that all sessions of this experiment receive the same information.

This is an experiment in decision-making. In addition to a $10 participation fee for showing up on time, you will be paid any money you accumulate from interactions that will be described to you in a moment. You will be paid privately, by check, at the conclusion of the experiment. The exact amount you receive will be determined during the experiment and will depend on your decisions and the decisions of others. All monetary amounts you will see in this experiment will be denominated in dollars. If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Participants intentionally violating the rules may be asked to leave the experiment with only your show-up payment.

The Game

Today's experiment will consist of three phases.

Throughout the experiment, you will be making decisions in groups of six, three humans and three robots. The humans and robots in your group will each make a choice between two actions: Action A and Action B. Your payoff from this interaction will be determined by your choice and the choices of the others in your group. In each round, one action will give you a better base payoff than the other, but you will also generally earn more money if your action matches the actions of the other members in your group.

All members of the group will be making their choices sequentially. Those choosing earlier cannot see what those choosing later will do, while those choosing later can see what

the earlier members chose. The three humans will choose in the first three positions while the three robots will always choose in the last three positions. All robots are programmed to choose the action most common among humans in their group. After all three humans and all three robots have submitted their choice of action, your payoff will be determined.

Your payoff for this interaction is calculated in the following manner. In some rounds, action A will pay you better, and in some rounds action B will pay you better. When action A pays you better, you will receive $0.30 if you choose A and $0.10 if you choose B. When action B pays you better, you will receive $0.30 if you choose B and $0.10 if you choose A. Regardless of which base payment is higher, you will receive an $0.80 bonus on top of your base payoff if your choice of action matches the action most common in your group.

Please click "Continue" to move to the next screen. We will now go through an example to help you familiarize yourself with how the payoffs work. In this example, you will be selecting choices for the first three positions (the human positions, which are in red). Positions 4 through 6 are the robot positions (which are in blue), so they will just pick whatever action is most common among humans. Notice which actions pay each position better. In this example, action A pays Position 1 better, action B pays Position 2 better, and action B pays Position 3 better. As a first example, choose "A" for Position 1, choose "A" for Position 2, and choose "A" for Position 3. Notice how the robots have all chosen action A in this example. If you look at the bottom two rows, you will see the actions each position took as well as the payoffs each position would earn from this outcome. These two rows will not be on the main decision screen, they are only on this screen to help you familiarize yourself with how the payoffs work. The most common action is action A. Since all three human positions chose action A, they each received the bonus payment of $0.80. Action A pays Position 1 better, so he received $0.30 plus the bonus for a total of $1.10. Action B pays Position 2 and Position 3 better, so they each received $0.10 plus the bonus for a total of $0.90 each. To make sure you understand how the payoffs work, click the reset button for each position and we will do another example.

As second example, choose "A" for Position 1, choose "B" for Position 2, and choose "B" for Position 3. Notice how the robots have now all chosen action B. If you look back at the bottom row labeled "Payoffs", you will see the payoffs for each position from this new outcome. The most common action is now action B. Since Position 2 and Position 3 chose action B, they each received the bonus payment of \$0.80. Position 1 chose action A, so he did not receive the bonus. Action A pays Position 1 better, so he received \$0.30 but no bonus for a total of \$0.30. Action B pays Position 2 and Position 3 better, so they each received \$0.30 plus the bonus for a total of \$1.10 each. Are there any questions about how these payoffs were determined? If there are no questions, you can click "reset" for each position to reset your example screen.

For the next few minutes, you may continue to pick actions for each position to see what would happen in various possible outcomes. When you feel comfortable with your understanding of how choices lead to payoffs, please click "OK." Please feel free to ask questions if you do not understand how the payoffs work. Once everyone is finished, we will begin the first phase of the experiment.

Instructions for Stage 1 (Periods 1-10)

You will now begin your group interactions. In this first part of the experiment, you will play through 10 rounds of the interaction to help you familiarize yourself with making choices in each position. At the beginning of each round, you will be randomly assigned to a group of three humans and three robots. You will also be randomly assigned to one of three positions in your group's sequence: Position 1, Position 2, or Position 3. For every round, we will do a random draw for every person to determine which action pays each person better. Because we are making a random draw for every person, not every human in your group will necessarily be paid better by the same action as you. We conduct the draws so that there is a 70% (90%) chance that Action A will pay you better, and a 30% (10%) chance that Action B will pay you better. This means that, on average, 7 (9) out of 10 times Action

A will pay a player better, and 3 (1) out of 10 times Action B will pay a player better. In any round you may end up with having the same action that pays you better as both of the other human players, only one of them or neither of them.

When it is your turn to choose an action, we will show you the choices of any previous positions, and we will show you who in your group has yet to make a choice. Remember, those who have not chosen yet will be able to see your choice when it is their turn to choose. Once you have made your choice of action, you can press "OK" to move to the next position's turn to choose.

Once every position has made a choice of action, you will be presented with a results screen for that period. The summary screen includes the action chosen for each position along with a brief summary of the payoff you earned for the interaction. After every group has completed their interaction, you will be randomly assigned to a new group of three humans and three robots. You will also be reassigned to a potentially new position in your new group. We will also do a new random draw to determine which action pays each person better.

Do you have any questions before we begin? [After they are resolved] You may begin the first part of the experiment.

Instructions for Stage 2 (Period 11)

In the second part of the experiment, you will be making choices for each position, in every scenario which they could potentially face. You will be making the same choices as before, but now the payoffs are multiplied by 10. When action A pays you better, you will receive \$3.00 if you choose A and \$1.00 if you choose B. When action B pays you better, you will receive \$3.00 if you choose B and \$1.00 if you choose A. You will receive an \$8.00 bonus if your choice of action matches the most common action of your group. Also, instead of waiting for the other humans in your group to make their choices, you will be making your choices contingently were you in Position 1, Position 2, and Position 3. You will also

make your choices conditional on whether you were to be paid better by A and B and for the different choice sequences which could precede your choice.

Please click "Continue" to move to the next screen. In the next three screens, we will show you the three screens you will see for this part of the experiment. In the first screen, you will be making choices as though you were in Position 1 for all the cases a position 1 player may face. In the column labeled "Action That Pays You Better," you will see which action pays you better in each case. Notice that there are only two cases for the Position 1 player, one where Action A pays you better, and one where Action B pays you better, because you do not see the choices of the other players prior to your choice. We will ask you to indicate the choice you would prefer to take under each case.

Please click "OK" to move to the next screen. In the second screen, you will be making choices as though you were in Position 2 for all the cases a position 2 player may face. Notice that there are now four cases. The cases differ according to whether you are paid better by A or B and whether the Position 1 player has chosen A or B. On this screen you will be asked to make a choice for each of those four cases regarding what you would prefer to do were you faced with each separate case.

Please click "OK" to move to the next screen. In the third screen, you will be making choices as though you were in Position 3 for all the cases a position 3 player may face. Notice that now there are 8 cases which differ according to whether A or B pays you better and then conditional on that, for the different possible choice configurations of players in positions 1 and 2. Once again you will be asked to make a choice for each of these situations you might find yourself in as a player in Position 3.

Once everyone has made their choices, you will be assigned into a group, assigned a position and will receive a draw to determine which action pays you better. We will then use the choices entered on these screens from each group member to determine the outcome of the interaction. Therefore, on these choice screens it is important that for each case you enter in the choice you would prefer to make were you faced with that specific situation.

Do you have any questions before we begin? [After they are resolved] You may begin the second part of the experiment.

Instructions for Stage 3 (Period 12)

In the third part of the experiment, you will be choosing between two options: Option 1 and Option 2. Just as with the first two parts of the experiment, you will be in a group with two other humans. There are nine rows for which you will be choosing between Option 1 and Option 2. The maximum payoff you can receive for this part of the experiment is $4.00. Option 1 will always pay you the best possible payoff of $4.00 while the other members of your group would generally receive less. Option 2 will always pay each of the other two humans in your group the best possible payoff of $4.00 but will typically pay you less. For Option 1, the payoff for each of the other two humans in your group will start with $4.00 in Row 1, and decrease by $0.50 each row, ending with $0.00 in Row 9. For Option 2, your payoff and the payoff for the other humans in your group does not change from row to row. In each row for Option 2, your payoff is $2.00 and the payoff for each of the other two humans in your group is $4.00. For each of the nine rows, you will tell us whether you would prefer to choose Option 1, or Option 2 to determine the payoffs for you and your other two group members.

Once everyone has submitted their choices, we will randomly assign everyone into groups of three people. We will then randomly pick the choices from one person in each group to determine the payoffs for all three people in that group. For each group, we will randomly draw one of the nine rows. If the selected person from a group picked Option 1 for that group's row, then the three people in that group will be paid according to Option 1. If the selected person from a group picked Option 2 for that group's row, then the three people in that group will be paid according to Option 2.

Do you have any questions before we begin? [After they are resolved] You may begin the third part of the experiment.

REFERENCES

[Achen and Bartels, 2006] Achen, C. H. and Bartels, L. M. (2006). It feels like we're thinking: The rationalizing voter and electoral democracy. In *Annual Meeting of the American Political Science Association, Philadelphia*, volume 30. 1

[Algan and Cahuc, 2010] Algan, Y. and Cahuc, P. (2010). Inherited trust and growth. *American Economic Review*, 100(5):2060–2092. 27

[Anderson and Holt, 1997] Anderson, L. R. and Holt, C. A. (1997). Information cascades in the laboratory. *The American Economic Review*, pages 847–862. 60

[Andreoni and Bernheim, 2009] Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636. 58

[Andreoni and Sanchez, 2020] Andreoni, J. and Sanchez, A. (2020). Fooling myself or fooling observers? avoiding social pressures by manipulating perceptions of deservingness of others. *Economic inquiry*, 58(1):12–33. 31

[Barr and Serra, 2010] Barr, A. and Serra, D. (2010). Corruption and culture: An experimental analysis. *Journal of Public economics*, 94(11-12):862–869. 58

[Bartels, 2002] Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24:117–150. 1

[Bhargava et al., 2017] Bhargava, S., Loewenstein, G., and Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, 132(3):1319–1372. 27

[Bohm, 1972] Bohm, P. (1972). Estimating demand for public goods: An experiment. *European Economic Review*, 3(2):111–130. 26

[Bohnet and Zeckhauser, 2004] Bohnet, I. and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4):467–484. 30

[Bonetto and Arciszewski, 2021] Bonetto, E. and Arciszewski, T. (2021). The creativity of conspiracy theories. *The Journal of Creative Behavior*. 26

[Brotherton and French, 2014] Brotherton, R. and French, C. C. (2014). Belief in conspiracy theories and susceptibility to the conjunction fallacy. *Applied Cognitive Psychology*, 28(2):238–248. 49

[Brotherton et al., 2013] Brotherton, R., French, C. C., and Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in psychology*, 4:279. 28, 29, 30

[Brown et al., 2015] Brown, S., Gray, D., McHardy, J., and Taylor, K. (2015). Employee trust and workplace performance. *Journal of economic behavior & organization*, 116:361–378. 27

[Bullock et al., 2015] Bullock, J. G., Gerber, A. S., Hill, S. J., and Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, 10:519–578. 1, 2, 3, 4, 10

[Bullock and Lenz, 2019] Bullock, J. G. and Lenz, G. (2019). Partisan bias in surveys. *Annual Review of Political Science*, 22:325–342. 1

[Bursztyn et al., 2020] Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, 110(10):2997–3029. 61

[Campbell et al., 2004] Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., and Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of personality assessment*, 83(1):29–45. 32

[Caplan, 2001] Caplan, B. (2001). Rational ignorance versus rational irrationality. *Kyklos*, 54(1):3–26. 60

[Chen and Li, 2009] Chen, Y. and Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1):431–457. 33

[Cichocka et al., 2016] Cichocka, A., Marchlewska, M., de Zavala, A. G., and Olechowski, M. (2016). 'they will not control us': Ingroup positivity and belief in intergroup conspiracies. *British Journal of Psychology*, 107(3):556–576. 26

[Coffin, 2003] Coffin, B. (2003). Breaking the silence on white collar crime. *Risk Management*, 50(9):8–9. 27

[Cox, 2004] Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, 46(2):260–281. 30

[Cummings et al., 1995] Cummings, R. G., Harrison, G. W., and Rutström, E. E. (1995). Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible? *The American Economic Review*, 85(1):260–266. 26

[d'Adda et al., 2020] d'Adda, G., Dufwenberg, M., Passarelli, F., and Tabellini, G. (2020). Social norms with private values: Theory and experiments. *Games and Economic Behavior*, 124:288–304. 58

[Dickinson and McEvoy, 2021] Dickinson, D. L. and McEvoy, D. M. (2021). Further from the truth: The impact of moving from in-person to online settings on dishonest behavior. *Journal of Behavioral and Experimental Economics*, 90:101649. 35

[Douglas and Sutton, 2011] Douglas, K. M. and Sutton, R. M. (2011). Does it take one to know one? endorsement of conspiracy theories is influenced by personal willingness to conspire. *British Journal of Social Psychology*, 50(3):544–552. 32

[Douglas et al., 2019] Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., and Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology*, 40:3–35. 42

[Dutcher and Rodet, 2018] Dutcher, E. G. and Rodet, C. S. (2018). Which two heads are better than one?: Uncovering the positive effects of diversity in creative teams. *Uncovering the Positive Effects of Diversity in Creative Teams (November 12, 2018)*. 38

[Elster, 1989] Elster, J. (1989). Social norms and economic theory. *Journal of economic perspectives*, 3(4):99–117. 61

[Enders et al., 2021] Enders, A. M., Uscinski, J. E., Klofstad, C. A., Seelig, M. I., Wuchty, S., Murthi, M. N., Premaratne, K., and Funchion, J. R. (2021). Do conspiracy beliefs form a belief system? examining the structure and organization of conspiracy beliefs. *Journal of Social and Political Psychology*, 9(1):255–271. 26, 29

[Esarey et al., 2012] Esarey, J., Salmon, T. C., and Barrilleaux, C. (2012). What motivates political preferences? self-interest, ideology, and fairness in a laboratory democracy. *Economic inquiry*, 50(3):604–624. 26

[Fehr and Fischbacher, 2004] Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in cognitive sciences*, 8(4):185–190. 58

[Galliford and Furnham, 2017] Galliford, N. and Furnham, A. (2017). Individual difference factors and beliefs in medical and political conspiracy theories. *Scandinavian Journal of Psychology*, 58(5):422–428. 43

[Goertzel, 1994] Goertzel, T. (1994). Belief in conspiracy theories. *Political Psychology*, pages 731–742. 26, 29

[Gong et al., 2013] Gong, Y., Zhou, J., and Chang, S. (2013). Core knowledge employee creativity and firm performance: The moderating role of riskiness orientation, firm size, and realized absorptive capacity. *Personnel Psychology*, 66(2):443–482. 27

[Graham et al., 2002] Graham, C., Litan, R. E., and Sukhtankar, S. (2002). *The bigger they are, the harder they fall: an estimate of the costs of the crisis in corporate governance.* Brookings Institution Washington, DC. 27

[Harambam and Aupers, 2017] Harambam, J. and Aupers, S. (2017). 'i am not a conspiracy theorist': Relational identifications in the dutch conspiracy milieu. *Cultural Sociology*, 11(1):113–129. 26, 36

[Harvey and Harris, 2010] Harvey, P. and Harris, K. J. (2010). Frustration-based outcomes of entitlement and the influence of supervisor communication. *Human Relations*, 63(11):1639–1660. 27

[Harvey et al., 2014] Harvey, P., Harris, K. J., Gillis, W. E., and Martinko, M. J. (2014). Abusive supervision and the entitled employee. *The Leadership Quarterly*, 25(2):204–217. 27

[Holt and Laury, 2002] Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655. 26

[Holt and Laury, 2005] Holt, C. A. and Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, 95(3):902–912. 26

[Iyengar et al., 2012] Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideologya social identity perspective on polarization. *Public opinion quarterly*, 76(3):405–431. 1

[Iyengar and Westwood, 2015] Iyengar, S. and Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3):690–707. 1

[Jacquemet et al., 2021] Jacquemet, N., James, A. G., Luchini, S., Murphy, J. J., and Shogren, J. F. (2021). Do truth-telling oaths improve honesty in crowd-working? *PloS one*, 16(1):e0244958. 35

[Jolley et al., 2019] Jolley, D., Douglas, K. M., Leite, A. C., and Schrader, T. (2019). Belief in conspiracy theories and intentions to engage in everyday crime. *British Journal of Social Psychology*, 58(3):534–549. 26

[Kates et al., 2022] Kates, J., Tolbert, J., and Rouw, A. (2022). The red/blue divide in covid-19 vaccination rates continues: An update. 2

[Krizan and Herlache, 2018] Krizan, Z. and Herlache, A. D. (2018). The narcissism spectrum model: A synthetic view of narcissistic personality. *Personality and Social Psychology Review*, 22(1):3–31. 26

[Merlo and Schotter, 1999] Merlo, A. and Schotter, A. (1999). A surprise-quiz view of learning in economic experiments. *Games and Economic Behavior*, 28(1):25–54. 72

[Miller et al., 2016] Miller, J. M., Saunders, K. L., and Farhart, C. E. (2016). Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust. *American Journal of Political Science*, 60(4):824–844. 26

[Mullin and Reiley, 2006] Mullin, C. H. and Reiley, D. H. (2006). Recombinant estimation for normal-form games, with applications to auctions and bargaining. *Games and Economic Behavior*, 54(1):159–182. 78

[Neill et al., 1994] Neill, H. R., Cummings, R. G., Ganderton, P. T., Harrison, G. W., and McGuckin, T. (1994). Hypothetical surveys and real economic commitments. *Land Economics*, pages 145–154. 26

[Prior et al., 2015] Prior, M., Sood, G., and Khanna, K. (2015). You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, 10(4):489–518. 1, 2, 3, 4

[Rasmussen, 2021] Rasmussen, D. (2021). The disillusionment of george washington: Partisanship. 1

[Robbett and Matthews, 2018] Robbett, A. and Matthews, P. H. (2018). Partisan bias and expressive voting. *Journal of Public Economics*, 157:107–120. 1, 39

[Sinaiko and Hirth, 2011] Sinaiko, A. D. and Hirth, R. A. (2011). Consumers, health insurance and dominated choices. *Journal of health economics*, 30(2):450–457. 27

[Smallpage et al., 2020] Smallpage, S. M., Drochon, H., Uscinski, J. E., and Klofstad, C. (2020). *Who are the Conspiracy Theorists?: Demographics and conspiracy theories*, pages 263–277. Routledge handbook of conspiracy theories. Routledge. 42

[Torrance, 1966] Torrance, E. P. (1966). Torrance tests of creative thinking. *Educational and Psychological Measurement*. 38

[Wood et al., 2012] Wood, M. J., Douglas, K. M., and Sutton, R. M. (2012). Dead and alive: Beliefs in contradictory conspiracy theories. *Social psychological and personality science*, 3(6):767–773. 29

[Zhou and Hoever, 2014] Zhou, J. and Hoever, I. J. (2014). Research on workplace creativity: A review and redirection. *Annu.Rev.Organ.Psychol.Organ.Behav.*, 1(1):333–359. 27