University of South Carolina

Scholar Commons

Fall 2020

Morals 'Trump' Party: Make America Cooperate Again

Jered Abernathy

Morals 'Trump' Party: Make America Cooperate Again

by

Jered Abernathy

Bachelor of Science
University of Iowa, 2016

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Arts in

Sociology

College of Arts and Sciences

University of South Carolina

2020

Accepted by:

Brent Simpson, Director of Thesis

Matthew Brashears, Committee Member

Laura Brashears, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Abstract

Conflict between Democrats and Republicans is a central component of the contemporary American political system. Negative feelings and discrimination based on political orientation are at an all-time high, leading otherwise similar Americans to deeply distrust one another. Social identity theory provides a framework for not only understanding how this distrust between partisans persists, but how it may be negated. This study builds on recent work on moral judgments and trust games to create situations to increase trust across party lines. Using an online-experimental design, this study investigates the effects of two types of moral judgments on trust building: 1) moral judgments characterized by consensus, where agreement is expected for most Americans (e.g. cheating is bad), and 2) moral judgments characterized by dissensus, where there is a large amount of disagreement across party lines (e.g. abortion rights). This experiment uses a 2 (political group: same or different) x 4 (moral stance: pro-life, pro-choice, universal, or control) design, which allows me to explore how the effects of moral disagreement differ for ingroup members versus outgroup members. Results from this study will be presented and future directions will be discussed for improving political discourse and generalizing interpersonal trust to a group-level.

Table of Contents

List of Tables

Introduction

Political partisanship and animosity in American politics are the highest they have been in decades (Pew 2016). Americans are becoming increasingly polarized in their political views (Pew 2016) and implicit prejudice against the opposing party has grown to higher levels than implicit racial biases (Iyengar and Westwood 2014). This prejudice creates distrust between Americans who hold different political identities, which, in turn, leads to less cooperation between Democrats and Republicans (Carlin & Love 2011; Andris et al. 2015). One explanation for the trust gap is the underlying differences between Democrats and Republicans, including on core demographic dimensions (Taylor 2016), personality traits (Jost, et al. 2003; Pratto, et al. 1994), and moral foundations (Miles and Vaisey 2015). These differences are amplified by the tendency for people to view other groups as more extreme and ideological than their own, a tendency shown in both political and non-political groups (Krislov and Kiley 2016). Below, I develop hypotheses to explain the distrust between partisans and propose a mechanism to promote cooperation between political rivals. I then test the hypotheses in an experimental setting and end with a discussion of my future research agenda.

Chapter 1: Theoretical Background

Social Identity Theory

To explain distrust between members of different political parties, I first turn to a well-documented phenomenon in social psychology: ingroup favoritism (Chen and Li, 2009; Hewstone et al. 2002; Tajfel, 1982). As the name suggests, social identity theory describes the tendency for people to favor others who they believe belong to the same social groups as them. Tajfel (1971) first demonstrated the power of ingroup favoritism using experimentally-created "minimal groups" and went on to develop social identity theory (SIT) to explain his findings (Tajfel and Turner 1979). SIT states that once people identify themselves as belonging to a group, it becomes important to their self-concept and a basis on which they define themselves.

SIT argues that individuals organize their world into social groups via a 3-step process. First, people automatically sort things, including themselves and other people, into categories to more easily understand them. For instance, people subconsciously categorize others by their race, gender, and age group in milliseconds (Contreras, Banaji and Mitchell 2013). The second step, social identification, is where individuals move from categorizing others to categorizing themselves. In this stage, people's social identities become important to their concept of self and they start defining themselves according to their group or category memberships. These social groups can exist naturally outside the lab (e.g. gender, race, religion, political orientation) (Bernhard et al., 2006; Bouckaert and Dhaene, 2004; Fershtman and Gneezy, 2001) or be created in

experimental settings (Eckel and Grossman, 2005; Chen and Li, 2009; Guala et al., 2013). For instance, a person can come to think of themselves as a "mother" or a "vegan.".

The last stage, social comparison, is where members compare their group to other groups. These comparisons are done with the intention of viewing their group as superior, known as ingroup favoritism or ingroup bias (Brewer 1999). Through this process, members teach one another how to act in accordance with the group's values and people begin to internalize the group's norms as their own. Studies consistently find that people prefer their ingroup over the outgroup. This results in individuals giving ingroup members more positive evaluations (Brewer, 1979; LeVine & Campbell, 1972; Mullen, Brown, & Smith, 1992; Perdue et al., 1990), larger rewards (Tajfel et al., 1971), and being more motivated to work with ingroup – versus outgroup -- members (Ellemers, De Gilder, & Haslam, 2004; Worchel et al., 1998). This phenomenon has been shown to be powerful enough to override preexisting friendships in laboratory studies (Ahmed 2007). One key effect of ingroup favoritism is a heightened tendency to extend trust to, and cooperate with, ingroup members (Ahmed, 2007; Brewer, 1999; Dawes, van de Kragt, and Orbell, 1988; Irwin, Mcgrimmon, and Simpson 2008; Simpson 2006; Kramer, Brewer, and Hanna, 1995).

An important question regarding ingroup favoritism is whether the previously mentioned positive effects for members (ingroup love) is accompanied by similar negative effects for non-members (outgroup hate). There is a longstanding debate concerning what conditions lead to outgroup hate vs ingroup love, and the two are often hard to disentangle empirically. While Sumner (1906) argued that the two mechanisms

are directly linked, Allport (1954) later pushed back on the claim. Allport argued that ingroup favoritism need not include degradation of the outgroup. In an attempt to settle the debate, Brewer (1999) looked at forty years of psychological data and found that the link between ingroup love and outgroup hate depends on the structure of the society the groups exist in. Specifically, she points to constant competition and a zero-sum perspective.

The American political system's design is characterized by the conditions Brewer outlines. There is constant competition between Democrats and Republicans in modern-day America with 24/7 political news cycles and the average political campaign length increasing each year (Grosz 2019). The competition between groups also entails *zero sum gains*. This means that one group's success is tied to the failure of the opposing group (Porter and Kramer 2006). For example, the number of Republicans in congress is inversely related to the number of non-Republicans (Democrats and Independents). A few third-party candidates notwithstanding, for one to increase, the other typically must decrease. Therefore, opposing political identities represent competition, leading to outgroup hate and distrust between partisans (Brewer 1999; Brewer and Campbell 1976; Tajfel et al. 1971).

So far, only positive aspects of shared group membership have been discussed. However, being a fellow group member comes with higher expectations and pressures to conform to the group's ideals (Marques et al. 1988; Marques et al. 1994; Gaertner et al. 2011). When a member fails to meet those ideals or deviates from the group's goals, they are subject to harsher punishments than outgroup members (Coull et al. 2001; Khan and

Lambert 2010). The *Black Sheep Hypothesis* (BSH) builds on the assumptions of SIT and explains why this occurs (Marques, Yzerbyt and Leyens 1988).

The 'Black Sheep Hypothesis'

BSH asserts that frequent interactions lead members to have more complex and nuanced views of their own group and, inversely, the lack of contact with outgroup members leads people to exaggerate the outgroup's homogeneity. This is compounded by the fact that people more extensively use cognitive shortcuts such as stereotypes when evaluating outgroup members (Marques and Paez, 1994; Ouwerkerk 2005; Sayans-Jimenez et al. 2017). According to the BSH, agreement or disagreement with ingroup members will matter more than agreement or disagreement with outgroup members. In addition to the different amounts of attention paid to disagreement, people punish their fellow ingroup members more harshly because they have higher expectations for them (Balliet and Van Lange, 2013; Mendoza, Lane, and Amodio, 2014). Therefore, the amplified effects of ingroup disagreement result from the surprise that ingroup disagreement creates. Lastly, deviant ingroup members (or *black sheep)* are seen as a larger threat to the ingroup than are outgroup members. They have more influence over fellow members than outsiders and their existence indicates that the group's values may be weaker than those of the outgroup, who are assumed to be homogeneous in their beliefs (Linville, Salovey, and Fischer, 1986; Meissner & Brigham, 2001; Ostrom and Sedikides, 1992).

The Moral Judgements Hypothesis

The last theory used in this paper is, in a sense, the inverse of the black sheep hypothesis. While the BSH explains how ingroup disagreements can counteract ingroup

love, the moral judgements hypothesis (MJH) explains how outgroup agreements can counteract outgroup hate.

Recently, social scientists have begun looking at morality and its role in producing cooperation between individuals in situations requiring trust and trustworthiness. The focus has been on two main factors: (1) the motivation to view oneself as moral (Mazar, Amir, and Ariely 2008; Aquino and Reed 2002) and (2) the desire to be seen as moral by *others* (Feinberg, Willer, and Schultz 2014; Wu, Balliet, and Van Lange 2015). Simpson, Harrell, and Willer (2013) found that randomly assigning participants to take a stance on a moral issue (e.g. cheating is bad) led them to think of themselves as more moral and to act more trustworthy in economic games compared to a control group. Additionally, observing another take a moral stance led people to view the stance-taker as more trustworthy and prefer them as partners in these games (Everett, Pizarro, and Crockett 2016; Simpson, Harrell, and Willer 2013; Simpson, Willer, and Harrell 2017;).

While research on the moral judgment hypothesis highlights the importance of moral stance-taking, it has thus far focused solely on moral stances that are characterized by a high level of consensus, or agreement about what is right or wrong (e.g. arguments against cheating or plagiarism on college campuses). Polarizing morals, like those that characterize political disagreements (e.g., abortion), have not been investigated. The authors do hypothesize that ingroup status and agreement will moderate the effect of a moral stance when moral consensus is low (e.g. abortion, same-sex marriage, gun rights). This project extends the findings of the MJH by testing the predictions on polarizing moral arguments. Additionally, it examines how group membership, both ingroup love

6

and outgroup hate, interacts with moral stances to promote trust and cooperation between

outgroup members.

Chapter 2: Hypotheses

Derived from SIT, I expect people to have more positive views of ingroup members, all else equal.

Hypothesis 1: When no moral stance is taken, individuals will perceive ingroup members as more moral and trustworthy than outgroup members.

The moral judgements hypothesis (MJH) predicts that agreement will moderate the effect of a moral stance on ratings of trustworthiness.

Hypothesis 2a: Individuals will perceive stance takers they agree with as more moral and trustworthy than those who do not take a moral stance.

Hypothesis 2b: Individuals will perceive stance takers they disagree with as less moral and trustworthy than those who do not take a moral stance.

A conceptual model showing predictions from SIT and MJH is given in Fig. 2.1. The model shows how ingroup bias (following SIT) and shared morals (following the MJH) are predicted to each increase positive interpersonal attitudes, which will in turn increase trust.

While H1 reflects SIT's predictions for ingroup love and outgroup hate, the black sheep hypothesis predicts that moral disagreement will lead to more negative views of ingroup members, which will counteract the effects of ingroup love.

Hypothesis 3: Moral disagreements will have a larger negative effect on perceptions of ingroup members than outgroup members.
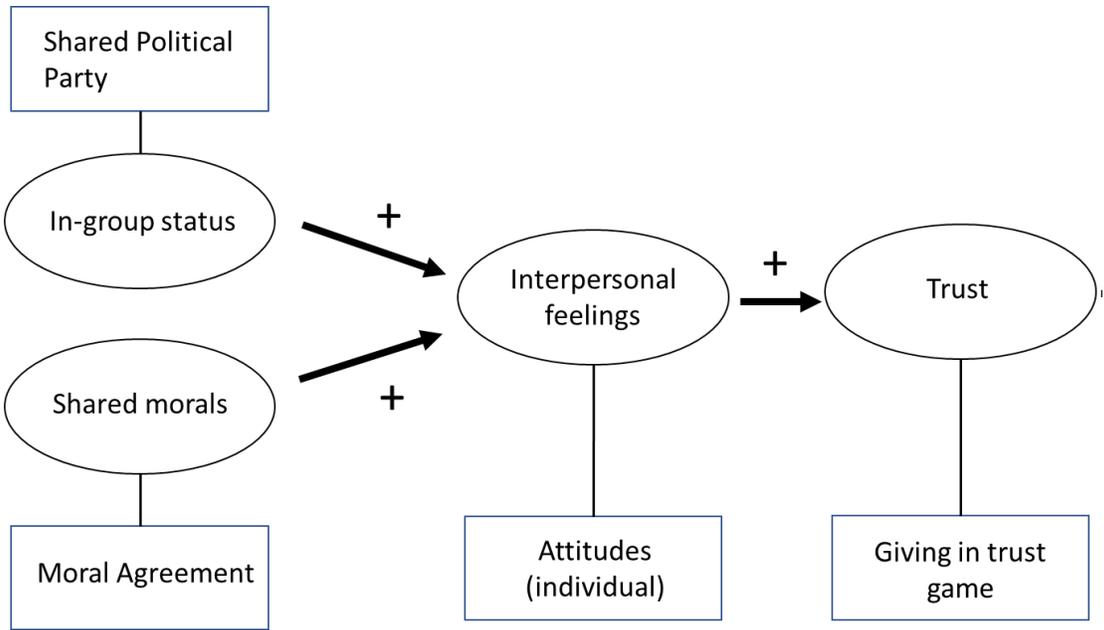
Fig. 2.1: Conceptual Model

It is important to note that H3 refers to the size of the effect and not the overall outcome since ingroup members have higher baseline ratings compared to outgroup members according to H1.

H1 can be thought of as a baseline. When no moral stances are being taken, in this study the control conditions, H1 predicts ingroup favoritism to occur. H3 predicts that the size of the moral disagreement effect will be larger for ingroup members. Lastly, H2 is split into two separate hypotheses to compare the effects that moral agreement (a) versus disagreement (b) have on feelings towards the stance-taker.

Chapter 3: Methods

Recruitment and sample information

Participants were recruited through Prolific Academic, an online database of adult workers who complete online tasks on a wide variety of topics (Peer et al. 2017). The sample consisted of 48% females, 53% Republicans, with a mean age of 35. See Table 3.1 for a more detailed description of the sample demographics.

Procedures

After opening the study on Prolific's task page, participants were redirected to a Qualtrics survey. Participants were told that the purpose of the study was to make quick judgements about others based on short paragraphs and limited information. After agreeing to participate, participants provided basic demographic information about themselves including gender, age, race, education, and political orientation.

Next, participants were asked to write a short paragraph about their views on one of four topics, depending on condition. The study uses a 2 (political group: same or different) x 4 (morals: pro-life, pro-choice, universal, or control) design. After writing their own paragraph, participants read a paragraph ostensibly written by a (fictitious) person with whom they were paired. Participants either read a pro-life argument (conservative morals), a pro-choice argument (liberal morals), an anti-cheating argument (moral universal), or a restaurant review (control, no moral). The writing prompts always matched the reading prompts, so participants read a paragraph responding to the same

Table 3.1: Demographics

Race
White	332 (77.39%)	
Black	18 (4.20%)	
Latinx	19 (4.43%)	
Asian	22 (5.13%)	
Other	38 (8.86%)	

Gender
Male	222 (51.99%)
Female	205 (48.01%)

Age
Min	18
Max	73

Education
<HS	6	(1.4%)
High School	58	(13.5%)
Some College	101	(23.5%)
2-year Degree	33	(7.7%)
4-year Degree	154	(35.9%)
M.A.	59	(13.8%)
PhD.	18	(4.20%)

Region
Northeast	98	(22.8%)
Midwest	95	(22.1%)
South	166	(38.7%)
West	70	(16.3%)

Income
<$15k	40	(6.3%)
$15k-$30k	100	(15.8%)
$30k-45k	98	(15.5%)
$45k-$60k	104	(16.5%)
$60k-$75k	80	(12.7%)
$75k-$90k	53	(8.4%)
$90k-$105k	40	(6.3%)
$105k-$120k	37	(5.9%)
$120k+	80	(12.7%)

prompt they had just written about. Participants were also told their partner's political orientation (Democrat or Republican) and region of residence.

The region of residence served as an attention check to ensure participants read the passages carefully and remembered their partner's demographics. Additionally, it was included to ensure that political orientation was not the sole piece of information given. This made it more difficult for participants to discover the purpose of the study. Lastly, the partner's region was set as Midwest due to its political diversity. This avoided unintended interactions between the region and political orientation that may have occurred with examples such as "a liberal from the South" or "a conservative from the West Coast."

After reading the prompt, participants were asked two manipulation check questions ("What was the paragraph you just read about?" and "Was the author a Democrat or a Republican?") as well as a comprehension check ("What region was the author from?") Comprehension rates were high, with 97% of participants answering all the checks correctly.

After the reading and writing portion of the study, participants took part in a token allocation task, specifically a one-shot trust game. Then, they answered questions about their feelings towards their own political group and the opposing political group to gauge attitudes towards the groups as a whole. Afterwards, participants were debriefed about the true purpose of the study and paid. The measures are described in further detail below.

Measures

Agreement and disagreement with a moral stance

After reading the prompt, participants answered a 7-point Likert scale in response to the question, "How much do you agree with the other participant's opinion?" The scale ranged from strongly disagree to strongly agree. The above scale was split into two separate measures, one for agree and the other for disagree. This split was necessary in order to compare the effects of agreement and disagreement on attitudinal and behavioral trust.

This study measured agreement on two types of moral stances: those with high levels of consensus and those with low levels of consensus. Following Simpson et al. (2013), the high consensus condition used an argument against cheating on college exams. The low consensus condition contained one of two arguments either in support or against abortion. To ensure realism, I visited multiple online forums and editorials to see the talking points and argument styles of each side. I then wrote a prompt for each side that reflected the reoccurring themes. The prompts were pilot tested to ensure they sounded authentic and no participants mentioned the prompt's writing style as reasons for suspicion.

Attitudes

Attitudinal trust is modeled as an index of 6 indicators: liking, trustworthy, responsible, moral, selfish, fairness. Except for liking, the question text was the same for all the measures. Participants were asked *"How ____ do you think the other participant is, in general?"* and each of the measures were paired with a 1-7 Likert scale For liking, participants were asked "How much do you like the other participant?" The 6 indicators

had a very high interitem reliability (0.94). The models used were also run using the indicators individually, rather than as a scale and the main findings remained unchanged.

Behavioral Trust

My measure of trust used a standard 2-person trust dilemma, also called an investment game (Kuwabara 2011; Berg, Dickhaut and McCabe 1995; Buchan, Croson and Dawes 2002). In this decision scenario, participants were assigned the role of Sender and given 10 tokens, each worth $0.03. As the Sender they could send 0-10 tokens to the other participant who was the Receiver. Any tokens sent by the Sender were *tripled* before given to the Receiver. Any tokens not sent were added to their point total at the end of the round. The Receiver then chose how many of the tokens to return to the Sender and the round is over. The number of tokens participants chose to send to the Receiver is the behavioral measure of trust in this study. Participants were told they would learn how many tokens were returned to them at the end of the study. I did not measure partner's trustworthiness, as it was outside the scope of this research.

Political Orientation

I measured political orientation in three ways. First, participants reported how they think of themselves politically (Democrat, Republican, Independent, or Other). Next, using 7-point Likert scales, participants described how liberal/conservative they view themselves both socially and economically. Lastly, I measured who participants voted for in the 2016 presidential election and how enthusiastic they were about their candidate of choice. These additional measures of political orientation allow for models to investigate how political extremism acts as a moderator on the impact of ingroup bias on attitudes (discussed more below).

Additional Variables of Interest

   I expect the strength of one's party affiliation to moderate the effect of in-group bias on trust and cooperation. Specifically, the more extreme liberals or conservatives should show the most ingroup love and outgroup hate. Since this study includes two measures of trust, behavioral and attitudinal, I model how attitudes mediate the effect of group status and moral agreement on behavioral trust. I expect ingroup status and moral agreement will increase attitudinal trust, which will lead to more behavioral trust.

Chapter 4: Results
Level of consensus

   Before testing the hypotheses, it is important to see if the level of consensus

around an argument alters the effect of agreement on either attitudinal or behavioral trust.

The first column of Table 4.1 and Table 4.2 labelled 'Consensus Check' show no

difference on the level of consensus for either attitudinal or behavioral trust. The non-

significant consensus coefficient shows that agreement with a moral stance has the same

positive effect on trustworthiness ratings regardless of whether the stance taken is an

abortion argument or an anti-cheating argument. Again, the level of consensus of the

moral stance has no direct or indirect effect on behavioral trust. Based on these results,

the following analyses look at agreement regardless of whether it was in the low or high

consensus conditions.

Hypothesis-testing for attitudes

   I use multivariate regression to test my hypotheses for two reasons; first, it allows

me to analyze how my continuous independent variables function along a spectrum,

rather than only using experimental conditions as indicators. For example, I analyze H2a

and H2b to see how *the degree of agreement or disagreement* affects attitude ratings of

stance takers (e.g. the more a person agrees with a stance, the higher they rate the stance

taker). This has the added benefit of accounting for partisans with more moderate beliefs.

Under the regression framework, a person who only moderately agrees with a moral

stance is predicted to rate the stance taker less trustworthy and moral than a participant

who strongly agrees with the moral stance. This linear effect would be categorized as

Table 4.1:Attitude Model

	Consensus Check	Ingroup bias	Black Sheep Hypothesis	MJH Agree	MJH Disagree
Same Party	0.039	0.812***	0.151	0.100	0.077
	(0.12)	(0.30)	(0.19)	(0.16)	(0.17)
Female	0.085	0.035	0.186	0.086	0.193
	(0.11)	(0.26)	(0.12)	(0.11)	(0.12)
Nonwhite	-0.150	-0.186	-0.163	-0.155	-0.173
	(0.13)	(0.31)	(0.15)	(0.14)	(0.15)
Republican	0.234	0.287	0.207	0.309	0.212
	(0.12)	(0.32)	(0.21)	(0.17)	(0.18)
Disagree			-0.800***		-0.850***
			(0.08)		(0.05)
Agree	0.826***			0.824***	
	(0.05)			(0.04)	
Same Party * Rep			-0.022	-0.136	-0.015
			(0.25)	(0.23)	(0.25)
Same Party * Disagree			-0.100		
			(0.11)		
High Consensus (0/1)	0.041				
	(0.12)				
Constant	3.53***	4.55***	5.46***	3.54***	5.49***
	(0.17)	(0.36)	(0.14)	(0.15)	(0.14)

*p<0.05 **p<0.01 ***p<0.001

Table 4.2: Behavior Model

	Consensus Check	Ingroup Bias	BSH	MJH
Same Party	0.628	-0.417	0.77	0.627
	(0.43)	(1.0)	(0.50)	(0.70)
Female	-1.08*	-0.894	-1.0**	-2.43**
	(0.42)	(.70)	(0.43)	(0.70)
Nonwhite	-0.114	-1.04	0.03	0.35
	(0.51)	(0.83)	(0.51)	(0.50)
Republican	-0.852	-1.9	-0.848	-0.84
	(0.44)	(1.0)	(0.44)	(0.44)
Disagree			-0.424	
			(0.26)	
Agree	0.643***			0.219
	(0.17)			(0.28)
Same Party * Rep		-0.8	-0.022	
		(1.6)	(0.25)	
Same Party * Disagree			-0.1	
			(0.11)	
Same Party * Agree				0.011
				(0.33)
Female * Agree				0.792*
				(0.33)
High Consensus (0/1)	0.061			
	(0.46)			
Constant	6.21***	8.95***	7.63***	6.97***
	(0.64)	(1.1)	(0.47)	(0.63)

standard error in an ANOVA framework. The second reason I use multivariate regression is that it allows me to control possible confounding variables.

Hypothesis 1 states that ingroup bias will occur when no moral argument is made. To test this hypothesis, I used a dummy variable for whether the participant and their partner were in the same political party. Participants' gender, race, and political party were included as controls. The results are listed in the Table 4.1's second column labelled 'ingroup bias.' The coefficient for political party and its interaction with ingroup status were not significant (not shown), indicating that ingroup-bias for both Democrats and Republicans was measured at similar levels. Ingroup status was the sole significant (p<.01) predictor of attitudes in the model. The positive and significant coefficient supports hypothesis 1's claims about ingroup biased attitudes in the absence of a moral argument.

Hypothesis 2a states that agreement on a moral stance will increase ratings of the stance-taker. I use the Likert scale of agreement (0-3) to assess how level of agreement impacts attitudinal ratings of stance takers. Consistent with MJH, Table 4.1's MJH Agree column shows the linear effect of moral agreement is highly significant (p<0.001) and positive. Specifically, the coefficient shows that for each 1 unit increase in agreement, the attitude ratings of the stance taker increase by 0.824. Similarly, Table 4.1's MJH disagree shows disagreement is highly significant (p<0.001) and negative. For each 1 unit increase in disagreement, the attitude ratings of the stance taker decrease by 0.85.

In both models, ingroup bias is non-significant. When moral stances are introduced, ingroup bias no longer influences attitudes. Therefore, ingroup status does not affect ratings of a partner's trustworthiness *outside of the control conditions*. Put another

way, when a participant read a moral argument from their partner, the only predictor of their attitudes towards that partner was how much they agreed with the stance. When a moral stance was taken, politics were not related to the ratings participants gave their partners. These findings show support for both of MJH's claims.

Next, Hypothesis 3 predicts that moral disagreement with ingroup members will have a greater decrease in attitudes when compared to moral disagreement with outgroup members. The BSH column of Table 4.1 shows that moral disagreement has a large, highly significant ($p<.001$), negative effect on attitudes. Importantly for hypothesis 3, the interaction between moral disagreement and ingroup is non-significant ($p=0.33$). Therefore, I find that moral disagreement did not affect ratings more (or less) for ingroup members than outgroup members. This does not reflect the predictions of the black sheep hypothesis, so hypothesis 3 is not supported.

Hypothesis-testing for behavior

So far, the outcome variable has been attitudes: how trustworthy, responsible, moral, non-selfish, fair, and likeable participants rate their partners. The next step is to change the focus from attitudes to behavioral trust as measured by how many tokens participants gave to their partner in a one-shot trust game. The first question is, like above, is there evidence for ingroup bias in behavioral trust? Table 4.2 shows the models with behavior as the outcome variable. Looking at the second column labelled 'ingroup bias' shows that ingroup status and all the controls fail to reach significance at the $p<0.05$ level. This stands in contrast to the findings on attitudes reported above. Therefore, even though participants report more positive attitudes towards ingroup members when no moral stance was taken (control condition), they do not entrust more tokens to their own party members.

21

Next, are ingroup members punished more harshly for moral disagreements via decreased giving in the trust game? The 'BSH' column shows that both the main effect of disagreement and its interaction with ingroup status are non-significant predictors of giving behavior. A significant negative effect of female gender was found ($p<0.05$) suggesting that females trusted their partners less than their male counterparts, a finding consistent with the literature on gender and trust (Buchan et al. 2008; Simpson and van Vugt 2009). Additionally, Republican status is marginally significant and negative ($p=0.057$) suggesting less trusting behavior by Republican participants. It is important to note that these effects were not present in the control condition, suggesting that these variables are only relevant to the decision of how many tokens to entrust a partner when that partner has taken a moral stance.

Lastly, looking at agreement's effect on giving in 'MJH' column shows a non-significant main effect, but a significant interaction with gender. Female gender also has a highly significant negative effect on giving in the model. In combination, these coefficients show that females give less than males when they disagree with their partners but the gap in giving between them decreases and ultimately disappears once they fully agree with their partner. To illustrate this effect, Fig. 4.1 graphs the relationship between gender and agreement on number of tokens given to their partner.

In whole, the behavioral data appears to tell a different story from the attitude results reported above. Participants don't trust others with their tokens differently based on their agreement or disagreement with their argument. Additionally, their ingroup status also failed to predict giving behavior. Overall, partner characteristics failed to
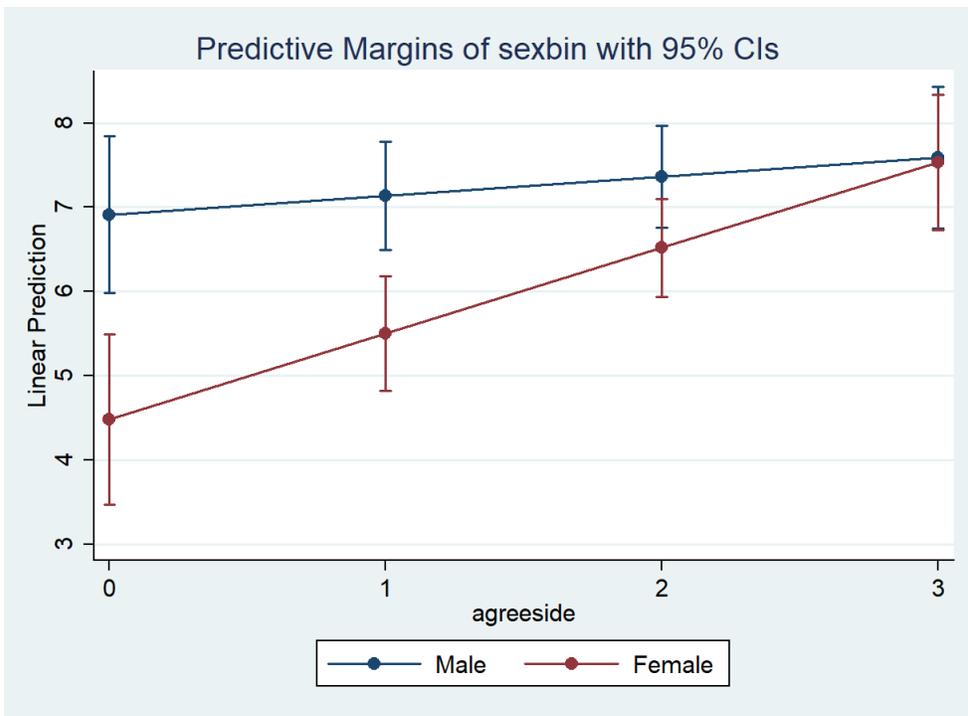
22

Fig. 4.1: Giving by gender

predict giving behavior, even though they succeeded at predicting attitudes towards the partner, which included a measure of trustworthiness.

Moderation analysis

I now return to the measure of political extremism that was discussed above. I measured political extremism by calculating the average distance from the midpoint each participant was on both social and economic matters. The measure ranged from 0 to 4, where 0 equals complete moderates, and 4 represents the most partisan individuals. This measure of extremism does not separate extreme liberals from conservatives, though they are separated in the models below using political party as a dummy variable. I hypothesized that extremism would moderate, and specifically amplify, the effects of ingroup bias.

Extremism and attitudes

First, does the level of political extremism amplify ingroup bias for attitudes? Table 4.3, column 1 shows the effect of introducing extremism into the model for control conditions. As a reminder, these participants did not make or see a moral stance, instead they read and wrote a short restaurant review. In these conditions, all the variables in the attitudinal model fail to reach significance. This non-significance is surprisingly considering the significant effect that ingroup status had on attitudes without including extremism in the model. Nevertheless, extremism did not moderate ingroup bias's effect on attitudes in the control condition.

The control conditions do not support extremism's status as a moderator of ingroup bias, but now I look at its effect when a moral stance is being taken. Table 4.3's second column shows that extremism has a significant negative effect on attitudes and interacts with both agreement and ingroup status. This is the first evidence we have that extremism

Table 4.3: Extremism and Attitudes

	Ingroup Bias	Agree	Disagree
Same Party (0/1)	0.859	-0.483	-0.142
	(0.54)	(0.26)	(0.28)
Female (0/1)	-0.015	0.107	0.181
	(.26)	(0.12)	(0.13)
Nonwhite (0/1)	-0.232	-0.177	-0.163
	(0.83)	(0.14)	(0.15)
Republican (0/1)	0.273	0.238[*]	0.199
	(0.32)	(0.12)	(0.13)
Extremism (0/4)	-0.17	-0.324[***]	-0.035
	(0.13)	(0.09)	(0.08)
Agree		0.623[***]	
		(0.11)	
Disagree			-0.836[***]
			(0.15)
Same Party * Extremism	-0.017	0.201[*]	0.088
	(0.18)	(0.09)	(0.10)
Same Party * Agree		0.074[*]	
		(0.04)	
Same Party * Disagree			-0.006
			(0.05)
Constant	5.00[***]	4.44[***]	6.97[***]
	(0.48)	(0.28)	(0.63)

*p<0.05 **p<0.01 ***p<0.001

is acting as a moderator, and it is significantly amplifying the effects of two variables of interest. The first interaction shows that political extremism increases the effect of ingroup bias on ratings when a moral argument is made. The second interaction shows that extremism amplifies the effect that agreement on a moral stance has on ratings of a partner. In other words, as extremist beliefs increase the effect of moral agreement on attitudes also increases.

With promising results from the moral stance agreement conditions, we should expect extremism to have similar if not greater impact when participants *disagree* with the moral stance. However, extremism has no effects, direct or indirect, in the disagreement model. Table 4.3's third column shows that the disagreement model has no significant effects other than the main effect of disagreement on attitudes.

The results change when the outcome variable is giving behavior. Once again, we turn to the control condition. Column 1 of Table 4.4 shows that extremism has a significant (p=0.05) negative direct effect on giving, but extremism did not moderate ingroup bias in the control condition for giving or attitudes. When no moral stance was taken, more extreme partisans sent fewer tokens to their partners regardless of their partner's political status. Ingroup status also *decreased* giving to politically similar others, which was an unexpected finding. Lastly, Republicans gave fewer tokens, though this did not interact with extremism.

Moving to the moral stance conditions, we see similar findings that I discussed above when the outcome variable was attitudinal. First, when participants agreed with a moral stance extremism decreased giving, as it did in the control condition. Extremism increased the impact of agreement on behavioral trust. This mirrors the effects found in

26

Table 4.4: Extremism and Giving

	Ingroup Bias	Agree	Disagree
Same Party (0/1)	-2.86*	0.182	0.241
	(1.42)	(0.95)	(0.97)
Female (0/1)	-0.88	-1.01*	-0.938*
	(.69)	(0.43)	(0.43)
Nonwhite (0/1)	-0.998	-0.02	-0.010
	(0.82)	(0.50)	(0.51)
Republican (0/1)	-2.25**	-0.788	-0.82
	(0.85)	(0.44)	(0.45)
Extremism (0/4)	-0.65	-0.767*	-0.090
	(0.33)	(0.34)	(0.28)
Agree		-0.21	
		(0.39)	
Disagree			0.064
			(0.53)
Same Party * Extremism	0.85	0.17	00135
	(0.48)	(0.35)	(0.35)
Agree * Extremism		0.31*	
		(0.14)	
Disagree * Extremism			-0.184
			(0.17)
Constant	10.8***	8.40***	7.91***
	(1.26)	(1.05)	(0.85)

*p<0.05 **p<0.01 ***p<0.001*

the attitudinal data of Table 4.3. Next, when we look at moral disagreement, we see that many variables of interest lose significance. Disagreement itself has no effect on giving behavior as well as extremism and political party of self and other. Lastly, in the moral stance conditions, females gave significantly fewer tokens to their partners than their male counterparts.

Overall, we see that in both the attitudinal and behavioral models, extremism only acted as a moderator when there was moral agreement. The behavior model shows that political extremism leads participants to give less as a baseline but amplifies the effect of agreement on tokens given.

Chapter 5: Discussion

This research presents key findings that help illustrate the underlying processes behind the lack of trust in and cooperation between Democrats and Republicans. It analyzes potential factors behind the large political trust gap in modern-day America using an online experimental setting. Specifically, this paper examines how group membership and moral agreement interact to affect perceptions of morality and trustworthiness. This research tested predictions derived from social identity theory, the black sheep hypothesis, and the moral judgements hypothesis and illustrated the conditions under which shared morals can outweigh 'outgroup hate' and when moral disagreement can outweigh 'ingroup love.'

Results from measures of attitudinal trust supported predictions from SIT on ingroup bias. In the control condition, where no moral stances were taken, both Democrats and Republicans rated members of their own party as higher on a scale consisting of liking, trustworthiness, responsible, moral, unselfish, and fairness. This ingroup bias effect was no longer present when a moral stance was taken. Additionally, the absence of ingroup biased attitudes when moral stances are taken supports recent findings that offering additional information about an interaction partner can outweigh parochialism (Hernandez-Lagos & Minor, 2015; Vermue, Meleady, & Seger, 2019).

These findings also supported MJH's predictions that agreement or disagreement on a moral stance leads to increased ratings of trust and morality. Agreement or disagreement with the moral stance were the only significant predictors of feelings

towards the stance-taker. This is consistent with prior findings that moral disagreement is the main mechanism behind partisan prejudice (Viciana et al. 2019). This effect was consistent regardless of whether the argument was characterized by high consensus (cheating is bad) or low consensus (prolife or prochoice arguments).

No support was found for the black sheep hypothesis. Results did not show evidence of increased negative attitudes or a decrease in trusting behavior when participants disagreed with an ingroup member when compared to outgroup disagreement. Instead, when a moral stance was taken ingroup status had no significant effects, neither direct nor indirectly through agreement, on attitudes. Therefore, agreement or disagreement's effect on view of stance-takers did not differ when the stance-taker was an ingroup or an outgroup member. This supports the argument above that when a moral stance is taken, agreement or disagreement with the stance is the sole determinant of attitudes towards the stance-taker and other factors no longer matter.

While the attitude results listed above are mainly consistent with predictions, the behavioral data were largely non-significant. In this study, the giving behavior did not support any of the hypotheses and the only significant predictors of giving behavior found were participant gender, political orientation, and political extremism. While there are mixed findings on political identity's role in trust and reciprocity via economic games (Hernandez-Lagos & Minor, 2015; Anderson, Mellor, & Milyo 2004; Wu et al. 2018) most published studies find evidence for ingroup favoritism (Balliet et al. 2018; Fowler & Kam, 2007; Van Lange et al. 2012; Carlin & Love, 2016; Iyengar & Westwood, 2014; Carlin & Love, 2013). This suggests that giving behavior is determined by a different mechanism than the ones outlined in this paper that accurately predicted attitudes,

including ratings of another's trustworthiness. Ultimately, the only significant predictors of giving behavior were participant-level variables. In summary, the choice of how much to trust another was based more on the one doing the trusting, than the recipient of that trust.

Chapter 6: Future Directions and Concluding Remarks

This project is the first in a research agenda focused on discovering the conditions where outgroup cooperation is possible. Using this project as a foundation, future projects could look at how repeated interactions with an outgroup member can build trust and overcome the partisan trust gap seen in the attitude measures. Additionally, future projects will focus on extending the current research to test if trust can be established in one-shot experiments and translated to future economic games with other outgroup members. This will answer the question, does the trust and liking established from agreeing with a moral stance generalize to the outgroup as a whole? Relatedly, does agreement with outgroup members cause participants to believe that the outgroup is more similar to their ingroup as a whole?

In modern-day America, we see an increasing gap between Democrats and Republicans. I find that politically opposed people can overcome partisan-biased attitudes towards each other by emphasizing their moral similarities rather than their differences. However, I also find that political extremism *increases the trust gap* by amplifying the effects of ingroup bias and decreasing behavioral trust across all domains. Americans are becoming increasingly ideological, so this finding should serve as a cautionary look into a potential future where citizens are even more distrusting of one another than they are currently. A beneficial next step would be to create more instances where people relate with politically different others over their similarities. This will expand the ingroup to be inclusive of all Americans, rather than only half of them.

References

Ahmed, Ali. 2007. "Group identity, social distance and intergroup bias." *Journal of Economic Psychology* 28 (3): 324-337.

Anderson, Lisa R., Jennifer M. Mellor, and Jeffrey Milyo. 2005. "Do liberals play nice? The effects of party and political ideology in public goods and trust games." Experimental and Behavioral Economics. Emerald Group Publishing Limited, 107-131.

Andris, Clio, David Lee, Marcus J. Hamilton, Mauro Martino, Christian E. Gunning, and John Armistead Selden. 2015. "The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives." *PloS ONE* 10 (4).

Brewer, Marilynn. 1999. "The Psychology of Prejudice:Ingroup Love or Outgroup Hate?" *Journal of Social Issues,* 55 (3): 429-444.

Carlin, Ryan E., and Gregory J. Love. 2013. "The politics of interpersonal trust and reciprocity: An experimental approach." *Political Behavior* 35.(1): 43-63.

Contreras, Juan, Mahzarin Banaji, and Jason Mitchell. 2013. "Multivoxel Patterns in Fusiform Face Area Differentiate Faces by Sex and Race." *PLoS ONE* 8 (7).

Eriksson, Kimmo, and Pontus Strimling. 2015. "Group differences in broadness of values may drive dynamics of public opinion on moral issues." *Mathematical Social Sciences* 77: 1-8.

Feinberg, Matthew, and Robb Willer. 2015. "From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence?" *Personality and Social Psychology Bulletin* 1-17.

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean Wojcik, and Peter Ditto. 2012. "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." *Advances in Experimental Social Psychology* 1-64.

Hernandez-Lagos, Pablo, and Dylan Minor. 2015. "Political Identity and Trust." Harvard Business School Strategy Unit Working Paper 16-012

Hunter, James. 1992. *Culture Wars: The Struggle to Define America.* New York: Basic Books.

Iyengar, Shanto, and Sean Westwood. 2014. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690-707.

Jost, John, Jack Glaser, Arie Kruglanski, and Frank Sulloway. 2003. "Political conservatism as motivated social cognition." *Psychological Bulletin* 129 (3): 339-375.

Koleva, Spassena, Jesse Graham, Ravi Iyer, Peter Ditto, and Jonathan Haidt. 2012. "Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes." *Journal of Research in Personality* 46: 184-194.

Krislov, Zachary, and Jocelyn Kiley. 2016. "Partisans see opposing party as more ideological than their own." *Pew Research Center.*

Kuwabara, Ko. 2011, "Cohesion, Cooperation, and the Value of Doing Things Together: How Economic Exchange Creates Relational Bonds." *American Sociological Review* 76(4): 560–580

Miles, Andrew, and Stephen Vaisey. 2015. "Morality and politics: Comparing alternate theories." *Social Science Research* 53: 252-269.

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research." *Journal of Experimental Social Psychology* 70:153-163.

Pew Research Center. 2016. *Partisanship and Political Animosity in 2016.*

Pew Research Center. 2019. *Trust and Distrust in America.*

Pratto, Felicia, Jim Sidanius, Lisa Stallworth, and Bertram Malle. 1994. "Social dominance orientation: A personality variable predicting social and political attitudes." *Journal of Personality and Social Psychology* 67 (4): 741-763.

Simpson, Brent, Ashley Harrell, and Robb Willer. 2013. "Hidden Paths from Morality to Cooperation: Moral Judgments Promote Trust and Trustworthiness." *Social Forces* 91 (4): 1529-1548.

Tajfel, Henri, and Jonathan Turner. 1979. "An Integrative Theory of Intergroup Conflict."
Chap. 3 in *The social psychology of intergroup relations*, 33-47.

Taylor, Paul. 2016. "The demographic trends shaping American politics in 2016 and
beyond." *Pew Research Center.*

Vermue, Marieke, Rose Meleady, and Charles R. Seger. 2019. "Member-to-member
generalisation in trust behaviour: How do prior experiences inform prosocial
behaviour towards novel ingroup and outgroup members?." *Current Psychology*
1-18.

Viciana, Hugo, Ivar Hannikainen, and Antonio Torres 2019. "The dual nature of partisan
prejudice: Morality and identity in a multiparty system" *PLoS ONE* 14(7)