ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 029

Hate Trumps Love: The Impact of Political Polarization on Social Preferences

Eugen Dimant

September 2020

www.econtribute.de

UNIVERSITÄT BONN

Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged.

Hate Trumps Love: The Impact of Political Polarization on Social Preferences

Eugen Dimant[a,b,c]

[a] *University of Pennsylvania*
[b] *CESifo, Munich*
[c] *Identity and Conflict Lab*

Abstract

Political polarization has ruptured the fabric of U.S. society. The focus of this paper is to examine various layers of (non-)strategic decision-making that are plausibly affected by political polarization through the lens of one's feelings of *hate* and *love* for Donald J. Trump. In several pre-registered experiments, I document the behavioral-, belief-, and norm-based mechanisms through which perceptions of interpersonal closeness, altruism, and cooperativeness are affected by polarization, both within and between political factions. To separate ingroup-love from outgroup-hate, the political setting is contrasted with a minimal group setting. I find strong heterogeneous effects: ingroup-love occurs in the *perceptional* domain (how close one feels towards others), whereas outgroup-hate occurs in the *behavioral* domain (how one helps/harms/cooperates with others). In addition, the pernicious outcomes of partisan identity also comport with the elicited social norms. Noteworthy, the rich experimental setting also allows me to examine the drivers of these behaviors, suggesting that the observed partisan rift might be not as forlorn as previously suggested: in the contexts studied here, the adverse behavioral impact of the resulting intergroup conflict can be attributed to one's grim expectations about the cooperativeness of the opposing faction, as opposed to one's *actual* unwillingness to cooperate with them.

Keywords: Identity, Norms, Political Polarization, Social Preferences, Trump

JEL: B41, D01, D9

[*] I am indebted to Simon Gächter, Daniele Nosenzo, and Silvia Sonderegger for their input on the initial project idea. This work has also benefited from conversations with Abraham Aldama, Danny Choi, Guy Grossman, Felix Kölle, Jeff Lees, Matthew Levendusky, Brendan Nyhan, Kyle Peyton, Nicholas Sambanis, and Sean Westwood. The three-part data collection and analyses were pre-registered at AsPredicted.org #42538, #42539, and #42540 entitled *'Hate Trumps Love'*. I acknowledge financial support from the German Research Foundation (DFG) under Germany's Excellence Strategy — EXC 2126/1– 390838866.

Email address: edimant@sas.upenn.edu (Eugen Dimant)

This version: September 8, 2020

1. Introduction

Rising political polarization coincides with and is often linked to fractured societies rife with racial inequality, factional conflict, and partisan animosity (Dixit and Weibull, 2007; Fiorina and Abrams, 2008; Iyengar and Westwood, 2015; Bénabou and Tirole, 2016; Reich, 2017; Autor et al., 2020; Graham and Svolik, 2020). At its core, polarization undermines social contracts that are necessary for a functioning society: it restrains social interactions across polarized clusters, impedes cooperativeness, trust, and altruism between political factions, and thus poses a credible threat to democratic values.[1] This is further amplified by *false polarization*, the perception of more polarization with respect to policy issues than actually exists (Levendusky and Malhotra, 2016; Flynn et al., 2017; Moore-Berg et al., 2020). In both strategic and non-strategic decision contexts that capture cooperativeness, altruism, and anti-social behavior, I examine the behavioral-, belief-, and norm-based mechanisms with which this political intergroup conflict materializes.

Understanding how political polarization may affect one's willingness to engage in altruistic behavior and the collective provision of goods within and between factions has direct social welfare implications (Henrich et al., 2001; Fehr and Fischbacher, 2003; Bowles and Gintis, 2013). Arguably, not all consequences of political polarization are created equal and can be tackled with the same policies, thus warranting the trifecta approach – as put forward in this study – of measuring beliefs, behaviors, and attitudes separately.

On a theoretical level, the aggregate social consequences of polarization are unclear. For one, polarization (and the resulting hostile climate) can produce enough outgroup animosity to reduce individual willingness to support and cooperate with members of the opposite faction. For another, it may also – or instead – increase intra-faction cooperation; for instance, by promoting a sense of shared identity. To assess the social impact of polarization, it is thus important to compare within- and between-group behaviors in a polarized environment that enables both strategic and non-strategic considerations. I attempt to quantify this phenomenon through the lens of multiple controlled online experiments.

Here, I focus on two layers of decision-making within and across political factions. The first is behavioral: how does polarization in the context of political identities affect

[1] Recent examples even include the opaque partisan divide over wearing face masks as a preventative measure against COVID-19 infections: https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html. See also Bhanot and Hopkins (2020).

pro-/anti-social decisions, cooperation, and social expectations in both strategic and non-strategic environments? Do these take shape in form of ingroup-love, outgroup-hate, or both simultaneously? Humans are known to bond over common identity markers (often exemplified by one's preference *for* something or someone), the study of which has its origins in social science in the form of ingroup-outgroup favoritism and social identity considerations (Tajfel and Turner, 1979; Alesina et al., 1999; Akerlof and Kranton, 2000; Bernhard et al., 2006; Efferson et al., 2008; Halevy et al., 2008). Although partisanship is found to exacerbate cross-faction discrimination, prior work has often attributed this behavior to a mix of ingroup-love and outgroup-hate by pitting one group against another.[2]

The second is perceptional: are these behavioral differences consistent with the observed variations in perceived interpersonal closeness and social norms and thus help to explain *why* we observe these differences within and between political factions?

My investigative approach is consistent with and speaks to the growing discussion on affective polarization – the animosity between and distrust towards members of the opposing faction (Druckman and Levendusky, 2019; Iyengar et al., 2019). Donald Trump is a polarizing figure and the current symbol of the Republican party (Jacobson, 2019), which is captured by the novel measures that I put forward here. By comparing settings with political identities to those with minimal group identities, I also contribute to the aforementioned *groupy* behavior (Kranton and Sanders, 2017) literature in that the impact of partisan animosity on ingroup-love and outgroup-hate can be examined separately.

Across these three settings (non-strategic decisions, strategic decisions, norm perceptions), I collect and analyze experimental data from a diverse set of almost 3,000 individuals. The results show that these are not necessarily two sides of the same coin and that depending on the decision environment, ingroup-love and outgroup-hate can co-exist independently from each other. More specifically, across several pre-registered experiments, I approach this research question from three angles, which allows me to disentangle the role of beliefs and preferences: First, I examine the impact of polarization in a *non-strategic* context by borrowing from the existing literature that has studied pro- and anti-social behavior through the lens of the extended dictator game (e.g., List, 2007; Bardsley, 2008; Dimant, 2019), to which I will refer to as the Take-or-Give (T-o-G) Dictator Game. One crucial feature of this game is that, in addition to being able to give money to the recipi-

[2] Greene, 1999; Abramowitz and Saunders, 2006; Mason, 2015; Michelitch, 2015; Orr and Huber, 2020; West and Iyengar, 2020. See also Yamagishi and Mifune, 2009; Amira et al., 2019; Iyengar et al., 2019.

ent, participants can exhibit anti-social behavior by taking money from the recipient. By employing a context in which strategic motives are eliminated by design, the experiment reveals how one's identity shapes altruistic preferences with ingroups and outgroups.

Second, I examine the impact of polarization in a *strategic* context by borrowing from the "Attitudes-Beliefs-Contributions (ABC) of cooperation" approach by Gächter et al. (2017). This method is nested in three variants of a public goods game: a one-shot sequential public goods game played with the strategy method to measure attitudes of cooperation, a belief-elicitation task to measure expectations of others' cooperation, and a one-shot simultaneous public goods game played with the direct response method to measure effective contributions. This approach allows me to answer a few important and policy-relevant questions: does the negative impact of polarization arise because people expect individuals from the opposite faction to be less cooperative (a belief channel)? Or is it the consequence of a lower willingness to cooperate with members of the opposite faction, no matter how cooperative they are (a preference channel)? Distinguishing these mechanisms is vital because it allow us to identify whether a society is truly fractured across factions or whether, in principle, cooperation might be sustained through appropriate belief management.

Third, I employ the social norm elicitation procedure by Krupka and Weber (2013) to examine the extent to which the observed behavioral differences driven by political polarization can be mapped onto the social norm perceptions within and between factions.

I introduce disparate feelings of polarization by using a participant's repugnance against (henceforth referred to as *hate*) or relish for (henceforth referred to as *love*) the 45th president: Donald J. Trump. This is a particularly expedient setting since Trump's actions during his 2016 presidency have been linked to increased social divergence and hate-related consequences.[3] To tease out the role of the emotional state that is produced by the partisan divide, I also run the same experimental conditions with a separate set of participants using the minimal group prime (following Chen and Li, 2009), where one's preferences for Klee or Kandinsky paintings are the identity markers, instead of one's opinion about Trump.

The results highlight that partisan animosity evokes an emotionally charged state that affects social preferences, beliefs, and attitudes of both a strategic and non-strategic nature

[3] See, e.g., Abramowitz and Webster, 2018; Mason, 2018; Müller and Schwarz, 2019; Klein, 2020. The consequences of hate are conspicuous and often erupt in form of social movements and protests (Meyer, 2004; Madestam et al., 2013; Mazumder, 2018; Cantoni et al., 2019). In other contexts, the dating app 'Hater', backed by Mark Cuban, has utilized joint hate as a social glue to facilitate love connections (with self-reported success) by matching people based on their joint hate, from food to lifestyle choices.

in different ways. In particular, by comparing the results between the Trump prime and minimal group prime treatments, I find that ingroup-love only occurs in the context of how one perceives *interpersonal closeness* to others; conversely, outgroup-hate is manifested in one's reduced *altruism* and *cooperativeness* with the opposing faction, as well as in the form of pessimistic beliefs about the opposing faction's cooperativeness. This confirms that the results are not driven by ingroup-outgroup considerations alone, but that the observed disparities in perceptions, beliefs, and own cooperativeness instead largely rest on the emotional state that is evoked by political polarization. Connecting this to insights from the norm-elicitation experiment, the scientific contribution and main takeaway is that partisan identity not only drives costly social behavior, in part due to gloomy beliefs, but it also comports with social norms that people suspect to exist.

From a policy perspective, these findings contribute to the affective polarization insights: correcting misguided beliefs can help to avoid harmful spillovers and intergroup animosities, bridge the 'dehumanization' divide – the misconception of how negatively we think others see us – and also address *negativity bias*, one's inaccurate first- and second-order beliefs about the outgroup's behaviors, intentions, and perception of us (Levendusky and Malhotra, 2016; Flynn et al., 2017; Lees and Cikara, 2020; Moore-Berg et al., 2020). This is particularly important since people are found to preferentially consume and engage with information that aligns with their prior beliefs, which can aggravate the partisan rift (Dorison et al., 2019; Shi et al., 2019; Schwalbe et al., 2020, but see also Becker et al., 2019). My results also connect to the findings by Lelkes and Westwood (2017): while they find (asymmetric) partisan bias that is consistent with my results, they find very little evidence for outgroup-hate in the context of discrimination (see also Westwood et al., 2019). The key differences are the contexts in which I employ incentive-compatible behavioral experiments, suggesting that a nuanced perspective of intergroup conflict is warranted.

Against this backdrop, while my findings indicate that the impact of political polarization can be picked up across all studied (perceptional and behavioral) measures, the results emphasize the nuanced composition of political polarization: the partisan rift might be not as forlorn as previously suggested. In the contexts studied here, the adverse behavioral impact of intergroup conflict can be attributed to one's grim expectations about the cooperativeness of the opposing faction, rather than one's actual unwillingness to cooperate with others. With that, my paper speaks to the effects of polarization and helps to understand how behavior has changed in response to it, providing evidence for how exacerbated the intergroup animosities are under the 45th presidency (Druckman and Levendusky, 2019).

Section 2 (Section 3) details Experiment 1 (Experiment 2), in which the research question is examined in a non-strategic (strategic) context. Section 4 presents Experiment 3, which contains the results from the norm elicitation for both contexts. Section 5 concludes.

2. Experiment 1: Hate and Love in a Non-Strategic Context

2.1. Data Collection and Experimental Design

In this experiment (same as in Experiment 2, see Section 3), data is collected without the use of deception for two types of primes using a between-subject design: the *Trump prime (TP)* and the *Minimal Group Paradigm (MGP) prime.* In TP, participants are asked to state their love/hate opinion about Trump after seeing a picture of Trump wearing a MAGA hat.[4] Participants are then randomized into one of three conditions that vary by the type information that they receive about their randomly assigned partner. That is, participants randomly observed that their matched partner either loved Trump, hated Trump, or had an undisclosed Trump opinion.

In MGP, participants also start by seeing the same picture of and stating their opinion about Trump. Immediately afterwards they are presented with Klee and Kandinsky paintings, asked to choose which they prefer, and these preferences – not their Trump opinion – are then used in the subsequent matching. That is, participants randomly observed that their matched partner either preferred Klee, Kandinsky, or had an undisclosed painting preference, while not knowing that partner's Trump preference.[5] This procedure allows me to keep the role of the Trump prime constant across treatments and focus on the sole effect of being matched according to one's (mis)matched Trump or painting preferences. Thus, conditional on their own Trump opinion/painting preference, participants were allocated to one of the partner preferences conditions at random. Consequently, the between-design captures the dimensions: 2 *(prime)* × 2 *(own Trump/painting preference)* × 3 *(partner's preference)*. Figure 1 illustrates the experimental design.

In sum, this variant of the Dictator Game allows me to take the first step towards investigating the impact of political polarization on altruism, which – unlike regular Dic-

[4] As per pre-registration (https://aspredicted.org/blind.php?x=wt9tg3), only participants who indicated to either hate or love Trump are analyzed, whereas participants who were indifferent are not analyzed. My reasoning for this is to align the analysis with the research question and focus on the role of polarization. This renders the indifferent participants (that MTurk cannot screen out ahead of time) obsolete.

[5] Data for recipients was collected separately and none of those participants took part in any of the experiments or performed any active tasks.

tator Games – also provides me with the opportunity to study both pro-social (giving) and anti-social (taking) behavior simultaneously. Moreover, the contrast with the minimal group prime adds an additional layer of detail in that I am able to distinguish whether the observed behavior with the political prime resembles ingroup-love, outgroup-hate, or both.

Note that, however, the condition in which the partner's opinion about Trump or preference for paintings is *not* revealed to the participant are relegated to the Online Appendix. I do this for two reasons: for one, the hate-love analysis is the main focus of this paper whereas the *unknown opinion/preference* condition is a robustness check. For another, as expected, both perceived closeness with and behavior towards a person with an undisclosed Trump/paintings preference fall right in between the results presented here.

Data from 810 participants[6] in the role of dictators have been collected for the conditions analyzed in the main text.[7] After dropping unusable data, data for 801 dictators is analyzed.[8] The design of this experiment is straightforward and consisted of two stages (details for all stages were announced sequentially): a belief elicitation stage (divided into two parts) followed by a take-or-give Dictator Game. The experiment lasted 10 minutes and dictators earned an average of $4 (including a show-up payment of $0.25). This translates to an hourly wage of $24 and is well above average on MTurk (Hara et al., 2018).

Stage I: Preference Elicitation

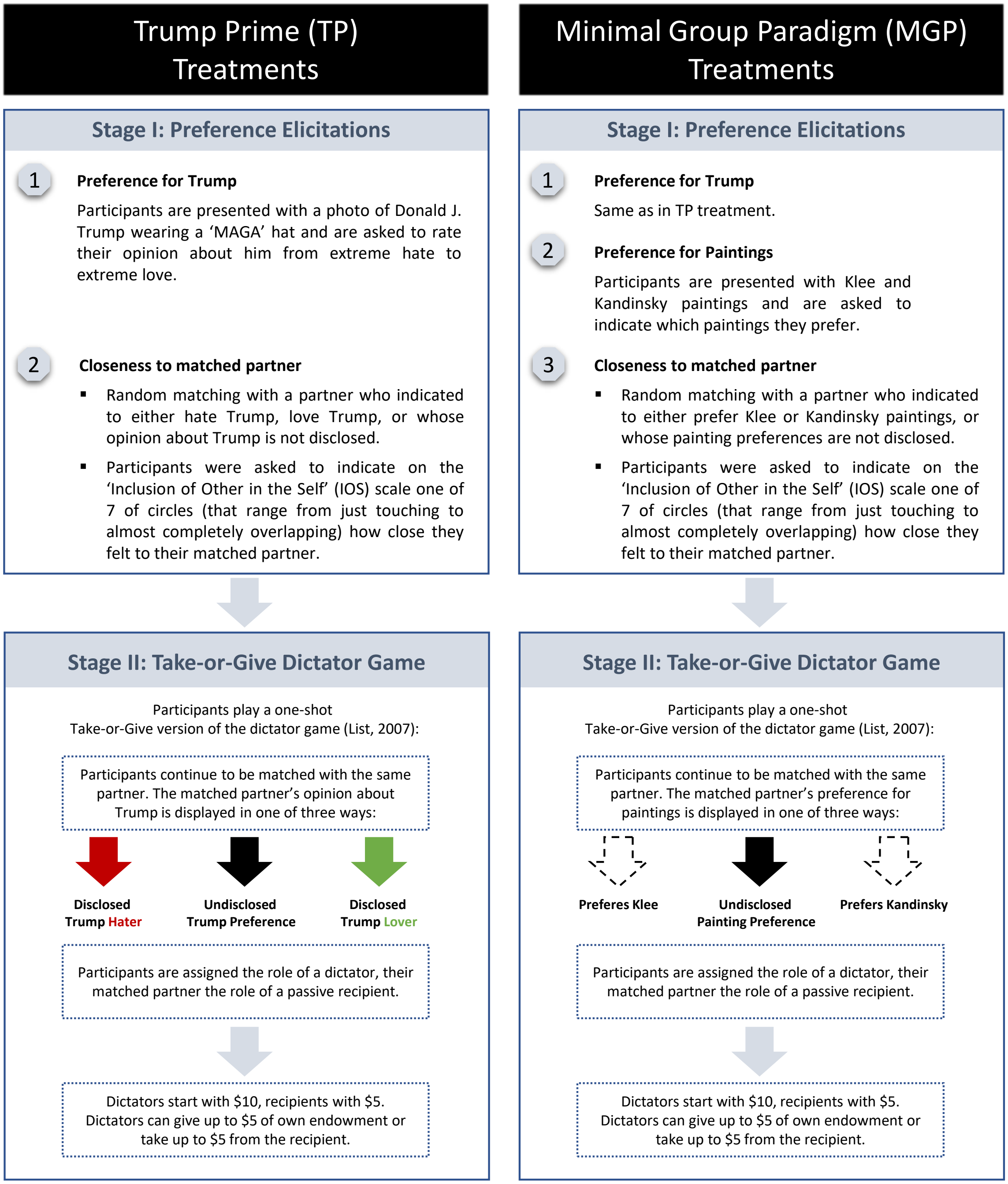
For the *Trump Prime* (TP) treatments, this stage was subdivided into two elicitations: one's opinion about Trump and one's perceived towards the matched partner.

1. In the first elicitation, participants were presented with a photo of Donald J. Trump and had to rate how they personally feel about him (with a focus on the time since

[6] Since MTurk is known to be liberal-leaning, I over-sampled in order to collect enough data for the Trump lovers. As correctly anticipated, those who indicated to love Trump appeared in the data about $\frac{1}{3}$ of the time. I calibrate the required sample size to obtain high statistical power based on a classroom pre-test that yielded an effect size of 0.54. Consequently, the power calculations yielded that 50 participants per cell are needed in order to achieve 80% at an alpha of 0.05. To ensure high quality data collection on MTurk, I utilized a combination of CAPTCHAs and screening questions to avoid pool contamination. I applied the following restrictions to the participant pool: U.S.-based, approval rate greater than 95%, and could participate only once in any of the three experiments presented in this paper. This corresponds to the recommended best practices to maximize data quality (Buhrmester et al., 2018).

[7] Data from participants in conditions that are not analyzed here are presented in the Online Appendix.

[8] Consistent with the pre-registered rules to identify unusable data (failure to pass attention checks, manipulation checks, and/or comprehension checks), Dictator Game data from a total of 18 participants is dropped from the analysis. The appearance of removed participants is uncorrelated with the treatments, and the presented results are not sensitive to the inclusion of these participants (available upon request).



Trump Prime (TP) Treatments

Stage I: Preference Elicitations

1 **Preference for Trump**

Participants are presented with a photo of Donald J. Trump wearing a 'MAGA' hat and are asked to rate their opinion about him from extreme hate to extreme love.

2 **Closeness to matched partner**

- Random matching with a partner who indicated to either hate Trump, love Trump, or whose opinion about Trump is not disclosed.
- Participants were asked to indicate on the 'Inclusion of Other in the Self' (IOS) scale one of 7 of circles (that range from just touching to almost completely overlapping) how close they felt to their matched partner.

Stage II: Take-or-Give Dictator Game

Participants play a one-shot Take-or-Give version of the dictator game (List, 2007):

Participants continue to be matched with the same partner. The matched partner's opinion about Trump is displayed in one of three ways:

Disclosed Trump Hater | **Undisclosed Trump Preference** | **Disclosed Trump Lover**

Participants are assigned the role of a dictator, their matched partner the role of a passive recipient.

Dictators start with $10, recipients with $5. Dictators can give up to $5 of own endowment or take up to $5 from the recipient.

Minimal Group Paradigm (MGP) Treatments

Stage I: Preference Elicitations

1 **Preference for Trump**

Same as in TP treatment.

2 **Preference for Paintings**

Participants are presented with Klee and Kandinsky paintings and are asked to indicate which paintings they prefer.

3 **Closeness to matched partner**

- Random matching with a partner who indicated to either prefer Klee or Kandinsky paintings, or whose painting preferences are not disclosed.
- Participants were asked to indicate on the 'Inclusion of Other in the Self' (IOS) scale one of 7 of circles (that range from just touching to almost completely overlapping) how close they felt to their matched partner.

Stage II: Take-or-Give Dictator Game

Participants play a one-shot Take-or-Give version of the dictator game (List, 2007):

Participants continue to be matched with the same partner. The matched partner's preference for paintings is displayed in one of three ways:

Prefers Klee | **Undisclosed Painting Preference** | **Prefers Kandinsky**

Participants are assigned the role of a dictator, their matched partner the role of a passive recipient.

Dictators start with $10, recipients with $5. Dictators can give up to $5 of own endowment or take up to $5 from the recipient.

Figure 1: Experimental design of the Take-or-Give Dictator Game for both the Trump Prime and the minimal group prime conditions. Note that, for the purpose of brevity, the results for the conditions in which participants were matched with a partner for whom the Trump/painting preference was not disclosed (indicated with a black arrow in this figure) are relegated to the Online Appendix.

he became president) on a 5-point Likert scale: extreme hate, moderate hate, indifferent, moderate love, or extreme love.[9] This method is adapted from the 'feelings thermometer' in the American National Election Study (ANES).

2. In the second elicitation, participants were randomly paired with another passive participant who indicated to either hate Trump (if they indicated either extreme or moderate hate), love Trump (if they indicated either extreme or love), or whose opinion about Trump was not disclosed. Participants were then asked to indicate on the 'Inclusion of Other in the Self' (IOS) scale one out of 7 of circles (ranging from just touching to almost completely overlapping) how close they felt to their matched partner (Aron et al., 1992; Gächter et al., 2015). In the analysis, this scale is converted to percentage (ratio of one's indicated value out of 7).

For the *Minimal Group Paradigm* (MGP) treatments, an additional elicitation stage was included: after eliciting one's opinion about Trump, participants were presented with several paintings from either Klee or Kandinsky and were asked to choose their favorite (design following Chen and Li, 2009). Subsequently, participants were matched with a partner at random whose painting preferences were either disclosed or remained undisclosed. Thus, the matching procedure mirrors exactly the procedure in the TP treatments, except that the matching and the subsequent IOS closeness elicitation were done on the basis of painting preferences instead of Trump opinions.

It is worth stressing why beginning both experiments, TP and MGP, with an elicitation of participants' opinion towards Trump is prudent, even if the matching of MGP does not utilize their opinions about Trump. For one, doing so enables me to hold any residual effect of the thought about Trump constant across both treatments. For another, I can break down and compare the data in both treatments by one's own opinion of Trump, which is a necessary comparison when studying ingroup-love/outgroup-hate.

Stage II: Take-or-Give Dictator Game

In order to capture both pro-social (giving) and anti-social (taking) behavior simultane-

[9] In accordance with the pre-registration, answers for moderate and extreme hate (love) were subsumed under 'hate' ('love') and were also done so in the matching procedure during the experiment. That is, in the treatments in which the matched partner's opinion about Trump was disclosed, participants only observed whether the partner indicated to hate or love Trump, but not the strength (extreme or moderate hate/love) of their opinion. Consequently, they will be treated as a bundled characteristic throughout the experiment. For a full distribution of opinions, see Figure A.1 in the Main Appendix.

ously, I employ a variant of a dictator game that was inspired by existing research (List, 2007; Bardsley, 2008; Dimant, 2019). In this variant, both the dictator and the recipient start with a non-zero endowment and the dictator's action space is augmented with one additional option: the opportunity to take some or all money away from the recipient. One of the many advantages of using this modified version of the game is the ability to measure both pro-social and anti-social tendencies simultaneously (see Dimant (2019) for a discussion). Prior to making the decision, participants are told that – on top of the show-up fee – half of all randomly determined dictator-recipient pairs would be paid a bonus corresponding to their in-game decisions. The remainder half only receive the show-up fee.

For the purpose of my study, I borrow the initial endowment structure from List (2007): the dictator starts with \$10 whereas the recipient starts with \$5.[10] The dictator makes one of the following three decisions exactly once:

1. Take up to \$5 from the recipient's endowment and add to one's own endowment.
2. Make no change to the initial distribution of money.
3. Give out up to \$5 from one's own endowment and add to the recipient's endowment.

2.2. Hypotheses

First, existing literature on identity, ingroup bias, and social proximity suggests that individuals will feel closer to participants who are more 'similar' to them, which will also show up in form of stronger pro-sociality (e.g., Akerlof, 1997; Akerlof and Kranton, 2000; Charness et al., 2007; Fowler and Kam, 2007; Chen and Li, 2009; Christ et al., 2014; Lees and Cikara, 2020). Compared to being matched with someone whose opinion of Trump is undisclosed, greater (lower) amount of pro-sociality and closeness towards a partner with the same Trump opinion will be labeled as *ingroup-love* (*outgroup-hate*).[11] Consequently:

$\mathbf{H_1}$: *Dictators will exhibit the largest closeness score and extent of pro-sociality towards a partner who has the same opinion of Trump (TH-TH or TL-TL), lowest when the matched partner's opinion is misaligned (TH-TL or TL-TH).*

What is more, as argued in the introduction, a contribution of this paper is to examine whether 'hate' is stronger than 'love'. If so, one would expect a disproportional effect for

[10] To retain incentive-compatibility, dictators were told that their allocation decisions are paid out in 50% of the time as bonus at the end of the experiment. If not selected, they only received the show-up fee.

[11] I follow Yamagishi and Mifune (2009) and define these terms as: *Ingroup-love* (*outgroup-hate*) indicates behavior that provides ingroup (outgroup) members with preferential (spiteful) treatment.

both closeness and displayed behavior that explains a host of existing phenomena, including asymmetries between positive and negative reciprocity as well as between the contagion of pro-/anti-social behavior (Offerman, 2002; Croson and Shang, 2008; Lelkes and Westwood, 2017; Dimant, 2019; Bicchieri et al., 2020).

H_2: *Dictators will exhibit disproportionately larger outgroup hate than ingroup love.*

2.3. Behavioral Results

In what follows, results for both TP treatments and MGP treatments are presented in the same Figure 2. Results will be broken up along multiple dimensions for both perception of closeness (henceforth referred to as *perception*, for illustrative purposes presented as % of indicated closeness on a scale from 1 to 7) and behavior in the T-o-G dictator game (henceforth referred to as *behavior*, measured as % of dollar amount given to or taken away from the recipient):[12]

- When the dictator is matched with a partner who has an aligned opinion about Trump (matching corresponds either to *Hate-Hate* or *Love-Love*) compared to being matched with a partner who has a **contrary opinion** about Trump (matching corresponds either to *Love-Hate* or *Hate-Love*). Results are presented in Figure 2.
- Same analysis, but broken down by a participant's opinion on Trump (hate or love). Results are presented in Figure 3.[13]

Comparing the perceived closeness and behavior between the Trump Prime treatments (top panel of Figure 2) and the minimal group prime treatments (bottom panel of Figure 2) yields a number of interesting patterns. First, it is evident that differences in both closeness and behavior only arise in TP and not in MGP, indicating that an ingroup-outgroup differentiation is evoked exclusively by the hate-love prime. Second, zooming in on the actual differences, one can observe that being matched with a partner with an aligned opinion leads to stronger perceived closeness (78.1% vs. 39.8%, BSM, p<0.001) as

[12] Consistent with the pre-registration, the following statistical analyses will be performed in all three experiments: bootstrap two-sample t-test method (BSM) approach as proposed by Moffatt (2015) with 9999 replications. The BSM procedure retains cardinal information without distribution assumptions. Robustness checks will be performed using non-parametric Mann Whitney-U ranksum tests. Unless noted otherwise, the results can be assumed to be consistent between the two methods.

[13] For ease of exposition, I present a more detailed breakdown of both perceived closeness and behavior when matched with another participant whose opinion of Trump was not revealed to the dictator in the Online Appendix (see, for example, Figure OA.2).

well as more pro-sociality (14.9% vs. -22.9%, BSM, $p<0.001$). Notably, I find a marked asymmetry between ingroup-love and outgroup-hate: the absolute negative average amount in the misaligned condition is over-proportionally larger than the positive average amount in the aligned condition ($\left|-22.9\%\right|$ vs. 14.9%, BSM, $p<0.01$).[14] For MGP, the observed differences are trivial in size and neither significant for perception of closeness (37.5% vs. 39.1%, BSM, p=0.51) nor for take-or-give behavior (12.0% vs. 13.5%, BSM, p=0.88). I conclude that the political polarization frame evokes an emotional state that produces traceable changes in both perceptions and behavior beyond the minimal group notion.

Lastly, taking these results together provides a clear indication of whether and where *ingroup-love*, *outgroup-hate*, or both simultaneously exist. In particular, perceptions of closeness towards participants with a contrary opinion (one's outgroup, 39.8%, red bar in top-left panel) are indistinguishable in TP and in MGP, whether compared to someone with the same preferences (37.5%, red bar bottom-left panel, p=0.40) or with the contrary preferences (39.1%, green bar bottom-left panel, p=0.87). Thus, the political priming produces no outgroup-hate with respect to perceived closeness. Conversely, one can clearly observe ingroup-love since participants felt much stronger closeness towards their own political faction (78.1%, green bar in top-left panel) compared to the perceived closeness in the MGP. Interestingly, for actual behavior, one can observe a clear indication of outgroup-hate but no indication of ingroup-love: with essentially fully overlapping error bars, the pro-social behavior towards one's own political faction (14.97%, green bar in top-right panel) is indistinguishable from the behavior in MGP, regardless of whether their partners held the same (13.54%, green bar bottom-right panel, p=0.78) or different (11.99%, red bar bottom-right panel, p=0.56) painting preferences. Unambiguously, all of this is in stark contrast to the behavior observed when matched with someone who has a contrary opinion about Trump: the inflicted harm on this group is large (-22.9%, red bar top-right panel). In sum, the results are compelling as they indicate that ingroup-love and outgroup-hate are context-specific in that they only appear in the political prime. In other words, ingroup-love occurs for perceived closeness and outgroup-hate occurs for actual behavior.

Next, the same two dimensions (perception of closeness and behavior) are broken down

[14] See also Lelkes and Westwood (2017). It is worth noting that this asymmetry is seemingly not *per se* driven by differences between Trump haters and Trump lovers. As Figure OA.1 in the Online Appendix shows, both types display the same average perception of closeness (57.5% vs. 59.0%, BSM, p=0.47). However, it is driven by the fact that the average perceived closeness towards Trump Haters (65.7%) is significantly larger than the perceived closeness towards Trump lovers (53.8%, BSM, $p<0.001$).

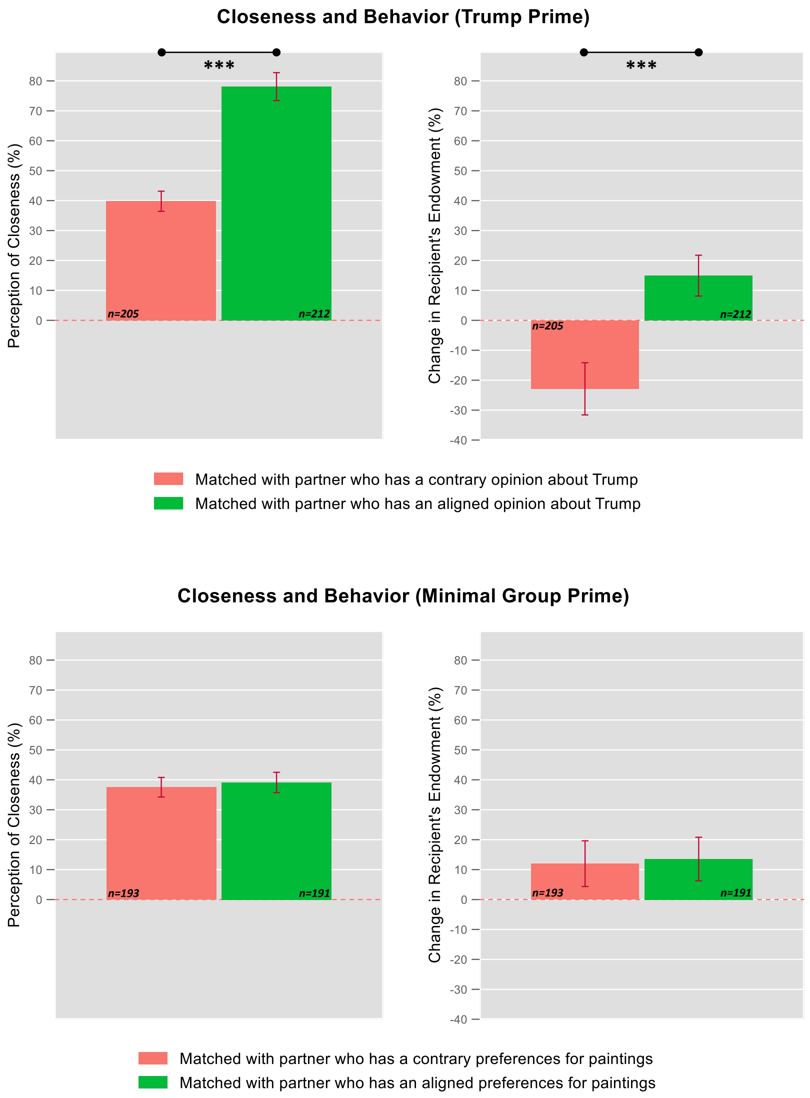


Figure 2: Closeness and behavior broken down by being matched with a partner who either has aligned or contrary opinions for both TP and MGP treatments. Perception of closeness is converted from a 7-point scale to % for illustrative purposes. All adjacent bars are compared and absence of significance stars indicates p-values above 0.05.

by one's own opinion of Trump and are presented in Figure 3. For the conditions using the Trump Prime, I continue to find the previously observed and highly significant differences along both dimensions: for the perception of closeness, participants feel the strongest (weakest) connection with the matched partner that has the same (opposing) view about Trump (all $p<0.001$). The magnitudes and differences are comparable for both Trump haters and Trump lovers. Evidently, a similar pattern also arises for the measured behavior: only those who are matched with a partner with the same opinion on Trump display significant pro-social behavior on average (increase in the recipient's endowment), whereas

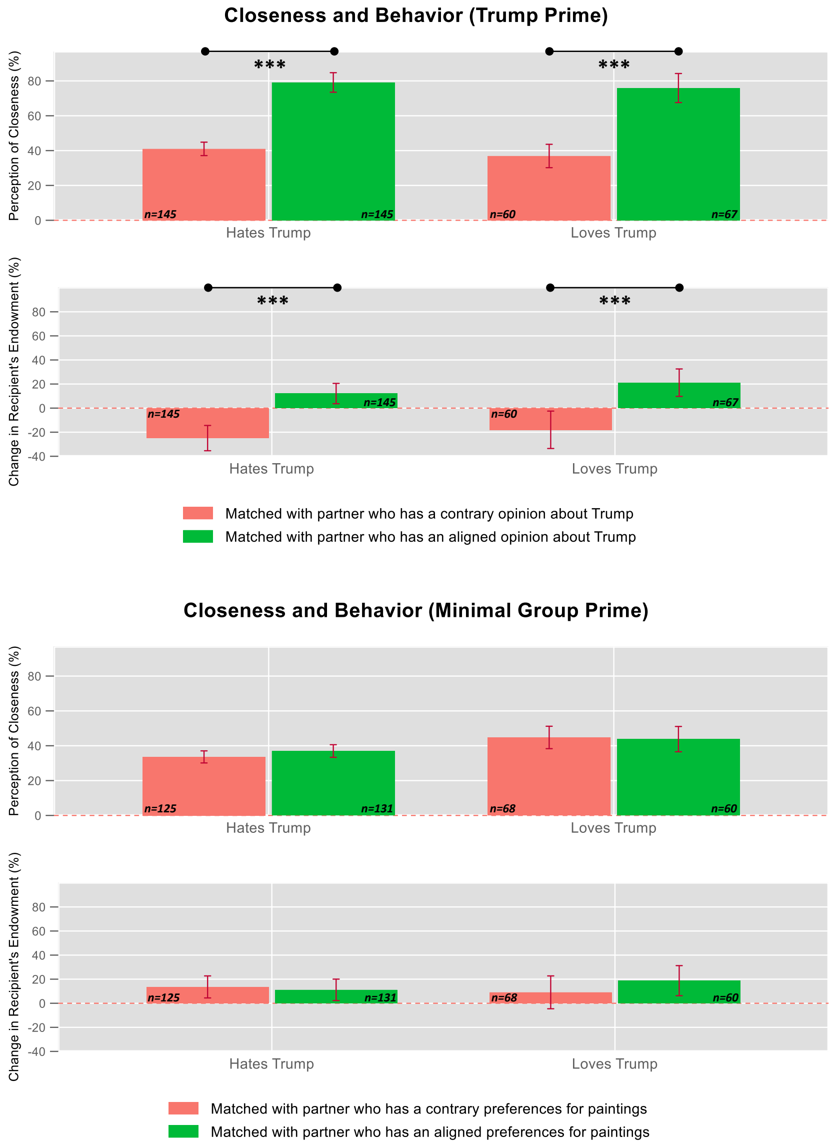


Figure 3: Closeness and behavior broken down by one's own opinion and being matched with a partner who either has aligned or contrary opinions for both TP and MGP treatments. All adjacent bars are compared and absence of significance stars indicates p-values above 0.05.

those who are matched with a partner with opposing views about Trump leads to significant anti-social behavior on average (decrease in the recipient's endowment).

It is worth pointing out that the previously observed asymmetry in behavior (see right-hand side in Figure 2) now reappears in a more nuanced way: one can now observe that the absolute magnitude of anti-social behavior outweighing the absolute magnitude of pro-social behavior is driven by those who hate Trump ($p<0.001$), whereas the absolute magnitudes are not statistically different for those who love Trump (p=0.146).[15]

I find none of these differences in MGP, neither for perceived closeness nor for behavior. Notably, again, anti-social behavior in form of taking only occurs in the Trump Prime conditions while the extent of pro-social behavior in the minimal group prime conditions is indistinguishable from the behavior in the Trump Prime conditions. As later shown in Figure 9, this maps well onto the different norm perceptions between Trump haters and lovers: the former make a clear distinction between harming their ingroup versus the outgroup, whereas the latter do not seem to make such a distinction.

Notably, I once again observe that the levels of perceived closeness in MGP are indistinguishable from that towards someone with a contrary opinion on Trump, and half as much than that towards someone with an aligned opinion on Trump. At the same time, the degree of pro-social behavior in MGP is the same as that of participants in the TP condition when matched with someone who has an aligned Trump opinion, and much higher than the overall anti-social behavior towards those with a contrary Trump opinion. Consistent with the previous findings from Figure 2, I conclude that the concept of ingroup-love and outgroup-hate is more nuanced than some existing research has suggested: in the context of non-strategic decisions, ingroup-love occurs with respect to perceived closeness, whereas outgroup-hate occurs with respect to altruistic behavior.

In a last step, the results are evaluated in a regression framework that includes the collected controls (age, gender, level of education, political affiliation, U.S. citizenship, whether one voted in the 2016 election, and race). Without qualifications, all previously presented findings hold and are presented in Table A.1 in the Main Appendix.

3. Experiment 2: Hate and Love in a Strategic Context

3.1. Data Collection and Experimental Design

I capitalize on a 2 × 2 × 3 experimental design (Trump/minimal group prime × own opinion about Trump × matched partner's opinion about Trump / preference for paintings). Consistent with the analysis in Section 2, a total of 2 × 2 × 2 treatments will be analyzed in the main body of the paper in that I only analyze the beliefs and behavior when participants were either matched with someone who had the same or contrary preferences

[15] As illustrated in Figure OA.2 in the Online Appendix, both the perception of closeness and behavior towards a matched partner whose opinion about Trump was not disclosed falls in between the perceived closeness and behavior towards someone with an aligned and misaligned opinion about Trump.

towards Trump/paintings. For these conditions, data from 752 participants are analyzed.[16] This leaves me with a total of 718 participants whose behavior will be analyzed shortly.

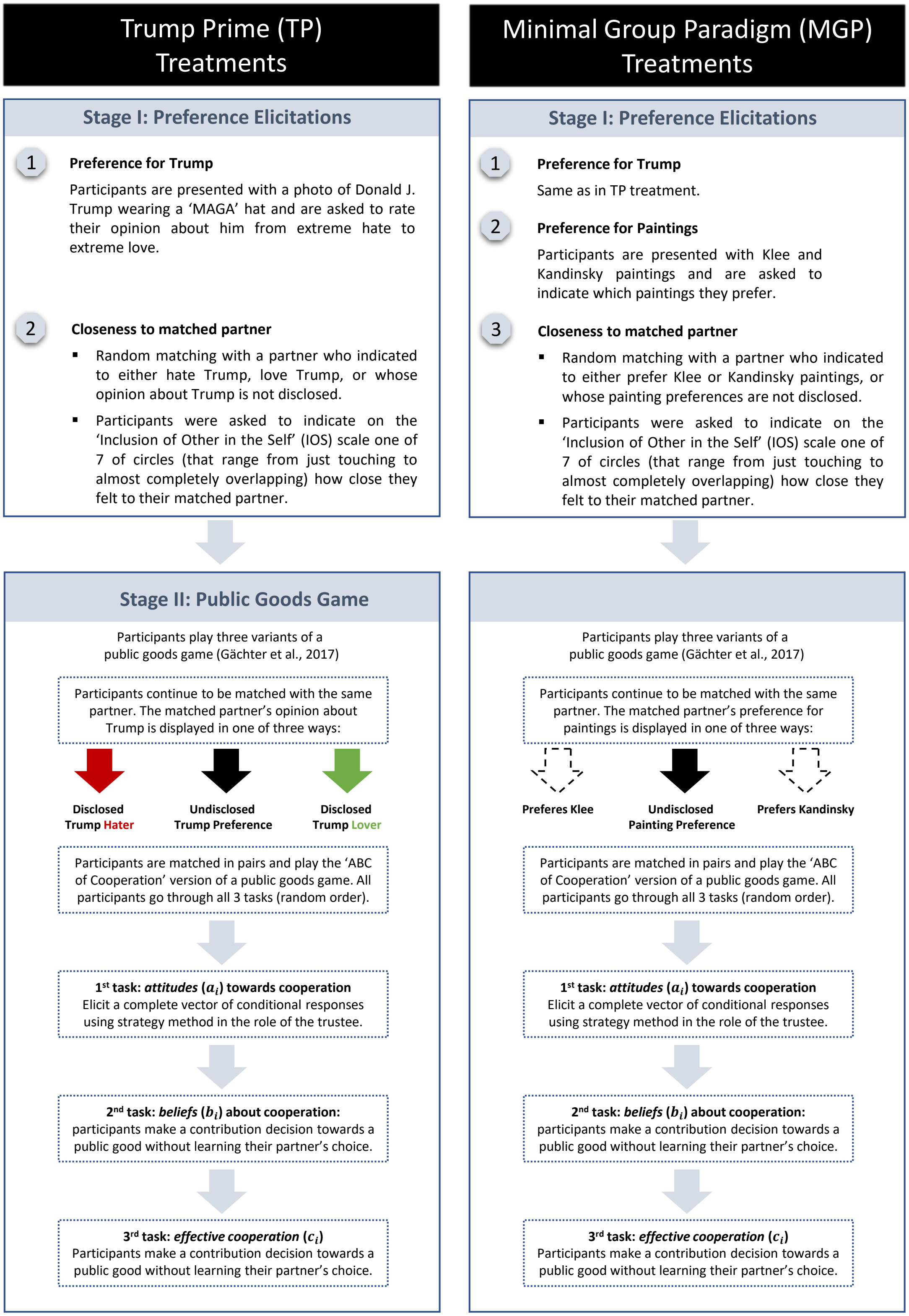
Treatment Variations

This experiment consists of three tasks to measure participants' cooperativeness using the "ABC of cooperation" approach (Gächter et al., 2017): a one-shot sequential public goods game played with the strategy method to measure attitudes of cooperation, a belief-elicitation task to measure expectations of others' cooperation, and a one-shot simultaneous public goods game played with the direct response method to measure effective contributions. The treatment variations closely follow those from Experiment 1 (see Section 2.1 for more details). For the Trump Prime (TP) treatments, the public goods games will be played either by a pair of subjects with the same opinion about Trump (TH-TH or TL-TL), with an opposing opinion about Trump (TH-TL or TL-TH), or by participants for which the opinion about Trump are not disclosed to the other participant (TH-TU or TL-TU). The same applies for the minimal group prime (MGP) treatments in which participants are randomly matched based on own and the partner's painting preferences.[17] In sum, this PGG variant creates the necessary environment to answer my research questions in that it allows me to disentangle the mechanisms through which political polarization operates. By contrasting this to the minimal group setting, I am able to distinguish between ingroup-love and outgroup hate. The design is detailed below and illustrated in Figure 4.

Subjects are matched in pairs and take part in three tasks that are based on the game explained above. All tasks are based on the following two-person, one-shot public goods game. To simplify the mental effort on the side of the participants, I follow the standard notion of the game and use an MPCR of 0.75 (Isaac and Walker, 1988): each player is endowed with $10 that she can either contribute to the public good or keep for herself. Participants are able to give any integer amount between 0 and 10, thus providing eleven options in total. Each dollar contributed to the public good is multiplied by 1.5 and then equally divided between the two participants, irrespective of the individual‘s contribution. As can be seen, the game embodies the classic tension between private and collective

[16] Data from participants in conditions that are not analyzed here are presented in the Online Appendix. The pre-registration protocol (https://aspredicted.org/blind.php?x=bq76mi) with respect to targeted sample size and removal of unusable data mirrors the Dictator Game protocol as explained in Section 2.

[17] Again, note that for the purpose of brevity, results for the conditions where one was matched with a partner for whom the opinions were undisclosed are relegated to the Online Appendix.



Trump Prime (TP) Treatments

Stage I: Preference Elicitations

1 **Preference for Trump**

Participants are presented with a photo of Donald J. Trump wearing a 'MAGA' hat and are asked to rate their opinion about him from extreme hate to extreme love.

2 **Closeness to matched partner**

- Random matching with a partner who indicated to either hate Trump, love Trump, or whose opinion about Trump is not disclosed.
- Participants were asked to indicate on the 'Inclusion of Other in the Self' (IOS) scale one of 7 of circles (that range from just touching to almost completely overlapping) how close they felt to their matched partner.

Stage II: Public Goods Game

Participants play three variants of a public goods game (Gächter et al., 2017)

Participants continue to be matched with the same partner. The matched partner's opinion about Trump is displayed in one of three ways:

Participants are matched in pairs and play the 'ABC of Cooperation' version of a public goods game. All participants go through all 3 tasks (random order).

1st task: *attitudes* (a_i) towards cooperation
Elicit a complete vector of conditional responses using strategy method in the role of the trustee.

2nd task: *beliefs* (b_i) about cooperation:
participants make a contribution decision towards a public good without learning their partner's choice.

3rd task: *effective cooperation* (c_i)
Participants make a contribution decision towards a public good without learning their partner's choice.

Minimal Group Paradigm (MGP) Treatments

Stage I: Preference Elicitations

1 **Preference for Trump**

Same as in TP treatment.

2 **Preference for Paintings**

Participants are presented with Klee and Kandinsky paintings and are asked to indicate which paintings they prefer.

3 **Closeness to matched partner**

- Random matching with a partner who indicated to either prefer Klee or Kandinsky paintings, or whose painting preferences are not disclosed.
- Participants were asked to indicate on the 'Inclusion of Other in the Self' (IOS) scale one of 7 of circles (that range from just touching to almost completely overlapping) how close they felt to their matched partner.

Participants play three variants of a public goods game (Gächter et al., 2017)

Participants continue to be matched with the same partner. The matched partner's preference for paintings is displayed in one of three ways:

Participants are matched in pairs and play the 'ABC of Cooperation' version of a public goods game. All participants go through all 3 tasks (random order).

1st task: *attitudes* (a_i) towards cooperation
Elicit a complete vector of conditional responses using strategy method in the role of the trustee.

2nd task: *beliefs* (b_i) about cooperation:
participants make a contribution decision towards a public good without learning their partner's choice.

3rd task: *effective cooperation* (c_i)
Participants make a contribution decision towards a public good without learning their partner's choice.

Figure 4: Experimental design of the Public Goods Game for both the Trump Prime and the minimal group prime conditions. Note that, for the purpose of brevity, the results for the conditions in which participants were matched with a partner for whom the Trump/painting preference was not disclosed (indicated with a black arrow in this figure) are relegated to the Online Appendix.

interest: while fully contributing to the public goods maximizes joint payoffs, each player's self-interest is maximized by contributing nothing. Subjects play the tasks sequentially and in random order, but receive no feedback on choices or earnings in any of the tasks until the end of the experiment. Only one of the three tasks is used to calculate earnings, and subjects are made aware of this fact at the beginning of the experiment. The task used for calculating payments is randomly selected at the end of the experiment, after subjects' choices in all tasks have been collected. The average pay resulted in about $6.15 (including a $0.25 show-up fee). It took participants on average 15 minutes to complete the experiment, which translates to an hourly payoff of about $24.6, and thus essentially identical to the average payoff in Experiment 1.

In the first task, I use a version of the game described above to measure players' attitudes towards cooperation. Subjects are randomly assigned to be either a first-mover or a second-mover, and only their choices in the relevant role are used to compute payoffs.[18] In order to do so, participants play the game sequentially and I use the strategy method to elicit the second-mover's choices. That is, second-movers are asked to submit a contribution decision for each possible contribution choice made by the first-mover. This ensures that, for each second-mover, one can observe a vector of contributions comprised of eleven choices. I denote subject's i contribution vector as a_{i}. I use this vector of contributions to classify subjects into "cooperation types" that reflect their underlying willingness to cooperate as a function of their opponent's cooperativeness. Following the standard procedure introduced by Fischbacher et al. (2001), subjects are classified into four types:

- *Free riders* if they contribute $0, regardless of the first-mover's contribution.
- *Conditional cooperators* if they have a vector of contributions that is either weakly monotonically increasing in relationship to the first-mover's contribution, or is not monotonically increasing but has a highly significant (at the 1% level) and positive Spearman rank correlation coefficient (between own and others' contribution).
- *Unconditional cooperators* if they contribute a positive amount that does not vary across different first-mover's contributions.
- *Other* if they cannot be classified according to any of the previous criteria.

[18] The task is played using role-uncertainty. That is, all subjects are asked to provide decisions in both the role of first-mover and second-mover, without knowing which role they will actually be assigned in the task until *after* all decisions have been collected.

In the second task, I elicit subjects' effective contributions using a simultaneous version of the public goods game described above. Subjects make a contribution decision in direct-response mode, without learning the contribution choice of their opponent. I denote subject's i effective vector as c_{i}.

In the third task I elicit subjects' expectations about the cooperativeness of their opponent. Subjects are asked to guess the contribution that their opponent has made in the simultaneous public goods game (i.e., the second task). Subjects are rewarded for the accuracy of their guess: if their guess is within \$2, they receive a bonus of \$0.50. I denote the subject's i belief regarding the opponent's contribution as b_{i}.

3.2. Hypotheses

In line with the previously reviewed literature and the hypotheses presented in Experiment 1, an ingroup-love/outgroup-hate effect can be expected.[19]

$\mathbf{H_3}$: *Participants will exhibit stronger closeness, more pronounced attitudes towards cooperation (a_i), higher beliefs about the partner's cooperativeness (b_i), and more effective cooperation (c_i) when matched with a partner who has the same opinion about Trump (TH-TH or TL-TL), whereas these numbers are the lowest when the matched partner's Trump opinion is misaligned (TH-TL or TL-TH).*

As before, this experiment will also examine the asymmetry between hate and love along the tested dimensions.

$\mathbf{H_4}$: *Participants will exhibit disproportionately larger outgroup-hate than ingroup-love.*

3.3. Results

The results from both the *Trump Prime* and the *minimal group prime* reveal a pattern that is consistent with my previous findings: when matched with a partner who has an aligned opinion about Trump, participants felt closer, had higher expectations of the partner's contribution, and effectively contributed more in the PGG (top panel of Figure 5) as compared to when participants were matched with a partner who had a contrary opinion about Trump (all $p<0.01$). In stark contrast, none of these differences appear in the minimal group prime conditions (bottom panel of Figure 5).

[19] As before, while the main focus of this experiment will focus on beliefs and behaviors towards own and opposing factions, the comparison with the participant for whom the opinion about Trump is not disclosed

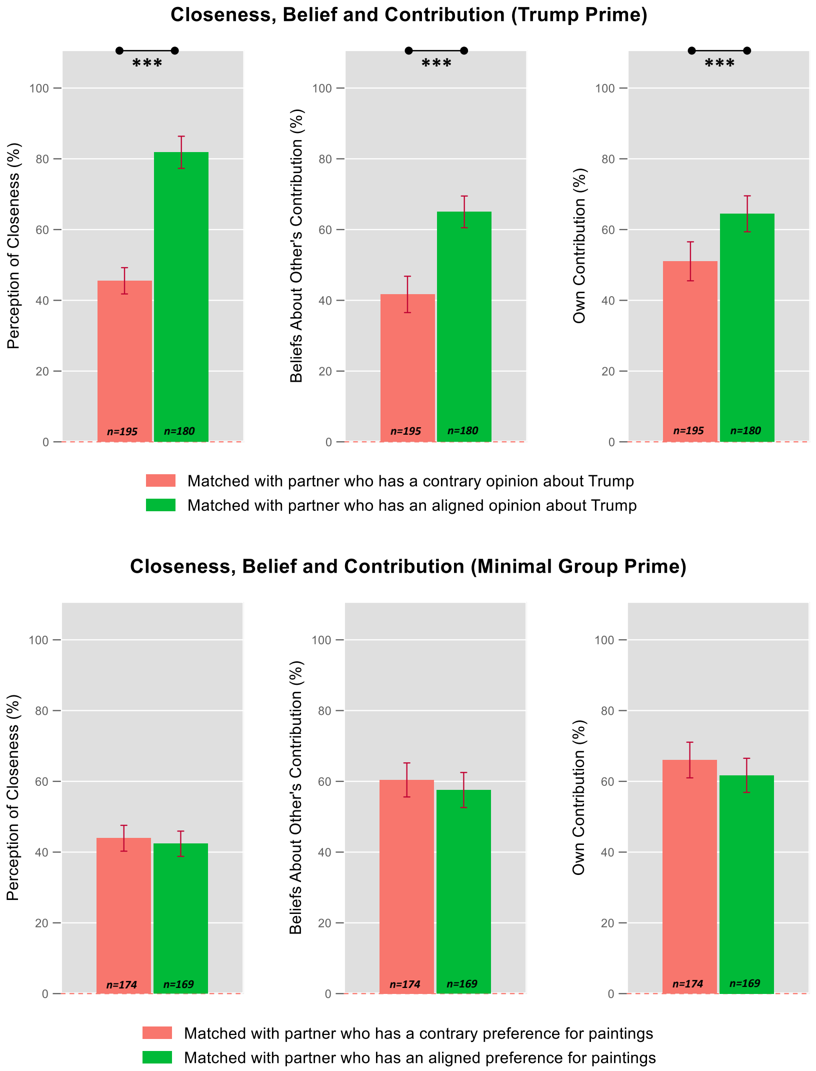


Figure 5: Closeness, belief, and behavior broken down by being matched with a partner who either has aligned or contrary opinions for both TP and MGP treatments. All adjacent bars are compared and absence of significance stars indicates p-values above 0.05.

As is the case for Experiment 1, by comparing the beliefs and behaviors in TP with those in MGP, one can also make statements about whether the observed results derive from ingroup-love, outgroup-hate, or both. Consistent with those previous results, one can

allows me to make to draw an inference about *whether* the observed behavior is ingroup-love or outgroup hate, or both. For brevity, I relegate these pairwise comparisons to the Online Appendix.

again clearly see that the pattern of perceived closeness is one of ingroup-love (red bars have about the same height, whereas the green bars are much higher in TP than in MGP). On the contrary, for both beliefs about other's contributions and one‘s own contribution, we observe outgroup-hate rather than ingroup-love (green bars have about the same height, whereas the red bars are much lower in TP than in MGP).

Next, I zoom in on the perception of closeness, the beliefs (b_i) about other's contributions and one's own effective contributions (c_i) in the PGG, and present the results in Figure 6. For Trump haters, the results are remarkably consistent: for all three measures, the magnitude is significantly higher when matched with another Trump hater (all comparisons p<0.001). For the Trump lovers, on the other hand, one can observe that the discrimination between ingroup and outgroup only holds for the perceived closeness and b_i, but not for c_i. There, Trump lovers contribute a statistically indistinguishable amount of about 55-62% of the maximum amount, irrespective of whether they were matched with another Trump lover or a Trump hater (p=0.253). As later shown in Figure 11, these findings are consistent with the norm elicitation: Trump lovers do not display an ingroup/outgroup differentiation in terms of free riding or cooperation, whereas Trump haters do. The regressions in Table A.2 (presented in the Main Appendix) confirm these results, and it is evident that the observed behavioral differences are driven by the differences in perceived closeness. As before, no significant differences occur along all these dimensions for the minimal group prime conditions (bottom Figure 6). This confirms again that the results are not driven by ingroup-outgroup considerations alone, but that the observed disparities in perceptions, beliefs, and own cooperativeness largely rest on the emotional state that comes with the political polarization.

As before, I examine whether the results reflect ingroup-love, outgroup-hate, or both by comparing both beliefs and behaviors in TP to those in MGP. I reach the same conclusions as above for both those who hate Trump and those who love Trump: ingroup-love occurs for perception of closeness, whereas outgroup-hate occurs for beliefs about other's contributions and one's own contribution.[20] Consistent with the theme and the previous findings

[20] This finding is further confirmed when comparing one's beliefs about other's contribution and one's own contribution in the treatment where participants were matched with someone whose Trump opinion is not disclosed. This comparison is largely insignificant (significant) when tested against beliefs and behaviors towards a partner with the same (contrary) opinion on Trump. A detailed breakdown including the treatment in which participants were matched with a partner whose preference towards Trump was not disclosed is illustrated in Figure OA.4 in the Online Appendix.

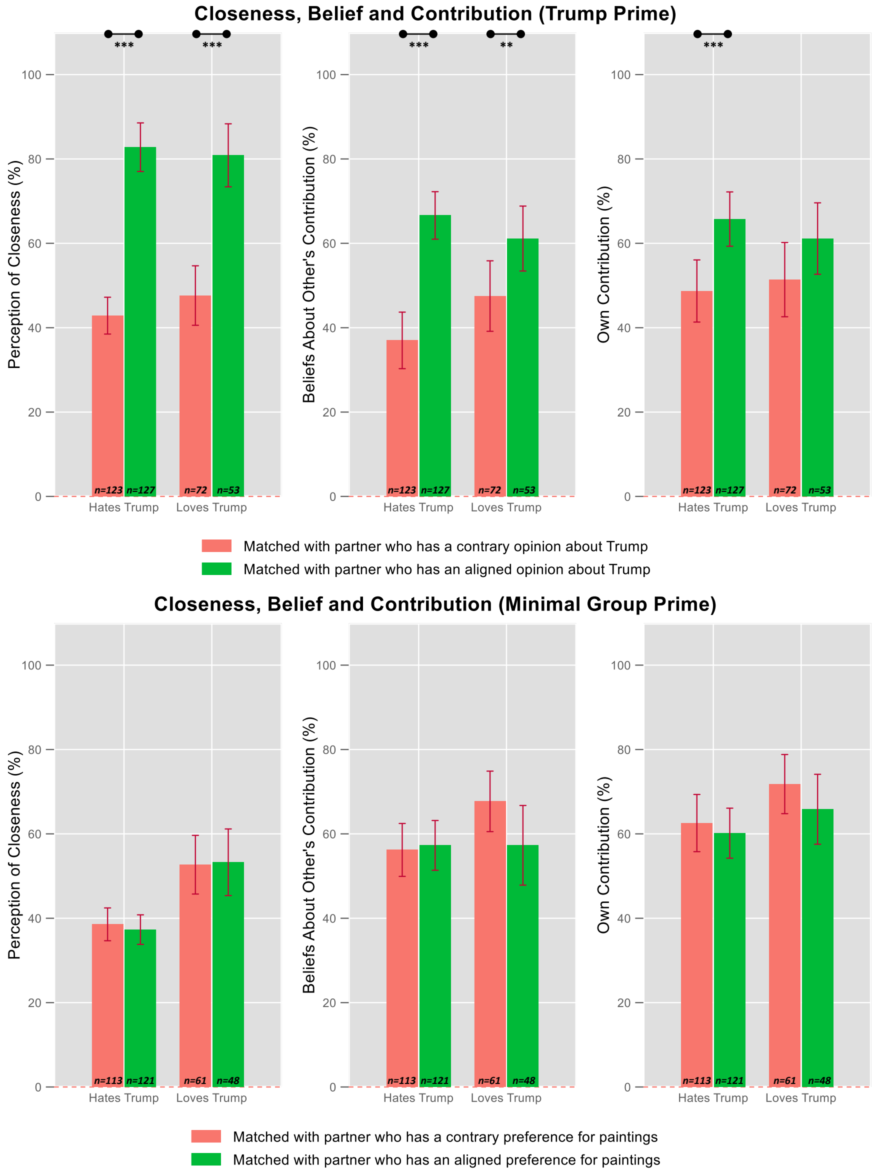
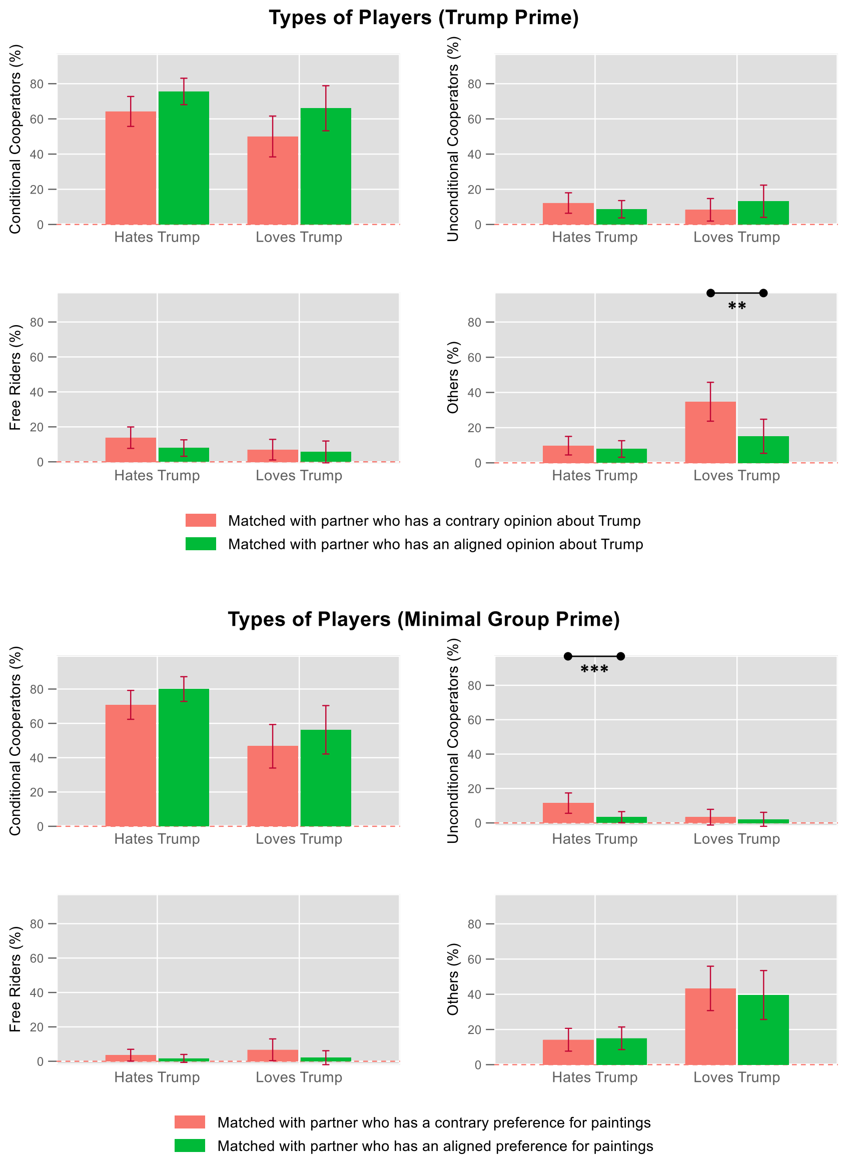


Figure 6: broken down by being matched with a partner who either has aligned or contrary opinions for both TP and MGP treatments. All adjacent bars are compared and absence of significance stars indicates p-values above 0.05.

of this paper, ingroup-love and outgroup-hate are nuanced concepts and their occurrence depends on what is investigated.

For the final part of the investigation, I follow the tradition of Fischbacher et al. (2001) and analyze the distribution of 'types' across the various treatments. I follow the previously introduced classification and distinguish between *Conditional Cooperators* (CC), *Unconditional Cooperators* (UC), *Free Riders* (FR), and *Others* (see Section 3.1 for details).

Zooming in, one can observe that essentially all type classification are insensitive to



Types of Players (Trump Prime)

Types of Players (Minimal Group Prime)

Figure 7: Closeness, belief, and behavior broken down by one's own opinion and being matched with a partner who either has aligned or contrary opinions for both TP and MGP treatments. All adjacent bars are compared and absence of significance stars indicates p-values above 0.05.

whom one is matched with, regardless of the treatment prime (Figure 7).[21] In combination with the previous insights from the Trump prime conditions, this is a key result worth

[21] Most differences do not achieve the pre-registered alpha level of 5%. For TP, the only significant differences is observed for the 'Others' group with p=0.011. Although visually distinct, the differences for Conditional Cooperators only achieve significance at the 10% level at p=0.06 and p=0.09 for Trump-hater and Trump-lover, respectively. For MGP, the only reliably significant difference (p<0.01) can be observed for the Unconditional Cooperators among Trump haters. For Conditional Cooperators, the differences reach p=0.11 and p=0.34 for Trump haters and Trump lovers, respectively.

stressing: the observed inter-faction animosity in form of ingroup/outgroup variability in contributions is a result of fallacious beliefs about the other's behavior (see top panel of Figure 6) and *not* of adverse preferences per se (see top panel of Figure 7). In the latter, it becomes apparent that participants are perfectly fine to cooperate with the opposing faction, regardless of one's partisanship. The implication is that the impact of political polarization should be counteracted by correcting the bleak expectations that the factions have about each other.

4. Experiment 3: Social Norms in Non-Strategic and Strategic Contexts

4.1. Data Collection and Experimental Design

In this experiment, I analyze the norm perceptions of Trump haters and Trump lovers utilizing the incentive-compatible approach by Krupka and Weber (2013) across various contexts. The contribution of this experiment is to understand whether the social norm perceptions map onto the heterogeneous ingroup-love and outgroup-hate by contrasting the results of this experiment with the previously discussed results (Experiments 2 and 3). To achieve robustness and maximize statistical power, the design contains both between- and within-subject variation:

- **Between-subject variation:** As before, participants first see a picture of Donald J. Trump and are asked to indicate on a Likert scale how they feel about him. Thus, hate and love towards Trump constitute the between-subject dimension.

- **Within-subject variation:** In random order, participants were informed of the structure of the T-o-G dictator game and PGG exactly as it had been explained to participants in Experiment 1 and Experiment 2. Subsequently, using the elicitation technique of Krupka and Weber (2013), participants were asked to rate the appropriateness of various behaviors in those games that were presented in random order.[22]

[22] As is customary in this norm elicitation procedure, the participants were asked to rate the appropriateness of the observed behavior along four dimensions: *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SI), *Somewhat Socially Appropriate* (SA), and *Very Socially Appropriate* (VSA). For the dictator game, participants were asked to rate the appropriateness for three distinct behaviors: the dictator making no change to the initial endowments, the dictator taking money from the receiver, and the dictator giving money to the receiver. For the PGG, the participants rated the appropriateness for four distinct behaviors: contribute nothing, contribute nothing when others contribute something (= free-rider), contribute everything (= full cooperator), and contribute more the more the matched partner contributes (= conditional cooperator). Other results are relegated to the Online Appendix.

> To stay as close to the original designs as possible, these ratings were elicited from the perspective of being matched with other participants who either had the same, the opposite, or an unknown opinion about Trump. Participants observed all three variations in random order. Importantly, to ensure reliable norm-inferences, each participant's beliefs were elicited only from the perspective of one's own opinion about Trump.[23] These matching variations constitute the three within-subject dimensions.

In sum, the two within- and three between-subject variations represent exactly the same $2 \times 3 = 6$ dimensions, the same dimensions explored in Experiments 1 and 2. To achieve proper statistical power, observations from a total of n=298 participants were collected (leaving me with n=232 participants after applying pre-registered removal criteria), neither of which have previously participated in Experiment 1 or Experiment 2.[24] Out of these, 162 participants (70%) indicated to hate Trump and 70 participants (30%) indicated to love Trump – a split comparable to the one obtained in Experiments 1 and 2.

4.2. Results for the T-o-G Dictator Game

In this section, the results from the dictator game, as discussed in Section 2.3 (especially in Figures 2 and 3), will be analyzed through the lens of a norm elicitation that follows the method of Krupka and Weber (2013).[25]

The first set of results is presented in Figure 8 and paints a picture that is extremely consistent with both the observed closeness and dictator behavior behavior, as illustrated in Figure 2: when matched with a partner who has an **aligned** opinion about Trump, taking from (giving to) that partner is perceived as more inappropriate (more appropriate) compared to when matched with a partner who has a **misaligned** opinion about Trump (all differences significant at $p<0.001$ using BSM tests).

[23] For example, a participant who initially indicated to hate Trump would only be asked to rate the appropriateness of various behaviors based on the matching of a Trump-hater with either another Trump-hater, a Trump-lover, or someone with an unknown opinion of Trump. Similarly, a Trump-lover would only be asked to rate the appropriateness from the perspective of another Trump-lover having been matched with one of the three possible partners.

[24] As before and per pre-registration (https://aspredicted.org/blind.php?x=xn9vb3), unusable data were dropped either because participants indicated to be indifferent about Trump or because they did not pass the attention/manipulation checks.

[25] Consistent with the previous analyses, the main focus remains the behavior towards partners with the same or contrary opinion of Trump. In the Online Appendix, I present analyses that include perceptions when matched with a partner for whom the opinion about Trump remains undisclosed.

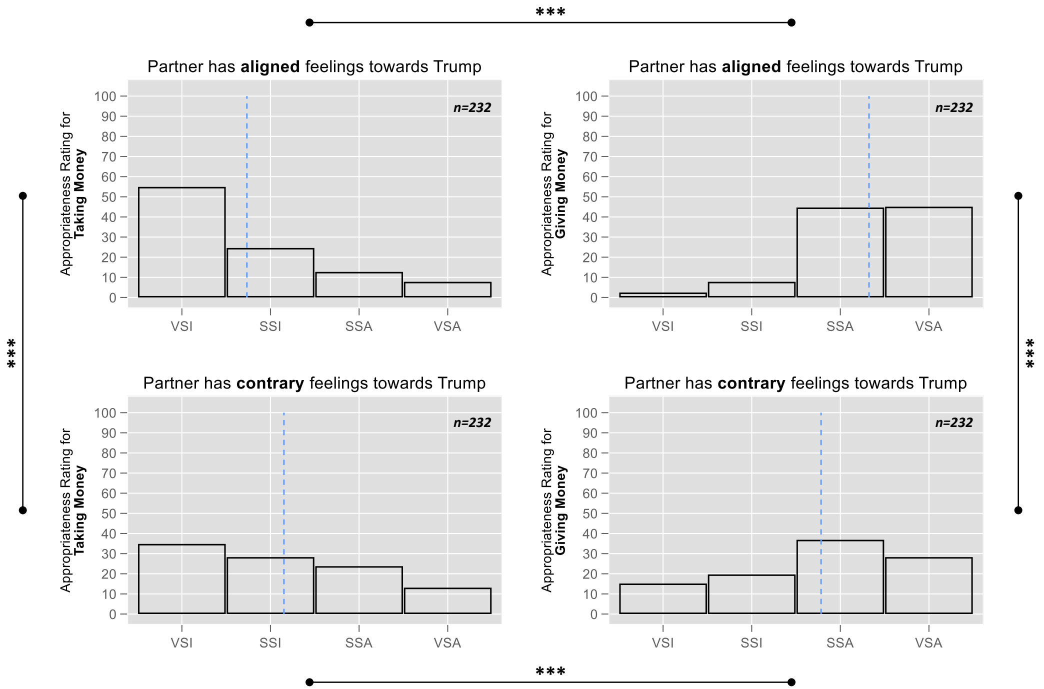


Figure 8: Norm perceptions for taking and giving money with partners who have aligned or contrary feelings towards Trump. All adjacent quadrants are tested and statistical significance (if either ***p<0.01 or **p<0.05) is indicated where applicable. *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SSI), *Somewhat Socially Appropriate* (SSA), and *Very Socially Appropriate* (VSA).

These insights complement the results from Figure 2 and suggest that the observed differences in feeling of closeness and pro-sociality towards a partner who has an aligned opinion about Trump go hand in hand with the norm perception that this is indeed the right thing to do, whereas it is perceived to be more appropriate to harm someone with a contrary opinion about Trump.

Next, following the previous analyses in Figure 3, I analyze the norm perceptions conditional on one's own opinion about Trump and present the results separately for taking behavior (top of Figure 9) and giving behavior (bottom of Figure 9). As before, the norm elicitations are consistent with the observed closeness and dictator game behaviors.

For taking behavior, those who identified as *Trump haters* indicate that it is more acceptable to take from a *Trump-lover* (TH-TL) than from a fellow *Trump-hater* (TH-TH), which is highly statistically significant (BSM, p<0.001). Conversely, I do not observe the same difference for those who identified as *Trump lovers* (BSM, p=0.81), which is primarily driven by the fact that those who are matched with their own kind have a substantially higher approval for taking money from their partner than Trump haters have when matched with their own kind (comparing TL-TL vs. TH-TH, BSM, p<0.001).

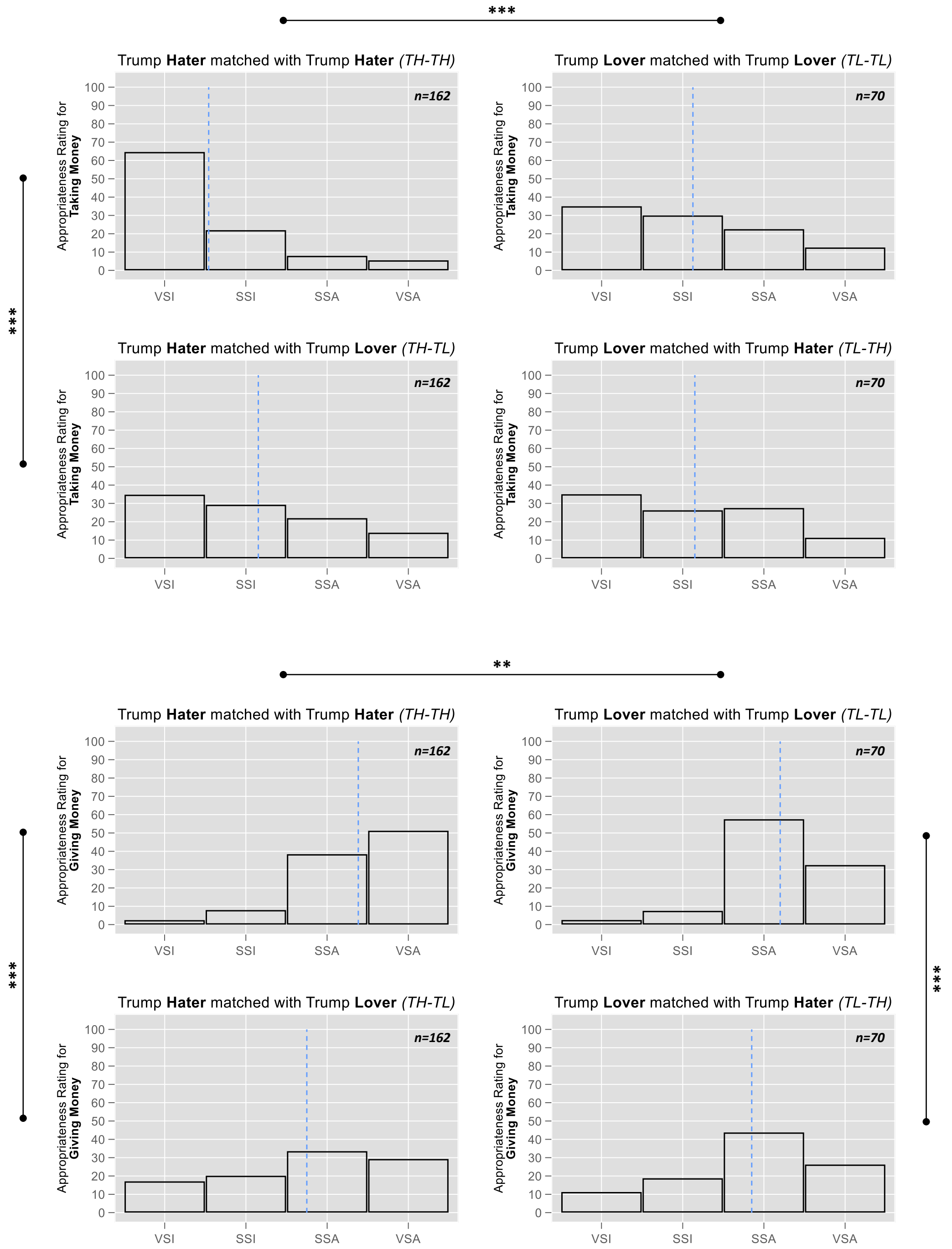


Figure 9: Norm perceptions for taking and giving money conditional on own and matched partner's Trump opinion. All adjacent quadrants are tested and statistical significance (if either ***p<0.01 or **p<0.05) is indicated where applicable. *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SSI), *Somewhat Socially Appropriate* (SSA), and *Very Socially Appropriate* (VSA).

Consistent with the theme of this paper, these results indicate that hate evokes stronger norms against harming each other and, additionally, those who love Trump do not distinguish between ingroup-love and outgroup-hate with respect to harming others.

In terms of giving behavior, one can observe that being matched with a participant with the same preference for Trump leads to a significantly higher appropriateness rating compared to giving to a participant with a misaligned opinion of Trump (comparing TH-TH vs. TH-TL and TL-TL vs. TL-TH, BSM, both p-vales <0.001). In addition, consistent with the previous results, joint hate for Trump evokes a stronger bond in the form of appropriateness for giving than joint love (comparing TH-TH vs. TL-TL, BSM, p=0.0199).

Taken together, one can conclude that the perceived social norms map convincingly onto the observed T-o-G dictator game behavior and can explain the observed behavioral differences between Trump haters and Trump lovers as well as their perceptions and attitudes towards people with aligned and misaligned opinions about Donald J. Trump.

4.3. Results for the Public Goods Game

Following the previous analyses, this section reports the norm perceptions across various possible behaviors in the PGG (free-riding and full cooperation) for the different treatments and reported Trump preferences.[26] These behaviors are defined as followed: **free-riding** refers to the decision to benefit from the public good by contributing nothing, even though one's matched partner contributes a non-zero amount. **Full cooperation** refers to the decision to contribute the full amount regardless of the partner's behavior.

The results presented in Figure 10 paint a clear picture: participants perceive it as *more* socially appropriate to free-ride on a partner who has a contrary opinion about Trump, but less socially appropriate to fully cooperate with the same partner (both p<0.01).

In addition, one can observe in Figure 11 that the previous results very much depend on one's stated preference towards Trump: the previously mentioned differential perception of appropriateness for *free-riding* is entirely driven by Trump haters (p<0.001), whereas there is no significant difference for Trump lovers (p=0.47). The same is true for full cooperation (bottom of Figure 11): Trump haters perceive it as more socially appropriate to fully cooperate with a partner who has an aligned Trump opinion (p<0.001). Again, Trump lovers do not make a distinction irrespective of whom they are matched with (p=0.72). This maps well onto the result presented in Figure 6 (top-right panel) showing that only Trump haters make an ingroup-outgroup differentiation in their level of contribution.

[26] In the Online Appendix, I also present the norm elicitations for two other behaviors: *Contribute Nothing* and *Conditional Cooperator*.

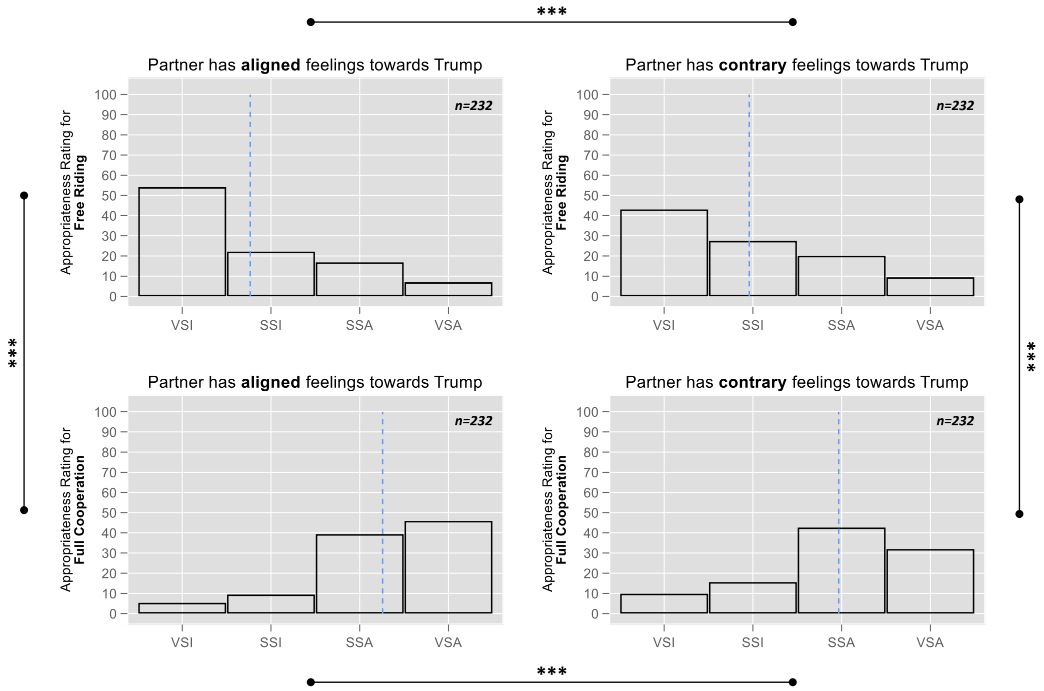


Figure 10: Norm perceptions for free riding and full cooperation with partners who have aligned or contrary feelings towards Trump. All adjacent quadrants are tested and statistical significance (if either ***p<0.01 or **p<0.05) is indicated where applicable. *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SSI), *Somewhat Socially Appropriate* (SSA), and *Very Socially Appropriate* (VSA).

Noteworthy, these findings are consistent with the results discussed in Section 3.3 in that Trump haters show a clear ingroup-love/outgroup-hate distinction, whereas Trump lovers do not seem to make this distinction and treat either participant in the same way. It is important to note that although Trump lovers do not discriminate between their matched partners, they perceive it as much more socially appropriate to free-ride on their partner than Trump haters do (p<0.001, not illustrated).

From a big picture perspective, the findings are in harmony with the existing social norms research and can be subsumed under the umbrella of *conditional norm followers* (Bicchieri, 2006; Bicchieri et al., 2019): people display a preference for cooperation that is conditional on *empirical expectations* (beliefs about the matched partner's behavior, as measured in Experiment 3) and *normative expectations* (as measured in this experiment using the method by Krupka and Weber, 2013). Combining both types of elicitations provides a comprehensive evaluation of beliefs, preferences, and behaviors.

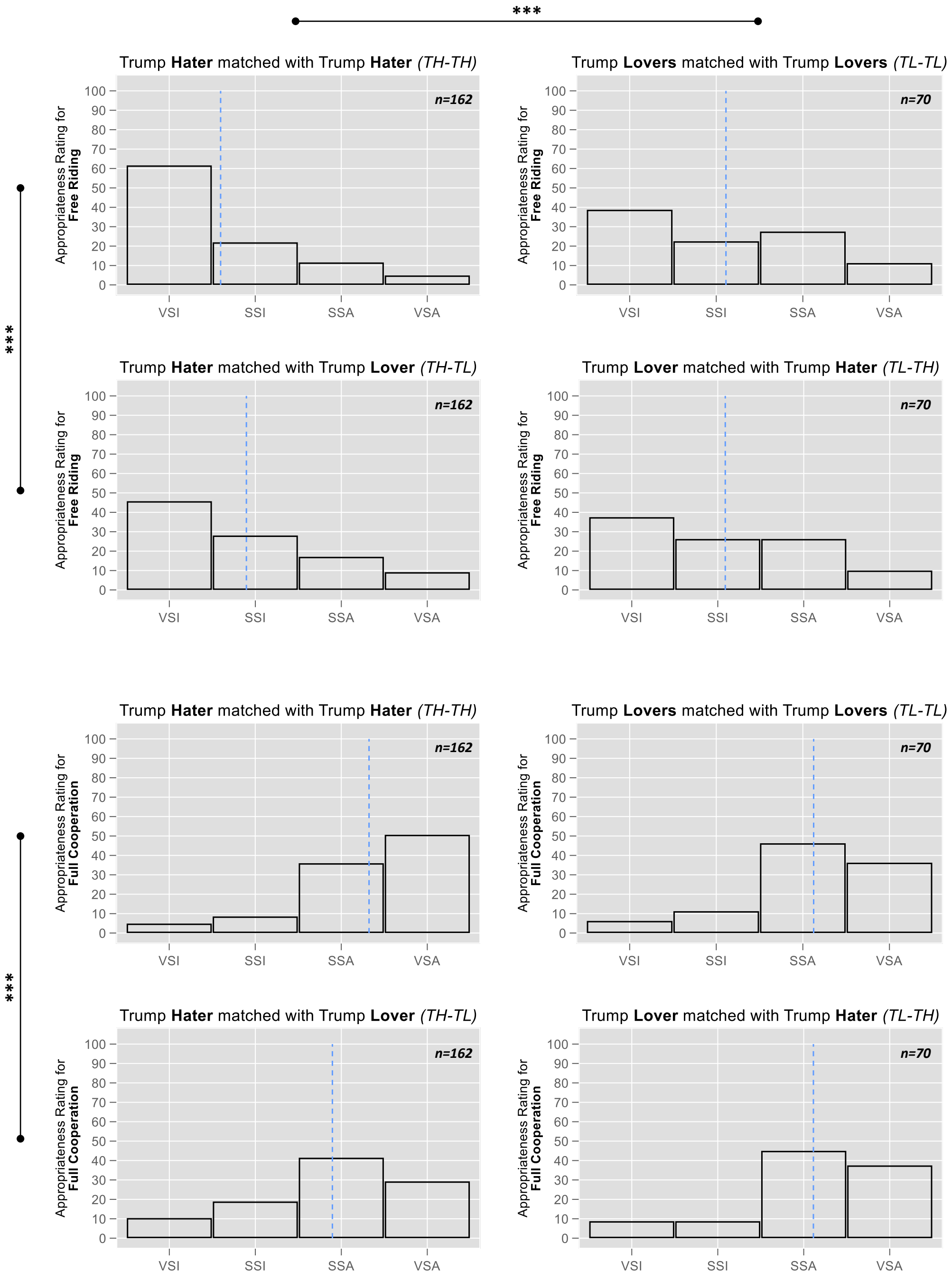


Figure 11: Norm perceptions for free riding and full cooperation conditional on own and matched partner's Trump opinion. All adjacent quadrants are tested and statistical significance (if either ***p<0.01 or **p<0.05) is indicated where applicable. *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SSI), *Somewhat Socially Appropriate* (SSA), and *Very Socially Appropriate* (VSA).

5. Conclusion and Discussion

I examine the impact of political polarization on behaviors, beliefs, and norms through the lens of three experiments of both strategic and non-strategic nature. This paper is concerned with the extent to which polarization succeeds in affecting pro- and anti-social behavior, cooperativeness, and the perception of social norms with respect to these behaviors. I embed polarization by capitalizing on participants' negative/positive opinions about Donald J. Trump and compare the outcomes to those observed in treatments using the standard minimal identity paradigm to disentangle ingroup-love from outgroup-hate.

Along all investigated dimensions, I obtain strong effects and the following results: for one, polarization produces ingroup/outgroup differentiation in all three settings (non-strategic, Experiment 1; strategic, Experiment 2; social norms, Experiment 3), leading participants to actively harm and cooperate less with participants from the opposing faction. For another, lack of cooperation is not the result of a categorical unwillingness to cooperate across factions, but based on one's grim expectations about the other's willingness to cooperate. Importantly, however, the results also cast light on the nuance with which ingroup-love and outgroup-hate – something that existing literature often takes as being two sides of the same coin – occurs. In particular, by comparing behavior between the Trump Prime and minimal group prime treatments, the results suggest that ingroup-love can be observed in terms of feeling close to one another, whereas outgroup hate appears in form of taking money away from and being less cooperative with each other. The elicited norms are consistent with these observations and also point out that those who love Trump have a much weaker ingroup/outgroup differentiation than those who hate Trump do.

The results yield actionable policy prescriptions: by correcting the misguided beliefs about the preferences and actions of the opposing faction, one may be able to alleviate the pernicious outcomes in the context of cooperative and altruistic behavior. Existing research (e.g., Bicchieri and Dimant, 2019; Bursztyn et al., 2020, for a discussion in the context of politics see Flynn et al., 2017 and for a cross-cultural perspective see Ruggeri et al., 2020) has pointed to ways in which such social beliefs can be corrected – or at the least abate their inaccuracy – and be utilized by future research in the context of political polarization. This is particularly important since people are found to consume political information selectively, which aggravates the partisan rift (Dorison et al., 2019). An avenue for future research is to put more weight on the role of context and methodological approaches in uncovering the various forms in which polarization occurs and can be contained.

References

Abramowitz, A. I. and Saunders, K. L. (2006). Exploring the bases of partisanship in the american electorate: Social identity vs. ideology. *Political Research Quarterly*, 59(2):175–187.

Abramowitz, A. I. and Webster, S. W. (2018). Negative partisanship: Why americans dislike parties but behave like rabid partisans. *Political Psychology*, 39:119–135.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica: Journal of the Econometric Society*, pages 1005–1027.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Amira, K., Wright, J. C., and Goya-Tocchetto, D. (2019). In-group love versus out-group hate: Which is more important to partisans and when? *Political Behavior*, pages 1–22.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4):596.

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2020). Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review*.

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11(2):122–133.

Becker, J., Porter, E., and Centola, D. (2019). The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences*, 116(22):10717–10722.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64.

Bernhard, H., Fischbacher, U., and Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442(7105):912–915.

Bhanot, S. P. and Hopkins, D. J. (2020). Partisan polarization and resistance to elite messages: Results from survey experiments on social distancing. *Journal of Behavioral Public Administration.*

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

Bicchieri, C. and Dimant, E. (2019). Nudging with Care: The Risks and Benefits of Social Information. *Public Choice.*

Bicchieri, C., Dimant, E., Gächter, S., and Nosenzo, D. (2020). Observability, social proximity, and the erosion of norm compliance. Working Paper Available at SSRN: https://ssrn.com/abstract=3355028.

Bicchieri, C., Dimant, E., and Sonderegger, S. (2019). It's not a lie if you believe the norm does not apply: Conditional norm-following with strategic beliefs. Working Paper Available at SSRN: https://dx.doi.org/10.2139/ssrn.3326146.

Bowles, S. and Gintis, H. (2013). *A cooperative species: Human reciprocity and its evolution*. Princeton University Press.

Buhrmester, M. D., Talaifar, S., and Gosling, S. D. (2018). An evaluation of amazon's mechanical turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, 13(2):149–154.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American Economic Review*.

Cantoni, D., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2019). Protests as strategic games: experimental evidence from hong kong's antiauthoritarian movement. *The Quarterly Journal of Economics*, 134(2):1021–1077.

Charness, G., Rigotti, L., and Rustichini, A. (2007). Individual behavior and group membership. *American Economic Review*, 97(4):1340–1352.

Chen, Y. and Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1):431–57.

Christ, O., Schmid, K., Lolliot, S., Swart, H., Stolle, D., Tausch, N., Al Ramiah, A., Wagner, U., Vertovec, S., and Hewstone, M. (2014). Contextual effect of positive intergroup contact on outgroup prejudice. *Proceedings of the National Academy of Sciences*, 111(11):3996–4000.

Croson, R. and Shang, J. Y. (2008). The impact of downward social information on contribution decisions. *Experimental Economics*, 11(3):221–233.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73:66–88.

Dixit, A. K. and Weibull, J. W. (2007). Political polarization. *Proceedings of the National Academy of Sciences*, 104(18):7351–7356.

Dorison, C. A., Minson, J. A., and Rogers, T. (2019). Selective exposure partly relies on faulty affective forecasts. *Cognition*, 188:98–107.

Druckman, J. N. and Levendusky, M. S. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly*, 83(1):114–122.

Efferson, C., Lalive, R., and Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science*, 321(5897):1844–1849.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960):785–791.

Fiorina, M. P. and Abrams, S. J. (2008). Political polarization in the american public. *Annual Review of Political Science*, 11:563–588.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.

Flynn, D., Nyhan, B., and Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38:127–150.

Fowler, J. H. and Kam, C. D. (2007). Beyond the self: Social identity, altruism, and political participation. *The Journal of Politics*, 69(3):813–827.

Gächter, S., Kölle, F., and Quercia, S. (2017). Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour*, 1(9):650.

Gächter, S., Starmer, C., and Tufano, F. (2015). Measuring the closeness of relationships: a comprehensive evaluation of the 'inclusion of the other in the self' scale. *PloS one*, 10(6).

Graham, M. H. and Svolik, M. W. (2020). Democracy in america? partisanship, polarization, and the robustness of support for democracy in the united states. *American Political Science Review*, 114(2):392–409.

Greene, S. (1999). Understanding party identification: A social identity approach. *Political Psychology*, 20(2):393–403.

Halevy, N., Bornstein, G., and Sagiv, L. (2008). "in-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science*, 19(4):405–411.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 449. ACM.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2):73–78.

Isaac, R. M. and Walker, J. M. (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Inquiry*, 26(4):585–608.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, 22:129–146.

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.

Jacobson, G. C. (2019). *Presidents and Parties in the Public Mind*. University of Chicago Press.

Klein, E. (2020). *Why We're Polarized*. Avid Reader Press / Simon & Schuster.

Kranton, R. E. and Sanders, S. G. (2017). Groupy versus non-groupy social preferences: Personality, region, and political party. *American Economic Review*, 107(5):65–69.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Lees, J. and Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, 4(3):279–286.

Lelkes, Y. and Westwood, S. J. (2017). The limits of partisan prejudice. *The Journal of Politics*, 79(2):485–501.

Levendusky, M. S. and Malhotra, N. (2016). (Mis) perceptions of partisan polarization in the American public. *Public Opinion Quarterly*, 80(S1):378–391.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3):482–493.

Madestam, A., Shoag, D., Veuger, S., and Yanagizawa-Drott, D. (2013). Do political protests matter? evidence from the tea party movement. *The Quarterly Journal of Economics*, 128(4):1633–1685.

Mason, L. (2015). “I disrespectfully agree”: The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1):128–145.

Mason, L. (2018). *Uncivil agreement: How politics became our identity.* University of Chicago Press.

Mazumder, S. (2018). The persistent effect of us civil rights protests on political attitudes. *American Journal of Political Science*, 62(4):922–935.

Meyer, D. S. (2004). Protest and political opportunities. *Annu. Rev. Sociol.*, 30:125–145.

Michelitch, K. (2015). Does electoral competition exacerbate interethnic or interpartisan economic discrimination? evidence from a field experiment in market price bargaining. *The American Political Science Review*, 109(1):43.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for experimental economics.* Palgrave.

Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B., and Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences.*

Müller, K. and Schwarz, C. (2019). From hashtag to hate crime: Twitter and anti-minority sentiment. Working Paper Available at SSRN: https://dx.doi.org/10.2139/ssrn.3149103.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, 46(8):1423–1437.

Orr, L. V. and Huber, G. A. (2020). The policy basis of measured partisan animosity in the united states. *American Journal of Political Science*, 64(3):569–586.

Reich, M. (2017). *Racial inequality: A political-economic analysis.* Princeton University Press.

Ruggeri, K. et al. (2020). The general fault in our fault lines. Working Paper Available at: https://doi.org/10.31219/osf.io/xvksa.

Schwalbe, M. C., Cohen, G. L., and Ross, L. D. (2020). The objectivity illusion and voter polarization in the 2016 presidential election. *Proceedings of the National Academy of Sciences.*

Shi, F., Teplitskiy, M., Duede, E., and Evans, J. A. (2019). The wisdom of polarized crowds. *Nature Human Behaviour*, 3(4):329–336.

Tajfel, H. and Turner, J. C. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*, 33(47):74.

West, E. A. and Iyengar, S. (2020). Partisanship as a social identity: Implications for polarization. *Political Behavior*, pages 1–32.

Westwood, S. J., Peterson, E., and Lelkes, Y. (2019). Are there still limits on partisan prejudice? *Public Opinion Quarterly*, 83(3):584–597.

Yamagishi, T. and Mifune, N. (2009). Social exchange and solidarity: in-group love or out-group hate? *Evolution and Human Behavior*, 30(4):229–237.

Main Appendix

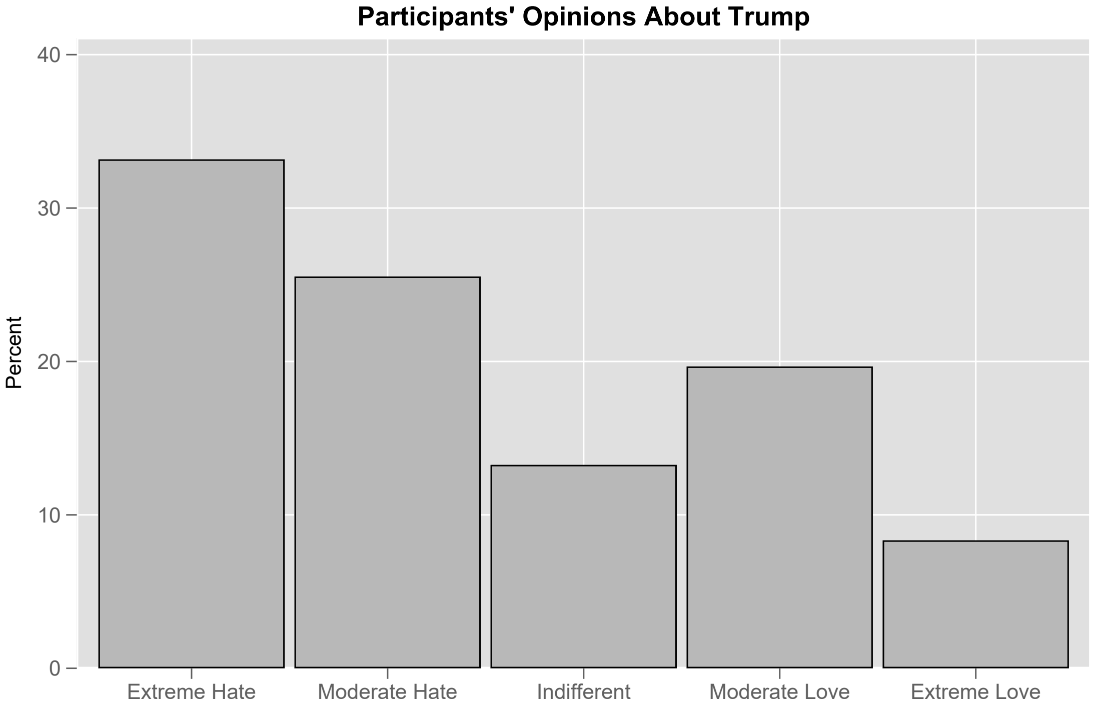


Figure A.1: Histogram of Trump opinions for both TP and MGP treatments.

Table A.1: OLS Regression Analysis of T-o-G Dictator Game Behavior

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Matched with Aligned	38.491***	31.762***	32.752***			
Trump Preference	(5.593)	(6.537)	(6.927)			
Closeness Score (%)		0.176*	0.182*		0.222***	0.244***
		(0.091)	(0.098)		(0.078)	(0.081)
Loves Trump				-32.514***	-23.298**	-17.790*
				(8.837)	(9.316)	(10.428)
Matched with Unknown				-15.075**	-7.334	-5.913
				(6.551)	(6.999)	(7.127)
Matched with Trump Lover				-36.979***	-28.519***	-29.041***
				(6.876)	(7.423)	(7.599)
Loves Trump ×				53.546***	44.381***	43.285***
Matched with Unknown				(12.054)	(12.269)	(12.441)
Loves Trump ×				78.557***	61.561***	58.618***
Matched with Trump Lover				(11.819)	(13.191)	(13.500)
Constant	-23.514***	-30.548***	-35.416	12.083***	-5.472	-31.785
	(4.398)	(5.623)	(36.886)	(4.302)	(7.286)	(30.163)
Controls	No	No	Yes	No	No	Yes
Observations	423	423	416	598	598	586

DV: Dictator game behavior (neg. = taking; pos. = giving). Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: OLS Regression Analysis of PGG Contribution Behavior

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Matched with Aligned	13.427***	4.355	5.281			
Trump Preference	(3.815)	(4.412)	(4.506)			
Closeness Score (%)		0.250***	0.196***		0.217***	0.156***
		(0.064)	(0.065)		(0.053)	(0.054)
Loves Trump				-10.873**	-3.703	-8.452
				(5.336)	(5.392)	(6.307)
Matched with Unknown				-7.643	-1.019	-4.873
				(4.870)	(5.081)	(5.175)
Matched with Trump Lover				-17.200***	-8.490	-10.530*
				(4.974)	(5.453)	(5.490)
Loves Trump ×				22.953***	12.896*	14.058*
Matched with Unknown				(7.593)	(7.751)	(7.897)
Loves Trump ×				23.904***	8.696	10.640
Matched with Trump Lover				(7.662)	(8.359)	(8.216)
Constant	51.029***	39.658***	-40.529**	65.748***	47.743***	-11.477
	(2.805)	(4.132)	(16.652)	(3.287)	(5.711)	(13.972)
Controls	No	No	Yes	No	No	Yes
Observations	388	388	373	537	537	519

Dependent variable is a participants contribution (%) to the PGG. Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix to "Hate Trumps Love: The Role of Rising Political Polarization in Social Preferences"

Eugen Dimant

Contents

I. Experiment 1: Additional Results and Robustness Checks

I.a. Main Experimental Conditions

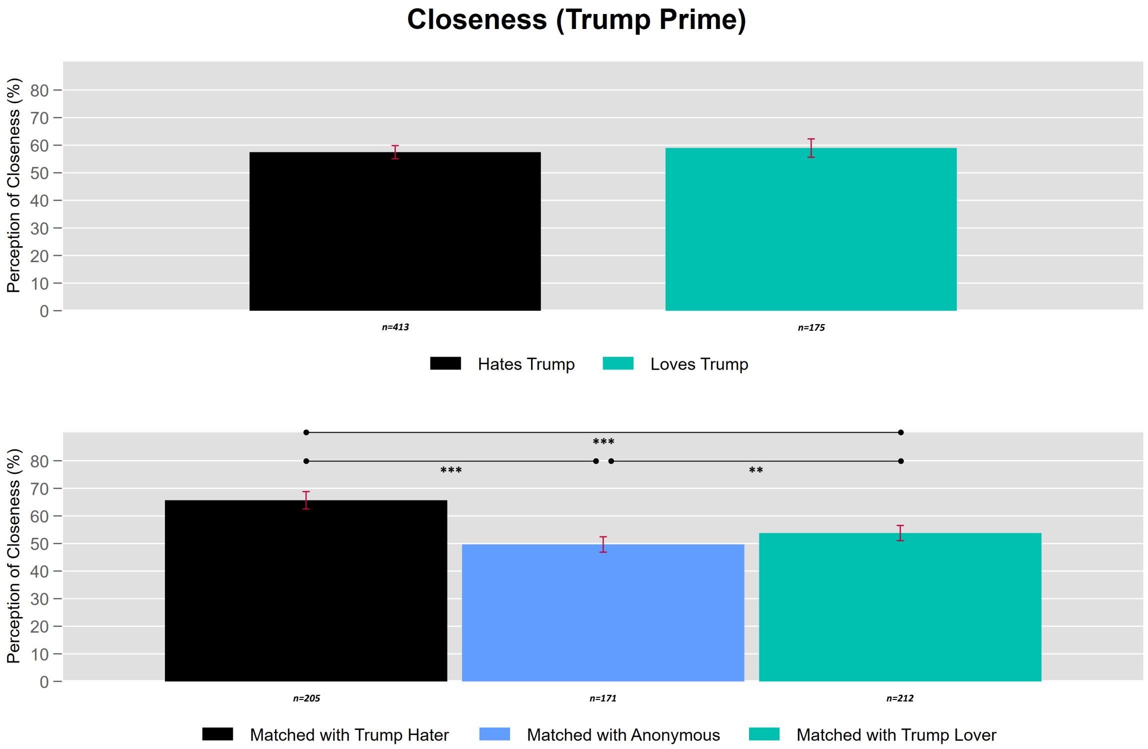


Figure OA.1: Closeness broken down by own opinion about Trump and being matched based on the partner's opinion about Trump. All adjacent bars are compared. Absence of significance stars ⇒ p-values > 0.05.

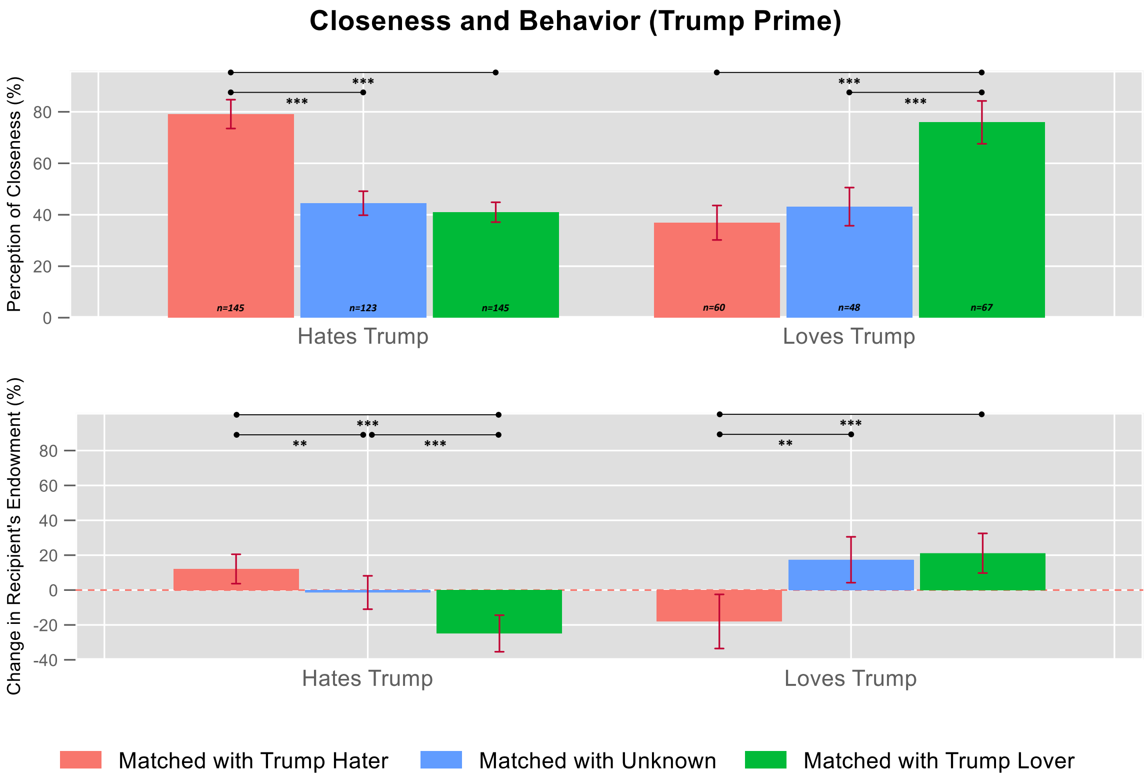


Figure OA.2: Closeness and behavior broken down by own opinion about Trump and being matched based on the partner's opinion about Trump. All adjacent bars are compared. Absence of significance stars ⇒ p-values > 0.05.

I.b. Minimal Group Paradigm Conditions

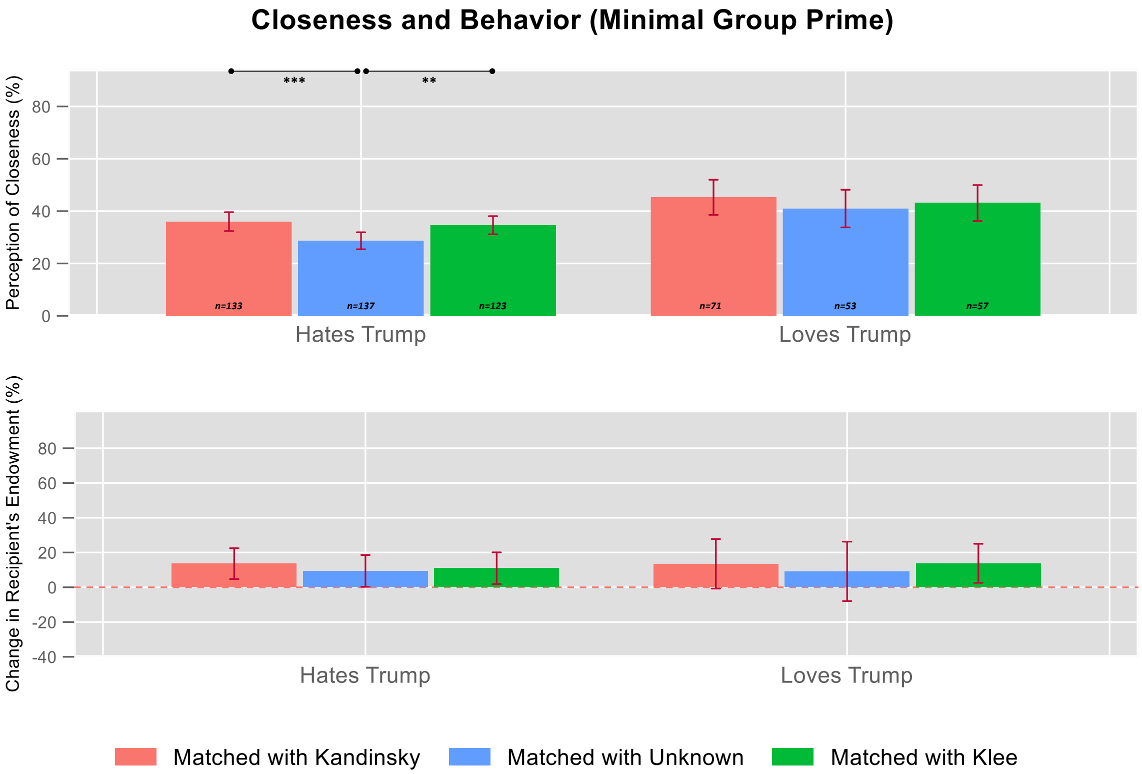


Figure OA.3: Closeness and behavior broken down by own opinion about Trump and being matched based the partner's painting preference. All adjacent bars are compared. Absence of significance stars ⇒ p-values > 0.05.

II. Experiment 2: Additional Results and Robustness Checks

Main Experimental Conditions

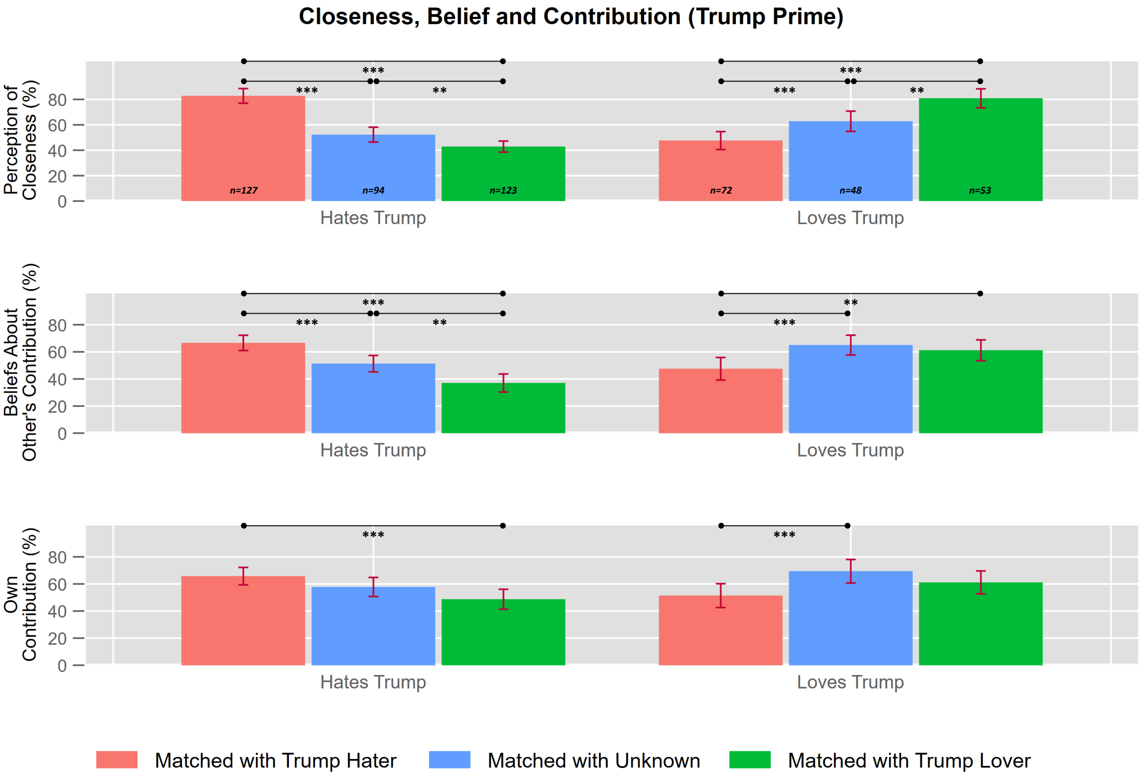


Figure OA.4: Closeness, belief, and behavior broken down by own opinion about and being matched based on the partner's opinion about Trump. All adjacent bars are compared. Absence of significance stars ⇒ p-values > 0.05.

III. Experiment 3: Additional Results and Robustness Checks

III.a. Dictator Game

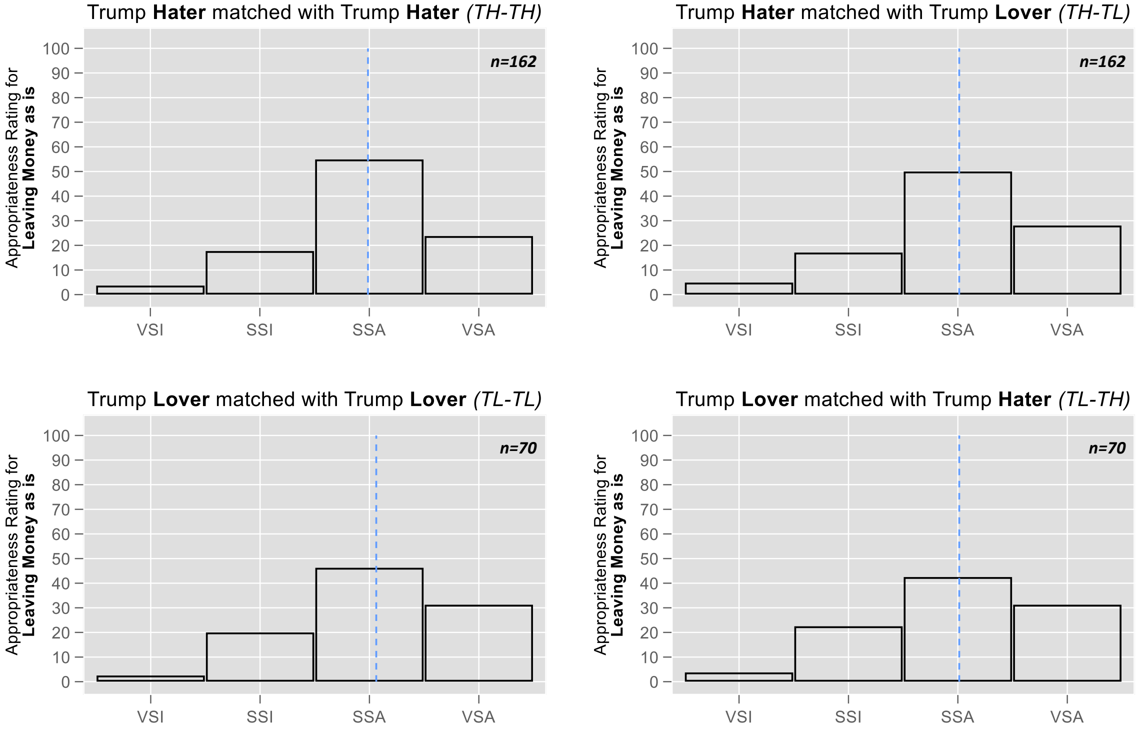


Figure OA.5: Norm perceptions for leaving initial split as is conditional on own and matched partner's Trump opinion. All adjacent quadrants are tested and statistical significance (if either ***p<0.01 or **p<0.05) is indicated where applicable. *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SSI), *Somewhat Socially Appropriate* (SSA), and *Very Socially Appropriate* (VSA).

III.b. Public Goods Game

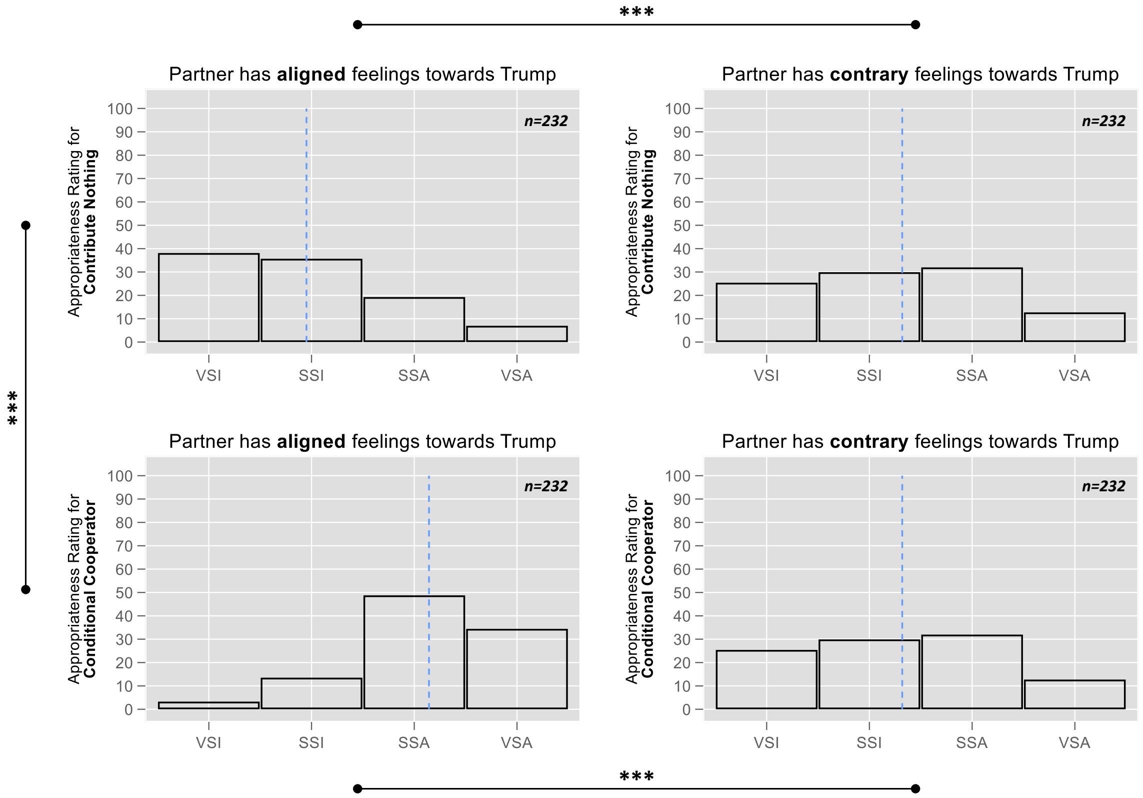


Figure OA.6: Norm perceptions for 'contribute nothing' Conditional Cooperators with partners who have aligned or contrary feelings towards Trump. All adjacent quadrants are tested and statistical significance (if either ***p<0.01 or **p<0.05) is indicated where applicable. *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SSI), *Somewhat Socially Appropriate* (SSA), and *Very Socially Appropriate* (VSA).

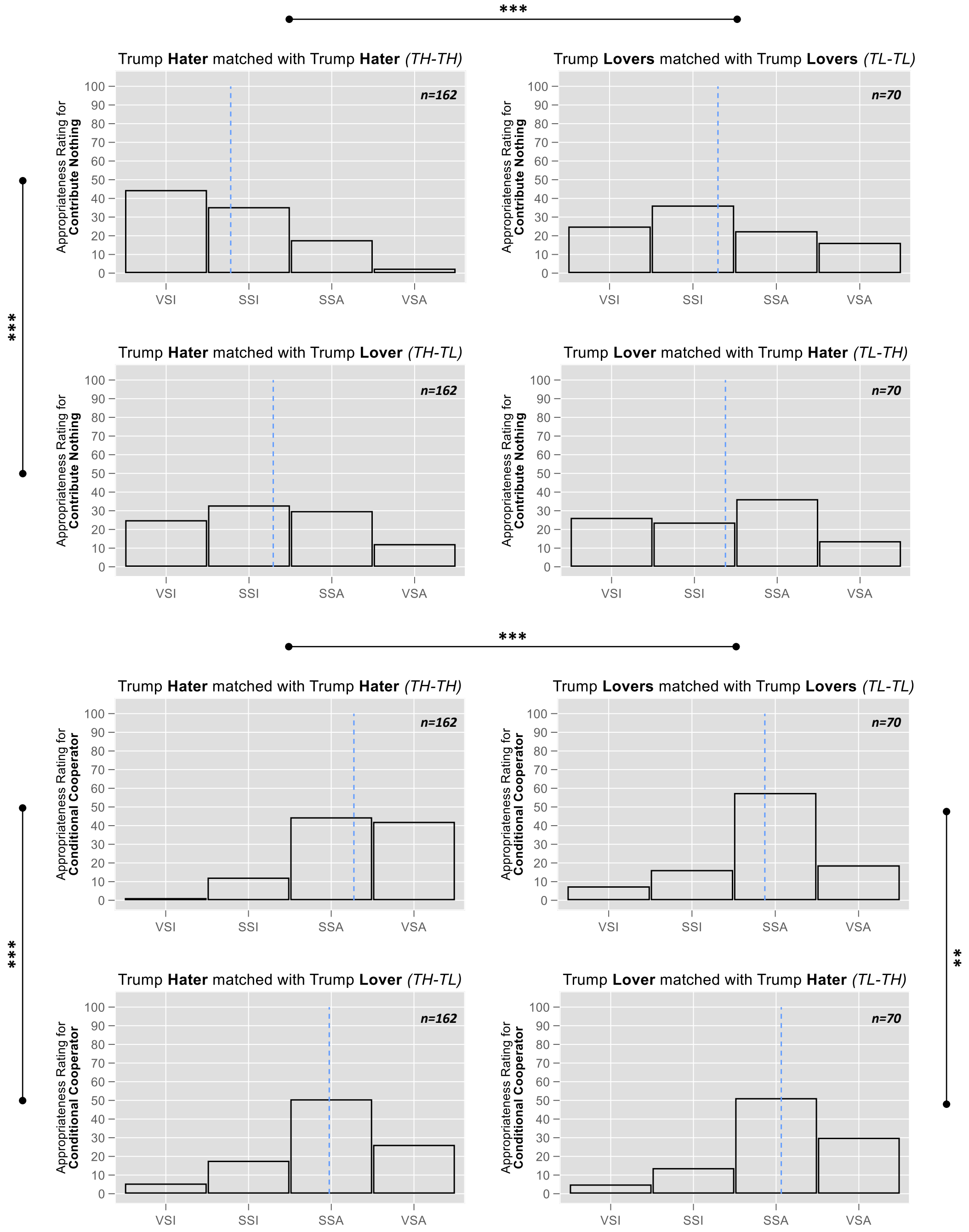


Figure OA.7: Norm perceptions for 'contribute nothing' Conditional Cooperators conditional on own and matched partner's Trump opinion. All adjacent quadrants are tested and statistical significance (if either ***p<0.01 or **p<0.05) is indicated where applicable. *Very Socially Inappropriate* (VSI), *Somewhat Socially Inappropriate* (SSI), *Somewhat Socially Appropriate* (SSA), and *Very Socially Appropriate* (VSA).

IV. Experimental Screenshots

All original experimental screenshots can be downloaded from: https://osf.io/tgku9/