Political Polarization and Selection in Representative Democracies *

Dominik Duell† and Justin Valasek‡

While scholars and pundits alike have expressed concern regarding the increasingly "tribal" nature of political identities, there has been little analysis of how this social polarization impacts political selection. In this paper, we incorporate social identity into a principal-agent model of political representation and characterize the impact of social polarization on voting behavior. We show that identity has an instrumental impact on voting, as voters anticipate that political representatives' ex post policy decisions have an in-group bias. We also conduct a laboratory experiment to test the main predictions of the theory. In contrast to existing work that suggests social polarization may have a positive impact by increasing participation, we show that social polarization causes political representatives to take policy decisions that diverge from the social optimum, and voters to select candidates with lower average quality.

Keywords: Social identity, political selection, political polarization.
JEL Codes: C92, D72.

*Thanks to Charlotte Cavaille, Steffen Huck, Rachel Kranton, Anselm Rink, Arturas Rozenas, and Paul Seabright for their valuable comments and suggestions. Support through the ANR - Labex IAST and WZB Berlin is also gratefully acknowledged.

†University of Essex.

‡Norwegian School of Economics (NHH), WZB Berlin, CESifo. Contact e-mails: dominik.duell@essex.ac.uk, justin.valasek@nhh.no

1 Introduction

While partisan identity has long been the strongest predictor of American political behavior, recent research indicates that partisan political identity is increasingly taking on a social dimension, resulting in an in-group/out-group mentality that is comparable in strength to racial identity (Iyengar and Westwood, 2015). This social polarization along partisan lines has caused pundits and academics alike to express concern that American politics has entered a new era of partisan tribalism. Such concerns are well-founded—a large body of literature has documented that social identity affects distributional preferences and political decisions (Charness, Rigotti and Rustichini, 2007; Chen and Li, 2009; Shayo, 2009; Klor and Shayo, 2010) and directly impacts voting behavior by creating an expressive preference (or even norm) to vote for in-group candidates (Green, Palmquist and Schickler, 2002; Bassi, Morton and Williams, 2011). In addition to an expressive effect, however, social polarization—defined as the existence of strong partisan identities—may also have an instrumental impact on voters' actions by influencing their beliefs regarding the partisan bias that political representatives will display once in office.

This instrumental effect can be clearly illustrated using the principal-agent model of political representation: Political representatives are often tasked with choosing between different policies that result in different distributions of voter payoffs. When facing such a choice, they could choose policies that are in the best interest of the general population, or policies that favor a particular group. Accordingly, in an environment where political representatives cannot commit to decisions prior to being elected, voters must anticipate how prospective representatives will choose between different policies before deciding whom to vote for; i.e. voters form beliefs over whether candidates will pursue a utilitarian objective, or instead choose policy to advance their identity group's specific interests. In such a situation, identity may play an important instrumental role in the voter's decision. Simply put, if a candidate's distributional preferences depend on their identity, e.g if candidates' distributional preferences are biased towards in-group voters, then identity will provide voters with an informative signal of that candidate's ex post policy choices, and voters will account for this in their voting decisions.

In this paper, we explore this instrumental impact of social polarization on partisan voting in a principal-agent model of political representation. First, we develop a formal theory that incorporates the group identity model of Charness, Rigotti and Rustichini (2007) into a simple principal-agent model of elections. In particular, we characterize the trade-off voters face between selecting political representatives based on identity and selecting based on quality (valence). Our framework focuses on detailing the instrumental impact of identity on voter behavior, and also informs our experimental strategy for distinguishing between the expressive and instrumental impact.[1] Importantly, our model shows that the instrumental impact of social polarization depends on the underlying degree of polarization in voters' policy preference. Intuitively, when there is little difference between the policy preferences of the different identity

[1] The existing literature describes, but does not experimentally separate cleanly, the expressive impact of social polarization on partisan voting and partisanship-guided political participation as driven by expressive motivations through rising partisan loyalty or increased concern for in-group status and welfare, and instrumental motivations to support policies and ideologies shared by those voters who identify with the same party (Hamlin and Jennings, 2011; Mason, 2015; Huddy, Mason and Aarøe, 2015).

groups, then there is little scope for representatives to bias policy in favor of in-group voters. Therefore, candidates will choose centrist policies when voters all have similar policy preferences, even when social polarization is high. Conversely, when policy preferences are strongly polarized and correlated with identity, then social polarization will lead representatives to select policies that favor in-group voters. If voters anticipate this interaction between social and policy polarization, then they will disproportionately vote for in-group candidates when policy polarization is high.

To test the theoretical predictions, we conduct a laboratory experiment. Our findings are highly consistent with the theory: Candidates consistently choose policy that is biased towards their partisan in-group, and voters display a willingness to vote for in-group candidates even when the out-group candidate has a higher valence. Importantly, we also find that the in-group bias of both candidates and voters is increasing in the degree of polarization in policy preference. This finding confirms that partisan in-group voting is more than an expressive phenomenon. Instead, the fact that voters respond to higher polarization of policy preferences indicates that they anticipate the increased bias of candidates' ex-post policy choices, and respond by voting in an increasingly partisan manner.

Our approach also allows us to measure the level of instrumental partisan voting. Since the expressive motive to vote for the in-group candidate does not increase with higher polarization of policy preferences, we can measure the instrumental impact of identity by comparing the degree of partisan voting when policy polarization is low to the degree of partisan voting when policy polarization is high. Using this approach, we find that when policy choice is equivalent to a zero-sum game between voters of the two partisan groups, then fifty-five percent of partisan, in-group voting can be attributed to the instrumental impact of social identity.

These findings have important implications regarding the impact of increasing social polarization on the electoral process: in contrast to existing studies, we show that the impact of social polarization in our setting is unambiguously socially harmful since it leads to policies that diverge from the social optimum and shifts the emphasis of political selection from selection based on quality to selection based on partisan identity.

Our theoretical framework considers voters who choose between two candidates via majority rule. Each candidate is characterized by membership in one of two identity-groups, and by a valence term that functions as a universal public good. Additionally, conditional on being elected, the candidate makes an ex post policy choice in a three-point policy space (left, center, right). Since candidates only receive benefits from holding office, their choice of policy depends only on their preferences over voters' payoffs: candidates can choose a centrist policy to maximize aggregate payoffs, or an extreme policy to favor a particular partisan group. Voters' payoffs are a function of both the valence of the winning candidate and the policy that the candidate sets when in office. Importantly, voters' policy preferences are partisan, in the sense that voters' group identities are correlated with the location of their ideal point in the policy space. Therefore, while all voters prefer a candidate with higher valence ceteris paribus, instrumentally, they will favor the co-partisan candidate to the extent that they expect candidates to select a partisan policy.

Our study also highlights the problem of separately identifying the expressive and instru-

mental impact of identity: if we empirically observe partisan voting, this could be due to either the expressive effect of identity as identified in Bassi, Morton and Williams (2011), the instrumental effect that we highlight in this paper, or a combination of the two. Therefore, to identify the instrumental link we rely on our models prediction that the degree to which social polarization influences voting behavior is a function of the magnitude of polarization in voters' policy preferences. Intuitively, when policy preferences are homogeneous, then voters will expect the candidates to adopt a centrist policy if elected and hence have a dominant instrumental incentive to vote for the higher-valence candidate. As policy preferences become more polarized, however, voters will expect candidates to take partisan policy positions to cater to the interests of their in-group, in which case partisan identity becomes a more important factor when selecting between the candidates. Again, since expressive motives to vote for the co-partisan candidate are not conditional on the underlying degree of polarization in voters' policy preferences,[2] the prediction of a positive correlation between policy polarization and partisan voting provides a clear test of the instrumental effect of social polarization.

While our model is a simplified setting, it captures important features of political competition that are affected by social identity and social polarization. Importantly, ex post policy discretion implies that voters face uncertainty regarding which policy the candidates will select once in office. Absent identity cues, voters expect all candidates to maximize aggregate utility and choose a centrist policy. In a setting with identity divisions, however, the group identity model predicts that candidates will favor policy positions that disproportionately benefit the in-group. This implies that voters will interpret identity cues as a signal that the co-partisan candidate will select policies that are consistent with the political values and norms of the group, and hence rationally respond to these cues by voting in a partisan manner.

In the experiment, both candidates and voters belong to one of two identity-groups. We induce social polarization either by a standard minimal-group intervention or by appealing to a pre-existing identity, both of which have been shown to result in group conflict and an in-group preference in a controlled experiment (Tajfel and Billig, 1974; Goette, Huffman and Meier, 2006; Chen and Li, 2009; Landa and Duell, 2015).[3] Additionally, identity groups correspond to voters' ideal points in the three-point policy space; specifically, we precisely control the degree of policy preference polarization by changing the degree of correlation between group membership and ideal policy points: in the case of no policy polarization, all voters have ideal points at the center; in the case of full polarization in voters' policy preference, one group is located at the left extreme while the other is at the right extreme.

Our experiment robustly confirms an instrumental impact of social polarization on voting behavior: as the degree of correlation between group identity and extreme ideal policy preference increases, voters increasingly vote for their co-partisan candidate. The positive rela-

[2] Another condition is required for identifying the instrumental impact; namely, the probability of being pivotal must be non-decreasing in degree of polarization of policy preferences. We show that this condition is satisfied in a robustness check introduced after our main analysis of the experimental results.

[3] Throughout this paper we refer to *group identity* and *social identity* interchangeably. We acknowledge that the former only requires individuals subjective awareness of group membership but may not rise to the level of being a social identity while the latter goes beyond awareness of membership and demands the individual to attach value and emotional significance to the membership (Tajfel, 1981).

tionship between policy polarization and partisan voting allows us to conclude that in a world where identities are correlated with political preferences, in-group voting is more than just an expressive phenomena—the strategic response in voting behavior to the underlying degree of policy polarization shows that voters anticipate candidates' bias in policy choices, and respond to changes in the fundamental elements of the political competition (i.e. the degree of policy polarization) by adjusting their voting behavior to account for the degree of in-group favoritism displayed by the candidates.

Our paper makes several important contributions to existing literatures. Generally, we build on the pioneering work by Turner and Brown (1978) studying the effect of minimal groups on behavior, and the incorporation of social and group identity into formal models of choice by Akerlof and Kranton (2000) and Charness, Rigotti and Rustichini (2007). Most existing work on social identity in experimental economics has considered the direct effect of group membership on distributional preferences and cooperation (for example, Eckel and Grossman, 2005, Goette, Huffman and Meier, 2006, and Chen and Li, 2009). In contrast, we focus on the impact of group membership on the beliefs subjects hold about the choices made by others. We find that subjects rationally anticipate the impact of group membership on distributional preferences, and exhibit a willingness to pay to delegate agency to an in-group member.

In the realm of collective choice, Klor and Shayo (2010) and Bassi, Morton and Williams (2011) show that group identity influences subjects' voting behavior (also related, Tyran, 2004 shows that social norms may influence voting behavior). However, these papers consider voting over fixed policy alternatives—in contrast, we consider the impact of social identity in a principle-agent model of voting. This allows for an instrumental impact of social identity, as social identity impacts voter behavior by influencing their beliefs regarding the ex post actions of political representatives.

Lastly, within political science a large, influential literature has emerged studying the impact of partisan social identities on political behavior (Green, Palmquist and Schickler, 2002; Iyengar and Westwood, 2015; Huddy, Mason and Aarøe, 2015; Mason, 2015; Huddy, Bankert and Davies, 2018). This literature, however, has been largely silent on the impact of social polarization on welfare. In this paper, we formally model the impact of social polarization of voting behavior in a principle-agent model of political behavior and predict that social polarization will have a strictly negative impact on welfare. Additionally, in an experimental test we confirm the predictions of the theory and show that existing social polarization leads to a situation in which political representatives take policy decisions that diverge from the social optimum, and voters select candidates with lower average quality.

2 A Principal-Agent Model of Elections with Social Identity

Here we introduce a simple formal structure that reflects our experimental design, and allows us to detail identity-contingent (partisan) political behavior when the electorate is socially polarized. While the model informs our experimental strategy for separately identifying the expressive and instrumental impact of social polarization on partisan voting, readers may also skip straight to Section 2.4 for an overview of the theoretical findings.

For clarity, we begin with an overview of the basic structure of the model before presenting the model in detail: Voters and candidates belong to one of two identity groups. Voters vote for one of two candidates, who belong to different identity groups and have individual valence terms that are publicly observable. The winning candidate then implements a policy that results in a distribution of payoffs that may be skewed towards one of the two identity groups.

Importantly, candidates chose policy ex post and do not receive policy payoffs. That is, candidates' monetary payoffs do not depend on their choice of policy; however, candidates' have preferences over the distribution of voters' policy payoffs. Accordingly, they implement the policy that maximizes their distributional preferences over voter payoffs. We consider the predictions of the model given two different assumptions regarding the candidates' distributional preferences: in the *Benchmark* model, candidates choose policy to maximize aggregate welfare; and in the *Identity* model, candidates' distributional preferences are skewed towards their respective in-group.

Voters only value their own payoffs and vote for the candidate that will maximize their individual payoffs.[4] Therefore, as we will establish below, voters preferences over the candidates will depend on the candidates' relative valence, the degree to which the candidates' distributional preferences are biased toward their in-group, and the degree of polarization between voters' policy preferences, as the latter will also influence the policy chosen the winning candidate.

Agents: There are n agents, denoted by the index set $N = \{1, ..., n\}$, with n even and greater than four. Agents either belong to (identity) group A or group B. Take $I_i \in \{A, B\}$ to be the identity group of agent i. Abusing notation, we will define group membership from the perspective of agent i when convenient; that is, i is a member of the in-group, denoted by set $I = \{j | j \in A \text{ if } i \in A \text{ else } j \in B\}$, while all other agents, j, are either in I or the out-group, denoted by set $I^- = N \backslash I$. Each identity group has an equal number of agents ($|A| = |B|$).

Actions and Payoffs: One agent in each group is a candidate; we denote these individuals by c^A and c^B. In addition to group membership, each candidate is endowed with a valence term denoted by α^A and α^B. After the election, the winning candidate implements a vector of policy choices, $\mathbf{p} = (p_l, p_m, p_r)$, over an ordered three-point policy space $\{l, m, r\}$. Each policy choice, represented by p_k for $k \in \{l, m, r\}$, consists of a value between zero and one ($p_k \in [0, 1]$) and available policy choices are constrained to the set of $\mathbf{p}$ that satisfy $p_l + p_m + p_r \leq 1$.

Agents who are not candidates are voters and, after observing the candidates' group membership and valence, submit a vote, v_i, for c^A or c^B (no abstention). The winner is chosen by majority rule with a random tie-breaking rule (50/50), and the winner affects voters' payoffs through the following two channels:

1. [Policy] voter i's payoffs are a function of the policy choice of the winning candidate, $\mathbf{p}^w$, and the voter's ideal point $p_i \in \{l, m, r\}$.
2. [Valence] voter i's payoffs are strictly increasing in the winning candidate's valence, α^w.

[4] We do not include expressive payoffs in the model, and refer interested readers to Duell and Valasek (2018) for a version of the paper that explicitly models expressive payoffs.

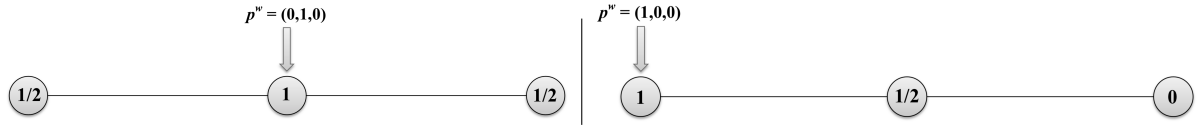


Figure 1: Payoffs as a function of the voters' ideal point (payoffs for each ideal point are listed within the circle) when the winning candidate allocates all policy to the middle point (left figure), and allocates all policy to the left point (right figure).

Formally, voter payoffs are as follows:

$$x_i^v = \alpha^w + v(\mathbf{p}^w, p_i), \tag{1}$$

where p_i is the ideal point of voter i.

Policy payoffs are analogous to a simple linear loss function of the ideal point and the amount allocated to each p_k; voters receive 1 unit of payoff for every unit of p_k placed at their ideal point, and 1/2 units of payoff for every unit of p_k placed at a point next to their ideal point (see Figure 1 for an illustration). Formally:

$$v(\mathbf{p}^w, p_i) = \sum^{k} v_k(p_k, p_i), \tag{2}$$

where:

$$v_k(p_k, p_i) = \begin{cases} p_k & \text{if } k = p_i, \\ \frac{1}{2}p_k & \text{if } k, p_i \text{ contiguous}, \\ 0 & \text{if } k, p_i \text{ non-contiguous}. \end{cases}$$

Consistent with our motivation, we are concerned with the case where policy is partisan, in the sense that policy preferences are correlated with identity. For simplicity, we consider the case in which policy preferences are stochastic: voters' ideal points are unknown prior to the election, but the distribution from which ideal points are drawn is common knowledge. Formally, for voter $i \in A$, p_i is drawn from $\{l, m\}$ and for $i \in B$, p_i is drawn from $\{m, r\}$. Additionally:

$$\Pr(p_i = l | i \in A) = \Pr(p_i = r | i \in B) = q. \tag{3}$$

This structure implies that each voter in group A (B) has the same *expected* policy position (ex ante symmetry). The assumption of stochastic policy preferences is not substantive with respect to the formal model; however, it simplifies the experimental analysis substantially, since group membership correlates perfectly with expected policy preferences. Moreover, q provides a measure of the polarization of voters' policy preferences, where $q = 1$ corresponds to perfect polarization in policy preferences.

Candidates, on the other hand, are purely office motivated and only receive payoffs based on whether they win the election. Candidates receive payoffs $x_i = x^w$ if they win the election

and $x_i = x^l$ if they lose the election, where $x^l < x^w$; i.e.:

$$x_i^c = \begin{cases} x^w & \text{if } i \text{ wins election,} \\ x^l & \text{if } i \text{ loses election.} \end{cases}$$

Importantly, to isolate the channel of distributional preferences and candidate identity on voters' decisions, we assume that candidates *do not* receive policy payoffs. The qualitative predictions of the model would be similar if candidates received policy payoffs in addition to being biased towards in-group members; i.e. our main result that partisan voting increases with policy polarization (q) is robust to the assumption of candidates with policy preferences. However, in this case, is would be difficult to empirically separate between the impact of the candidates' identity and the impact of the candidates' policy preferences on voters' voting decision. Therefore, by considering the case where candidates *do not* receive policy payoffs, we are able to directly identify the impact of identity.

Preferences: Voters' preferences are represented by a utility function that is linear in own payoffs:

$$u^v(x_i) = x_i.$$

Candidates, on the other hand, have social preferences over the distribution of payoffs. In particular, since in our setting candidates take policy decisions before citizens' ideal points are revealed, we assume that candidates have distributional preferences over ex ante payoffs. This is consistent with observed behavior in previous laboratory experiment; quoting Andreoni et al., 2018, "the most common behavioral pattern is for subjects to select the ex ante fair alternative ex ante ..." Take $E[x_i]$ to be the expected payoffs of voter i given the implemented policy, p^w:

$$E[x_i] = E_{p_i}[x_i | p^w, I_i]$$

Candidates' preferences are represented by the following utility function:

$$u^c(x_i, \{E[x_j]\}) = x_i + g(\mathbf{E}[\mathbf{x_j}]^I, \mathbf{E}[\mathbf{x_j}]^{I^-}),$$

where $\mathbf{E}[\mathbf{x_i}]^I$ represents the set of expected payoffs of voters in group I and $g(\mathbf{E}[\mathbf{x_i}]^I, \mathbf{E}[\mathbf{x_i}]^{I^-})$ represents the candidate's distributional preferences.

We make the assumption that voters only value own payoffs and that only candidates have distributional preferences to illustrate the qualitative predictions of the model as simply as possible—in the experiment subjects will take the role of both candidates and voters, and it therefore may be more natural to assume that both voters and candidates value own payoffs and have preferences over the distribution of payoffs. E.g.:

$$u(\{x_i\}) = x_i + \delta g(\mathbf{E}[\mathbf{x_i}]^I, \mathbf{E}[\mathbf{x_i}]^{I^-}), \tag{4}$$

with $\delta > 0$.

However, the qualitative predictions of the model are unchanged if both voters' and candi-

dates' preferences are represented by the same utility function as long as δ is not overly large: voters will still consider their own expected payoffs when choosing which candidate to vote for, and since candidates do not receive policy payoffs they will choose $\mathbf{p}^w$ to maximize their distributional preferences regardless of the size of δ.

We consider the predictions of the model under two models of distributional preferences: the *Benchmark* model where candidates choose policy to maximize aggregate welfare, and the *Identity* model where candidates' distributional preferences are skewed towards their respective in-group. In both models, we assume that distributional utility is a weakly concave function of others' expected payoffs to allow for fairness considerations (similar to the model of Cox, Friedman and Gjerstad, 2007 that allows for concavity in inequality aversion), in addition to in-group bias.

Benchmark Model: Given that our aim is to test the impact of group identity and social polarization, we must also define an appropriate benchmark for comparison. A natural candidate for distributional preferences is social efficiency: as highlighted in Charness and Rabin (2002), efficiency concerns can explain many experimental data. Therefore, to generate predictions regarding the behavior of candidates and voters in a world without social polarization, we also consider a benchmark case of distributional preferences for ex ante efficiency:

$$g(\mathbf{E}[\mathbf{x_i}]^I, \mathbf{E}[\mathbf{x_i}]^{I^-}) = \sum_{N-2} E[x_i]^{\frac{1}{m}}. \tag{5}$$

where $m \in [1, \infty)$—we use this family of utility functions to represent distributional preferences since it represents a broad range of strictly increasing and concave functions while also being tractable within the context of our model.[5]

Identity Model: Following the literature on minimal groups in social psychology, Charness, Rigotti and Rustichini (2007); Chen and Li (2009); Goette, Huffman and Meier (2012) document that even minimal group frames can significantly skew distributional preferences to favor payoffs for in-group members. As in Chen and Li, we formalize the group identity model by allowing for distributional preferences that are a function of group membership:

$$g(\mathbf{E}[\mathbf{x_i}]^I, \mathbf{E}[\mathbf{x_i}]^{I^-}) = \lambda \sum_{i \in I} E[x_i]^{\frac{1}{m}} + (1-\lambda) \sum_{i \in I^-} E[x_i]^{\frac{1}{m}}, \tag{6}$$

where $\lambda \in [0.5, 1]$ is a measure of the degree of social polarization, and $m \in [1, \infty)$ again. In what follows, we will characterize the predictions under both the *Identity* and *Benchmark* (efficiency) models.

Timing: The timing of the game is as follows

1. Candidates c^A, c^B are drawn and their abilities, $\{\alpha^A, \alpha^B\}$, are publicly revealed.
2. Voters simultaneously submit votes, v_i, for c^A or c^B.

[5] As we detail in the Appendix in the proof of Proposition 1 (introduced below), this family of utility functions allows us to derive a closed-form solution for $E[x_i]$ given the best reply of the winning candidate—other utility functions do not yield a closed-form solution, which renders part of our analysis intractable.

3. The winning candidate (by simple majority) chooses $\mathbf{p}^w$.
4. Voter policy preferences, p_i, are drawn and payoffs, $\{x_i\}$, realized.

Equilibrium and Welfare: The equilibrium concept is SPNE.[6] That is, an equilibrium, $\{\mathbf{v}; \mathbf{p}^A, \mathbf{p}^B\}$, maximizes the candidates' distributional preferences and, given $\{\alpha^A, \alpha^B\}$ and $\{\mathbf{p}^A, \mathbf{p}^B\}$, v_i maximizes $E_{p_i}[u^v(x_i)|\alpha_w, \mathbf{p}_w]$ for each i. We impose the selection criteria that, when they are indifferent, candidates choose a centrist policy and voters vote for their co-partisan candidate; these assumptions are for convenience only, and are not substantive. We consider a social efficiency benchmark mirroring the candidates' distributional preferences: i.e. the first-best solution maximizes $\sum_{N-2} E[x_i]^{1/m}$.

2.1 Analysis

We begin by characterizing the outcome that maximizes social efficiency.

Lemma 1 (Efficiency)
Social efficiency is maximized when candidates choose centrist policies, $\mathbf{p}^A = \mathbf{p}^B = \{0, 1, 0\}$, *and all voters vote for the highest-valence candidate,* $v_i = c^j$ *for all voters* i *if and only if* $\alpha_j \geq \alpha_{j'}$ *for* $j, j' \in \{A, B\}$.

First, note that a centrist policy maximizes aggregate expected monetary payoffs for any q, since a unit of policy allocated to the partisan extreme increases the expected payoffs of the in-group by $q - 1/2$, but decreases the expected payoffs of the out-group by $1/2 \geq q - 1/2$. Second, since a centrist policy results in an equal distribution of expected payoff, a centrist policy results in strictly higher social efficiency for all $m > 1$ (and weakly higher for $m = 1$). Lastly, given that both candidates choose the same policy, efficiency is maximized by selecting the candidate with the highest valence (social efficiency is always neutral with respect to candidate payoffs). Formal proofs of all results can be found in the appendix.

2.2 Candidates' policy choices:

Following backward induction, we begin with the candidates choice of policy, $\{\mathbf{p}^A, \mathbf{p}^B\}$. Since candidates choose policy after they are elected, the chosen policy has no direct or indirect impact on the candidates' probability of winning the election. Moreover, since candidates do not have policy preferences, the winning candidate will choose $\mathbf{p}^w$ to maximize their preferences of the distribution of *voters'* payoffs.

The following propositions partially characterize the equilibrium choices of $\{\mathbf{p}^A, \mathbf{p}^B\}$ under the Benchmark and Identity models.

Lemma 2 (Policy choices: Benchmark model)
If agents' distributional preferences are characterized by efficiency *then both candidates will choose centrist policies in equilibrium,* $\mathbf{p}^A = \mathbf{p}^B = \{0, 1, 0\}$.

Note that Lemma 2 follows directly from Lemma 1.

[6] While p_i is unknown to agents, the game is equivalent to a game of complete information since agents' maximize expected payoffs and p_i is not drawn until the final stage of the game.

Lemma 3 (Policy choices: Identity model)
If agents' distributional preferences are characterized by group identity*, then both candidates choose policies that are weakly asymmetric, in the sense that* $p^l \geq p^r$ *for* $\mathbf{p}^A$ *and* $p^l \leq p^r$ *for* $\mathbf{p}^B$*.*

Lemma 3 stems from the inter-group conflict over the partisan policy space: Under the group identity model, candidates put a higher weight on the payoffs of their group members, and hence will take policy decisions that favor the partisan position of their group. That is, while under the Benchmark model the candidate's group identity is irrelevant and the only distinguishing characteristic is their relative valence, under the Identity model group identity is an important predictor of the decisions the candidates will take when in office.

The next result details the comparative statics of the candidates' policy choices in the Identity model, and will be key to our strategy for identifying the instrumental impact of social polarization. However, instead of detailing the specific policy choices of the candidates, it will be more helpful to characterize the expected policy payoffs of the voters given the equilibrium policies $\{\mathbf{p}^A, \mathbf{p}^B\}$. Accordingly, we define Δ^x as the difference in expected policy payoffs between the two candidates for a voter with group identity I:

$$\Delta^x = E[x_i|\mathbf{p}^I] - E[x_i|\mathbf{p}^{I^-}]$$

Note that Δ^x is well-defined since the equilibrium policies of the candidates are uniquely defined given λ and q.

This definition allows us to formulate the following proposition:

Proposition 1 (Comparative statics of the Identity model)
(i) For $\lambda = 0.5$*,* $\Delta^x = 0$ *for all* q*.*
(ii) For $\lambda > 0.5$*,* $\Delta^x \geq 0$*:*

1. Δ^x *is increasing in the degree of policy polarization.*
2. Δ^x *is increasing in the degree of social polarization.*

Proposition 1 shows that the difference in expected payoffs is weakly increasing in both the degree of social polarization and the degree of polarization in policy preferences. The comparative statics of the model with respect to q and λ are also illustrated in Figure 2. Note that we do not formulate Proposition 1 in terms of the partial derivative with respect to q and λ. This is due to the fact that with $m = 1$, Δ^x is not a continuous function of q and λ. Instead, linear utility predicts that candidates allocating all policy at the middle for low levels of q, λ, and switch to allocating all policy to the partisan extreme for higher levels of q, λ.[7] Therefore, as illustrated in the right graph of Figure 2 there is a discrete upward jump in Δ^x at the point where candidates switch from centrist to extreme policy. However, the main comparative static predictions are similar with $m = 1$: Δ^x is increasing in λ since the switching point shift left for higher levels of λ, and Δ^x is increasing in q both due to the discontinuity, and due to the increasing probability that voters have an ideal point at the partisan extreme.

[7] Holding λ constant, this switching point can be characterized as $q^* = \frac{1}{2\lambda}$.

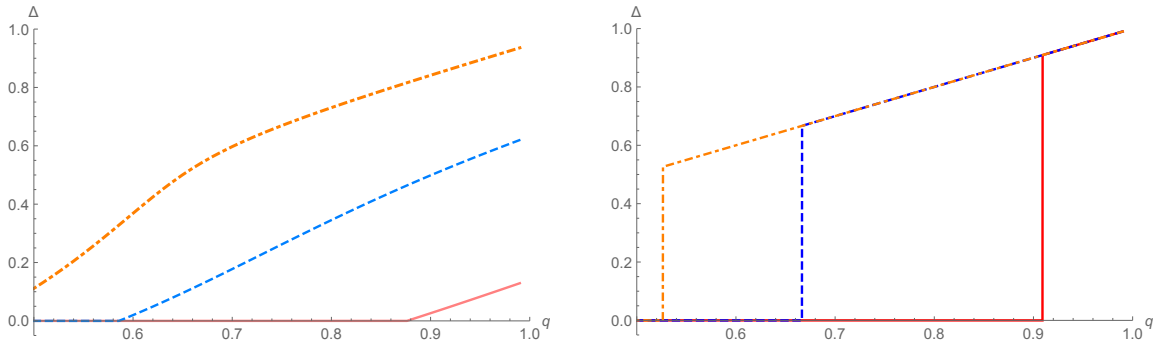


Figure 2: These graphs shows the relative policy payoffs for electing the in-group candidate (Δ^x) as a function of the polarization in policy preferences (q), for different values of λ: 0.95 (orange, dot-dashed), 0.75 (blue, dashed), and 0.55 (red, solid), for square root utility ($m = 2$; left graph) and linear utility ($m = 1$; right graph).

2.3 Voters' electoral decisions:

Having detailed candidates' equilibrium behavior under the two different behavioral models, we turn to the main object of interest: voters' decisions. Note that while voting, citizens only observe the candidates' identity and valence. While valence directly enters citizens' payoff functions, identity only enters indirectly through the correlation between identity and policy choices. Therefore, the relationship between identity and distributional preferences is the key component used by citizens to form beliefs over the ex post policy decisions of the candidates.

Take Δ^α to be equal to the relative valence of the *out-group* candidate:

$$\Delta^\alpha = \alpha^{I^-} - \alpha^I.$$

Note that Δ^x is calculated as in-group payoffs minus out-group payoffs, while Δ^α is calculated as out-group valence minus in-group valence; however, given this definition both Δ^x and the cutoff value of Δ^α at which voters prefer the out-group candidate (this cutoff value provides a measure of partisan voting and will be major element of our analysis) are both positive for $\lambda > 0$.

The following proposition characterizes voting behavior under the benchmark model.

Proposition 2 (Voting: Benchmark model)
If candidates' distributional preferences are characterized by efficiency, *then voters will vote for the in-group candidate if and only if:*

$$\Delta^\alpha \leq 0. \tag{7}$$

Since the candidates will both choose a centrist policy, voters maximize their payoffs by voting for the candidate with the highest valence.

As implied by Lemma 3 and as illustrated in the following proposition, the voting calculus becomes more complicated when agents have distributional preferences that favor the in-group.

Proposition 3 (Voting: Identity model)
If agents' distributional preferences are characterized by group identity*, then voters will vote for the in-group candidate if and only if:*

$$E[x_i|\mathbf{p}^I] - E[x_i|\mathbf{p}^{I^-}] = \Delta^x \geq \Delta^\alpha. \tag{8}$$

Proposition 3 shows that, relative to the Benchmark model, voters may have an instrumental incentive to vote for an in-group candidate with lower relative valence; we refer to this behavior—voting for an in-group candidate with lower relative valence—as *Partisan Voting*.

Next, we define $\tilde{\Delta}^\alpha$ as the maximum value of Δ^α such that voter i prefers to vote for the in-group candidate:

$$\tilde{\Delta}_i^\alpha = \begin{cases} 0 & \text{under Benchmark model,} \\ \Delta^x & \text{under Identity model.} \end{cases}$$

In the following corollary, we utilize this definition to compare the comparative statics of voting behavior under the Benchmark and Identity models.

Corollary 1 (Voting: comparative statics)
(i) Under the Benchmark model, $\tilde{\Delta}^\alpha$ is constant for all q.
(ii) Under the Identity model, $\tilde{\Delta}^\alpha$ is increasing in q.
(iii) For $q = 0$, $\tilde{\Delta}^\alpha = 0$ under both the Benchmark and Identity models.

The formal theory that we present in this section illustrates a novel insight regarding the relationship between social polarization and polarization in policy preferences. Namely, the impact of social polarization on partisan voting is a function of the degree of underlying polarization in policy preferences (see Corollary 1 (ii)). Importantly, this relationship is driven by the indirect, instrumental impact of social polarization: Intuitively, when policy preferences are homogeneous, there is little scope for choosing a policy that favors the in-group, and voters will expect the candidates to take a centrist policy if elected. Therefore, voters will prioritize the valence dimension when voting. As policy preferences polarize, however, voters will expect candidates to take partisan policy positions and partisan identity becomes the dominant concern when selecting between the candidates.

2.4 Hypotheses

Proposition 1 and Corollary 1 summarize the main theoretical findings that inform our empirical strategy for testing the suitability of the two models for explaining our experimental data. As we explain in detail below, our experimental setting closely replicates our theoretical model, and we gather data on both candidates' policy choices and voters' electoral decisions given different levels of policy polarization (q). Additionally, we attempt to vary the salience of social identity underlying the degree of social polarization by using different natural identities (λ).

Our theoretical predictions concern the candidates' choices regarding the relative payoffs of in-group voters versus out-group voters, Δ^x, and the maximum valence difference for which voters will vote for the in-group candidate, $\tilde{\Delta}^\alpha$. Beginning with the candidates' decision, the predictions for Δ^x are summarized in Table 1.

	Social polarization		
Policy	Benchmark model	Identity model	
Polarization	$\lambda = 0.5$	$\lambda = low$	$\lambda = high$
$q = 0$	$\Delta^x = 0$	$\Delta^x = 0$	$\Delta^x = 0$
$q = low$	$\Delta^x = 0$	$\Delta^x = low$	$\Delta^x = medium$
$q = high$	$\Delta^x = 0$	$\Delta^x = medium$	$\Delta^x = high$

Table 1: Qualitative predictions for candidates' policy choices under the Benchmark Model and Identity Model given different levels of policy polarization (q) and social polarization (λ). Relative levels (low, medium, high) are specified for purposes of illustration.

Table 1 follows directly from Proposition 1: When $q = 0$, candidates cannot allocate policy in a way that favors the in-group since all voters have the same ideal point, and therefore $\Delta^x = 0$ mechanically. Likewise, when $\lambda = 0.5$ (the benchmark model), $\Delta^x = 0$ since candidates maximize aggregate welfare and choose a centrist policy. If candidates' distributional preferences are skewed towards the in-group, however, then $\Delta^x \geq 0$ for $q > 0$, and Δ^x is increasing in q and λ.

For the candidates' choice of policy, we will consider the following two hypotheses to test the suitability of the identity model:

Hypothesis 1
Given $q > 0$, $\Delta^x > 0$.

The benchmark model predicts that $\Delta^x = 0$ for all q, while the identity model predicts $\Delta^x \geq 0$ for $q > 0$; therefore, the level of Δ^x when $q > 0$ provides evidence of whether candidates' preferences are better characterized by the benchmark or identity model.[8]

Hypothesis 2
Δ^x is increasing in q.

Additionally, since we are able to directly control q in our experiment, we test the second hypothesis derived from the identity model that Δ^x is increasing in q.

Lastly, on the candidate side, we consider the prediction that Δ^x is increasing in λ. Note that in our experimental setting, we will not be able to directly control the magnitude of λ, in the sense that λ is induced through the group manipulation rather than directly assigned. Instead, we rely on the comparison of minimal-group identities and naturally-occurring identities—arguably, subjects will be more biased towards in-group members (higher λ) if the subjects are sorted into groups based on salient natural identities. Therefore, the following hypothesis is not a test of the identity model per se; rather, it indirectly tests whether our experimental treatments are successful in increasing λ.

Hypothesis 3
Δ^x is higher with natural identities than with minimal-group identities.

[8]Note that this hypothesis essentially considers whether the results of earlier studies that have found an in-group bias, such as Chen and Chen (2011), replicate in our setting.

Next, we consider the predictions regarding voters' choices. Table 2 summarizes our theoretical predictions for partisan voting, $\tilde{\Delta}^{\alpha}$.

	Social polarization		
Policy	Benchmark model	Identity model	
Polarization	$\lambda = 0.5$	$\lambda = low$	$\lambda = high$
$q = 0$	$\tilde{\Delta}^{\alpha} = 0$	$\tilde{\Delta}^{\alpha} = 0$	$\tilde{\Delta}^{\alpha} = 0$
$q = low$	$\tilde{\Delta}^{\alpha} = 0$	$\tilde{\Delta}^{\alpha} = low$	$\tilde{\Delta}^{\alpha} = medium$
$q = high$	$\tilde{\Delta}^{\alpha} = 0$	$\tilde{\Delta}^{\alpha} = medium$	$\tilde{\Delta}^{\alpha} = high$

Table 2: Qualitative predictions for partisan voting under the Benchmark Model and Identity Model given different levels of policy polarization and social polarization.

Note that Table 2 replicates the predictions of Table 1, since the instrumental incentive to vote for the in-group candidate depends directly on Δ^{x}.

However, as discussed in the introduction, voters may also have an expressive motive to vote for the in-group candidate. Accordingly, our main hypothesis to distinguish whether subjects' voting decisions are consistent with the identity model is not that $\tilde{\Delta}^{\alpha} > 0$ for $q > 0$, since this result can be explained with a model of expressive payoffs. Instead, we consider the prediction that partisan voting is increasing in q, which cannot be explained by an constant expressive incentive.[9]

Hypothesis 4
$\tilde{\Delta}^{\alpha}$ is increasing in q.

Additionally, we also consider the prediction that partisan voting is increasing in λ. However, we are only able to test the following hypothesis if we are able to reject the null hypothesis that Δ^{x} is higher in the natural identity treatments relative to the minimal-group identity treatment (Hypothesis 3), which would indicate that our natural identity treatments were successful in inducing a higher λ.

Hypothesis 5
$\tilde{\Delta}^{\alpha}$ is increasing in λ.

Lastly, to preview our results, we do not find that our natural identity treatments result in stronger identities (which would have meant, arguably, a higher degree of social polarization). Therefore, we are unable to provide a direct test of Hypothesis 5. However, absent variation in social polarization from the natural identity treatments, we explore two alternative metrics for social polarization by comparing "strong" and "weak" partisans and analyzing the impact of multiple identities. These alternative measures provide suggestive evidence that partisan voting is increasing in social polarization (see Section B in the appendix, "Suggestive evidence on the impact of higher social polarization").

[9] In Section 4.4 following our main empirical analysis, we discuss the assumptions required for empirical identification of instrumental and expressive payoffs in detail and consider the empirical evidence.

3 Experimental Design

3.1 Protocol

Our experimental design largely mirrors the theoretical framework previewed above with the exception that, given that we use the strategy method, all voter and candidate choices are effectively simultaneous (in particular, candidate choices are taken before observing voter behavior). In our baseline identity treatment, the minimal-group *A vs B* treatment, subjects are randomly assigned into one of two groups of equal size, "Group A" or "Group B," after they receive instructions for the voting game but before the voting game commences. We exogenously induce social polarization by this standard minimal-group intervention because it has been shown to result in group conflict and an in-group preference (Tajfel and Billig, 1974; Goette, Huffman and Meier, 2006; Chen and Li, 2009; Landa and Duell, 2015) but also precludes that group membership is systematically correlated with other subject characteristics. Note that our minimal-group treatment does not actively promote a group identity, subjects are simply informed that they are assigned to Group A or Group B. Despite this very minimal intervention, however, we find that subjects systematically condition both their policy choices and voting choices on identity.

After subjects have been assigned to a group, the *voting game stage* implements the structure and payoffs as laid out in Section 2 and utilizes the strategy method. That is, each subject makes decisions in the role of a *candidate* and in the role of a *voter* for all potential distributions of voters' ideal points and, for voters, all possible combinations of candidate valence before election results are revealed.

As in the theoretical model, identity groups correspond to voters' ideal points in the three-point policy space. Specifically, we vary the degree of polarization in voters' policy preferences by changing the probability that voters will draw an ideal point at the partisan extreme. For example, in the case of no polarization in policy preferences ($q = 0$), all voters have ideal points at the center; in the case of full polarization ($q = 1$), subjects in one group have an ideal point at the left (Group A) while subjects in the other group have ideal point at the right (Group B).

Subjects take decisions for five different levels of policy polarization, $q = \{0, .25, .5, .75, 1\}$. All subjects know the value of q while making their decisions. As candidates, subjects chose how to allocate up to ten tokens to the three positions of a preference space, *Left*, *Center*, or *Right*. Candidates make one policy allocation for each q and do not observe their valence before making a policy allocation.

As voters, subjects make a choice between two candidates (one from each group) knowing both q and the valence of the candidates, but before drawing their individual ideal point. The valence of candidates is either low, medium, or high, which corresponds to a voter payoff (in tokens) of 2, 3 and 5. Subjects took decisions as voters for all possible valence combinations for each level of q. On the subject screens, the valence factor is referred to as "ability."

Subjects' decisions were organized into "blocks" for each level of q. At the beginning of each block q is announced. Subjects then make their allocation decision in the role of a candidate and then make 9 voting decisions between candidate pairs with varying valence. Both the order of decision blocks and order of the assigned candidate valence pairs within those blocks was randomized; therefore, subjects face decision environments in different orders. In total, subjects

make 5 decisions as candidate and 45 decisions as voter.[10] After the voting game, subjects also play one round of a dictator game and answer a questionnaire about basic demographics and the choices they made in the experiment.

Lastly, we conduct additional treatments using natural identities rather than minimal groups represents in an attempt to induce variation in social polarization (λ). Arguably, subjects will be more biased towards in-group members—increased social polarization—if the subjects are sorted into groups based on salient natural identities. (However, as we will detail in the results section, we do not find evidence of a higher λ in the natural identity treatments.) The *Bike vs Car*- and *Dem vs Rep*-treatments both feature endogenous sorting into groups based on natural identities. Importantly, in both the natural identities treatments, subjects are truthfully told that in the pool of subjects from which they were recruited to participate in the experiment, the distribution of subjects across the two groups is close to equal.

We conduct the experiment in the laboratories of Technical University Berlin and Florida State University. For the Florida sample, subjects were allocated to groups based on whether they reported to feel closer to Democrats or Republicans.[11] For the Berlin sample, subjects were allocated to groups based on whether they reported that they are more likely to use their bike or their car.[12] While this may seem a strange choice of an identity-group, a pre-experiment survey of the experimental pool in Berlin showed that students rank the car/bike-divide as being more important then religious or political affiliations, and that subjects are evenly split between car and bike.

To place the results from Florida in context, the sessions were run in June of 2016, shortly after Donald Trump had secured the nomination as the Republican presidential candidate. Given the contentious nature of the 2016 presidential election, we had the ex ante expectation that the level of in-group bias induced by subjects party identity would be significantly greater than a minimal-group group intervention.

3.2 Payoffs

Subjects are paid depending on their and other subjects' choices in one randomly chosen decision situation of the voting game (and their choices in the dictator game). That is, one subject from each identity group is chosen to be a candidate and is assigned a level of valence, and one distribution of voters' ideal points, q, is randomly selected. Next, the subjects' actual voting decisions for this value of q and pair of candidate valence are used to determine the winning candidate: the candidate with the largest vote share is the winning candidate and receives 15 tokens; the losing candidate receives 5 tokens.

Finally, the winning candidate's token allocation determines the payoffs of the subjects not chosen as candidates (voters): each subjects' ideal point was drawn randomly according to

[10] An exact overview over the decision environment for each round of the experiment can be found in Section C.1 in the appendix. We control for the order of the assignment to blocks in the analysis below by including *decision order* variables as appropriate.

[11] We ran this treatment at Florida since questionnaires showed that their subject pool was roughly evenly divided between Democrats and Republicans – we also informed subjects of this fact before they select a group.

[12] In German: "Sagen Sie uns bitte, ob Sie öfters Ihr Fahrrad oder Auto benutzen?

the distribution of ideal points selected to determine payoffs. Corresponding to the model, if assigned an extreme ideal point, A-voters (B-voters) receive 1 token for each 1 token allocated to Left (Right) and .5 tokens for each 1 token allocated to the Center.[13] Subjects assigned an ideal point of Center receive 1 token for each token allocated to center, and .5 tokens for each token allocated to Right or Left.

Additionally, all voters receive 2 tokens if the winning candidate has "low" valence, 3 tokens if "average" valence, and 5 tokens if "high" valence. Subjects also received a 5 Euro (7 Dollars) show-up fee, plus the tokens they earned at an exchange rate of 60 cents for 1 token.

3.3 Empirical strategy

The experiment presented above is designed to test the theoretical hypotheses specified in Section 2.4. It is set to identify whether identity has an instrumental impact on partisan voting and distinguish between the identity and benchmark models. As we discuss in the introduction, in-group bias in voting decisions can be due to either instrumental or expressive concerns. Therefore, our main identification strategy utilizes the within-subject variation in choices as a function of q—expressive payoffs are constant for all levels of q, while Proposition 1 shows that the instrumental impact of identity is increasing in q. We also examine the robustness of our empirical strategy to distinguish between the expressive and instrumental impact of identity in detail in Section 4.4 below.

We identify in-group bias in candidates' allocation decisions by the relative expected in-group token payoff that realizes given candidates' allocation decision (Δ^x). Since Δ^x measures the difference in payoffs to the in-group relative to the out-group, a positive value of Δ^x indicates that candidates' policy allocation biases payoffs to the in-group. We identify partisan voting at the individual level by the maximum difference in the token payoff from the valence of out-group and in-group candidate for which voters vote for the in-group candidate ($\tilde{\Delta}^\alpha$), top-coded at 3 for subjects who vote for the in-group candidate for all valence differences.[14]

Note that since candidates' policy choices do not impact their own payoffs, our experiment identifies the in-group bias in candidates' policy choices. Our experiment, however, does not specifically identify the mechanism for this in-group bias. In the theory section, we assume that—consistent with the findings of earlier experimental studies—candidates' distributional preferences are biased towards the payoffs of members of the in-group. An alternative explana-

[13] In the natural identity treatment, Democrat-voters and Bike-voters (Republican-voters and Car-voters) receive receive 1 token for each 1 token allocated to Left (Right) and .5 tokens for each 1 token allocated to the Center.

[14] In the appendix, we additionally report results associated with an alternative operationalization of ($\tilde{\Delta}^\alpha$) where we compute for each voter the minimum relative valence of the out-group candidate Δ^α at which the out-group candidate is still elected, the threshold limiting an in-group vote is then simply just below that minimum. We will refer to this measure as ($min(\Delta^\alpha)$, see Section D.4). Additionally, top-coding could obscure variance between our identity treatments if, say, more subjects have extreme in-group preferences in the natural identity treatment relative to the minimal-group treatment. This, however, does not appear to be the case, and the null findings regarding the additional impact of natural identities do not seem to stem from a ceiling effect: the proportion of subjects who vote for the in-group candidate across all valence differences is relatively similar in the minimal and natural identity treatments (in fact, more subjects systematically vote for the in-group candidate for the minimal-group treatment; see Figure D.3 in the Appendix).

tion for an in-group bias exists that is based on a reciprocity motive: irrespective of identity, candidates may reciprocate voter support by biasing policy in favor of their voters.[15] In our experiment, however, candidates do not observe votes prior to choosing policy, since we use the strategy method. Therefore, candidates cannot directly reward the individuals who voted for them. Instead, candidates may anticipate that voters will disproportionately vote for the in-group candidate, and may reciprocate 'in advance.' Note that identity is also key for this second mechanism—without identity to coordinate expectations, candidates would have no reason to systematically bias policy towards one of the two groups of voters. Therefore, while both mechanisms may contribute to candidates' in-group bias, we refer to the effect as an instrumental impact of identity.

For Hypothesis 1 we consider a simple difference in the average payoffs to in-group and out-group members that result from the candidate policy choices, pooled across q. We test Hypotheses 2-5, by estimating the following two linear models on the experimental data. The first model considers the impact of policy polarization, q, and natural identities, NI, on candidate i's partisan bias in allocation decisions:

$$\Delta_i^x = \beta^q \, q_i + \beta^{NI \times q} \, NI_i \times q_i + \beta^o \, o_i + \epsilon_i,$$

where o_i indicates the order in which decision situations (characterized by q) are presented to i. Note that this model does not feature an intercept since Δ^x is mechanically equal to zero when $q = 0$. Hypothesis 2 predicts a positive coefficient β^q, and Hypothesis 3 predicts an additional impact of natural identities ($\beta^{NI \times q} > 0$).

The second model considers the impact of policy polarization, q, on voter i's partisan bias in voting decisions:

$$\tilde{\Delta}_i^\alpha = \delta_0 + \delta^q \, q + \delta^{NI} \, NI + \delta^{NI \times q} \, NI \times q + \delta^o \, o_i + \epsilon_i.$$

Note that our second empirical model includes an intercept. As discussed above, voters might prefer to vote for the in-group candidate for expressive reasons—these expressive motives, however, are captured in δ_0 and δ^{NI}, since these coefficients estimate voters' willingness to vote for the in-group candidate when there is no possible bias in the candidates' policy allocations (again, when $q = 0$ policy payoffs are mechanically equal for all voters). Therefore, δ^q and $\delta^{NI \times q}$ cleanly identify the instrumental impact of identity on voters' decisions, and are used to test Hypotheses 4 and 5, respectively.

We estimate our two empirical models on the data from Berlin and Florida separately. We employ the ordinary least square estimator for all specifications of candidate and voter model, clustering errors at the subject level. The unit of analysis is the subject-q pair, that is for each subject i we observe five values of Δ_i^x and $\tilde{\Delta}_i^\alpha$, one value for each of the five realizations of policy polarization q.

[15] We thank Reviewer 1 for suggesting this alternative mechanism.

3.4 Session statistics

In 7 sessions, with 24 subjects each (one with 26), we collect, for each subject, 45 observations as voter and 5 observations as candidate. In total, we collect observations on 170 subjects with a total of 7650 voter-round and 850 candidate-round observations.[16] Given that subjects make decisions using the strategy method we have as many independent observations as subjects in the experiment. Subjects earning range from 7.7 to 20 Euro, average session earnings range from 12.9 to 18 Euros.[17] We ran 4 sessions in the laboratory at Technical University Berlin (2 sessions for the baseline *A vs B* treatment and 2 sessions for the *Bike* vs *Car*-treatment) and 3 sessions in the laboratory at Florida State University (1 session for the *A vs B* treatment and 2 sessions for the *Dem* vs *Rep*-treatment).[18]

4 Results

The main purpose of the experiment is to identify whether subjects' choices are consistent with the identity model and to identify the instrumental impact of social polarization on voters' identity-contingent voting. More specifically, we evaluate whether there is evidence to support the hypotheses laid out in Section 2.4 that distinguish between the identity and benchmark models. Following the order of our theoretical section, we first consider subjects' choices as candidates, followed by their decisions as voters.

Before we evaluate our specific theoretical predictions, we characterize subject behavior towards in-group and out-group. Overall, we find that behavior is highly dependent on the in-group/out-group designation. When in the role of a candidate, subjects make allocation decisions in a way that generats, on average, significantly higher expected token payoffs for in-group than out-group voters in the minimal group identity and natural identity treatments. This result holds true overall as well as in decision situations with $q = 1$, a situation where the candidate policy allocation decision is comparable to a "divide the dollar game." This also allows us to compare our results to earlier experimental findings: we find that our estimated in-group bias is a bit higher than the range found by Chen and Li (2009), who estimate an in-group bias of between $3.22 - 3.84$ (we normalize their results to 10 tokens). Additionally, as voters subjects are, on average, significantly more likely to vote for the in-group than the out-group candidate in all treatments. These results are listed in Table 3 below.[19]

[16] These numbers of observations result in 850 subject-level of policy polarization (q) observations relevant for the regression analysis of candidate and voter choices presented below.

[17] Table D.1 in the appendix provides an overview of these statistics.

[18] Additionally, Table D.2 in the appendix gives the summary statistics on *allocation decision*, Δ^x, *voting decision*, and $\tilde{\Delta}^\alpha$ by treatment.

[19] Whenever we report a p-values associated with the relevant one or two-sample t-test at a given level α we also check that the respective $1 - \alpha$ (subject-level clustered) bootstrapped confidence interval of the test statistic does not contain 0 and confirm our interpretation with the result from the appropriate difference-in-distribution test (clustered Wilcoxon sign rank or rank sum test). When we report p-values associated with the exact binomial test on vote choice, we check those against whether the respective $1 - \alpha$ confidence interval of differences in vote share in in-group and out-group, computed from bootstrapped vote choice (clustered at the subject-level), contains 0.

	A vs B			*Bike vs Car*			*Dem vs Rep*		
In-Group/Out-Group	In	Out	Diff.	In	Out	Diff.	In	Out	Diff.
Average expected	7.31	5.43	1.88***	7.59	5.37	2.22***	7.01	5.39	1.62***
policy payoff	(0.13)	(0.09)	(0.17)	(0.10)	(0.15)	(0.21)	(0.14)	(0.13)	(0.22)
Expected policy	7.38	2.47	4.91***	7.80	2.16	5.64***	6.73	2.83	3.90***
payoff $q = 1$	(0.27)	(0.25)	(0.32)	(0.50)	(0.31)	(0.63)	(0.31)	(0.28)	(0.54)
Average vote share	0.76	0.24	0.52***	0.76	0.24	0.52***	0.80	0.20	0.60***
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)

Standard errors clustered by subject

*Asterisk indicates difference in-group/out-group significant at: *p<0.1; **p<0.05; ***p<0.01*

Table 3: Mean of expected payoff from candidate allocation (overall and at $q = 1$) and candidate vote share. Standard errors in parentheses. Statistical significance of the differences between in-group and out-group are tested using a paired t-test/exact binomial test.

4.1 Candidate policy choices

First we consider whether, on average, candidates' policy choices are biased in the sense that they resulted in higher expected payoffs to in-group members (Hypothesis 1). As discussed above, the first row of Table 3 shows that candidates' policy decisions are systematically biased towards in-group voters' payoffs across all identity treatments. Moreover, the average difference between in-group and out-group payoffs are statistically significant at the one-percent level. These results provide evidence to support Hypothesis 1, and suggests that candidate choices are better explained by the identity model, since the benchmark model predicts that candidates will equalize the expected payoffs of the two groups of voters.

Summarizing,

Result 1

When policy polarization exists, candidate allocation decisions favor in-group voters; that is $\Delta^x > 0$ *when* $q > 0$.

4.1.1 The impact of policy polarization on candidates' choices:

Next, we consider the impact of policy polarization on the candidates' policy allocations. Figure 3 displays expected policy payoffs of in-group and out-group voters by the level of policy polarization, while Figure 4 shows Δ^x as a function of q. Both figures hint at support for Hypothesis 2 since the average level of Δ^x, the difference in expected payoff from candidates' allocations to in-group and out-group voters, is increasing in q. Two sample t-tests indicate a significantly higher payoff from the allocation of tokens to in-group than to out-group voters at every level of $q > 0$ with $p < .01$. In particular, at $q = .25$, the difference is $.34\,(.21, .47)$ and at $q = 1$ the

difference reaches $4.90\,(3.85, 5.91)$.

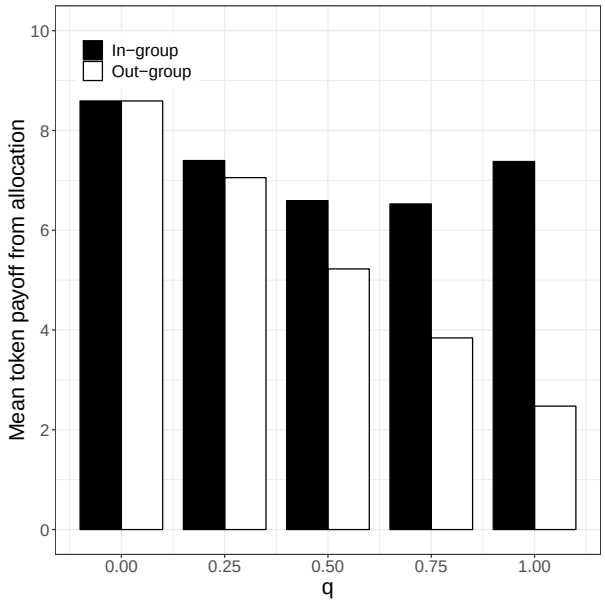


Figure 3: Average token payoff from the allocations by in-group (black) and out-group (white) candidates plotted over policy polarization (q) in the *A vs B* treatment.

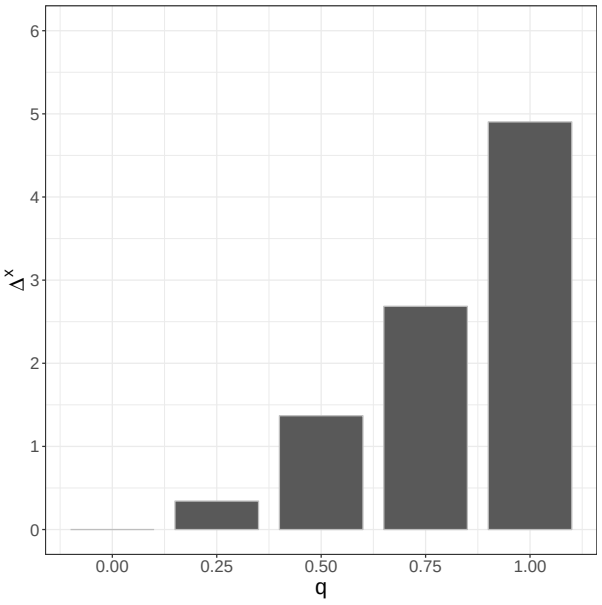


Figure 4: Difference in average token payoffs from the allocations by in-group and out-group candidates (Δ^x) in the *A vs B* treatment.

We investigate Hypothesis 2 more thoroughly by estimating our first empirical model. The results of the regression are presented in Table 4 below; our main specification is presented in columns (5) and (6), and we also estimate the model on the minimal-group and natural identity treatments separately for completeness (columns 1-4). The coefficient estimates of β^q in all treatments in both Berlin and Florida are all positive and significantly different from zero with $p < .01$. We find that for an increase in q from no policy polarization ($q = 0$) to perfect policy polarization ($q = 1$), the relative in-group token payoff from candidates' allocations increases by $4.20 - 4.58$ in the minimal group treatments.

These results strongly support Hypothesis 2:

Result 2

With rising policy polarization, candidates' policy choices increasingly favor in-group voters; that is, Δ^x is increasing in q.

4.1.2 The impact of natural identities on candidates' choices:

Lastly we explore whether our natural identity treatments resulted in a stronger in-group bias in candidates' policy choices (higher λ), relative to our minimal-group treatment. We hypothesize that natural identities will result in a higher level of social polarization, and hence a larger Δ^x at each level of policy polarization (q).

Surprisingly, we find no significant difference comparing mean of Δ^x in minimal group identity and natural identity treatments: the average Δ^x across all candidate choices with

$q > 0$ is 2.32 (1.93, 2.74) in the *A vs B* treatment and 2.40 (2.02, 2.74) in the natural identity treatments.[20]

Table 4: Relative in-group token payoff from candidate policy choices (Δ^x) regressed on policy polarization q (columns 1-6). Columns 5 and 6 list the estimates of a regression of Δ^x on q and the interaction of a dummy for the natural identity treatments (*Bike vs Car* in Berlin, *Dem vs Rep* in Florida) and q. We also include the decision order in all regressions.

	Dependent variable: Δ^x					
	A vs B		*Bike vs Car*	*Dem vs Rep*		
	Berlin	*Florida*	*Berlin*	*Florida*	*Berlin*	*Florida*
	(1)	(2)	(3)	(4)	(5)	(6)
q	4.52***	4.73***	4.78***	3.56***	4.62***	5.23***
	(0.54)	(0.79)	(0.56)	(0.51)	(0.56)	(0.85)
$q\times$ *Natural identity*					1.01	-1.29
					(0.80)	(1.01)
Observations	240	130	240	240	480	370
Subjects	48	26	48	48	96	74
Adjusted R^2	0.48	0.53	.57	.46	0.53	0.49
F-Statistic	38.30***	21.44***	65.02***	28.36***	62.30***	33.13***

Standard errors clustered by subject
$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Additionally, the identity model predicts a stronger impact of q on Δ^x with a higher degree of social polarization (see Figure 2 in Section 2 for an illustration). However, in the estimation of our first empirical model, presented in columns (5) and (6) of Table 4, the interaction term between q and the natural identity treatment is not significantly different from zero for either the Berlin or Florida samples.

Together, these findings show no support for Hypothesis 3:

Result 3
Candidate policy choices are not more favorable towards the in-group with natural identity than with minimal group identity; that is, we do not find evidence that social polarization, λ, is higher in the natural identity treatment.

4.2 Voter election decisions

Next, we turn to the main object of interest: the impact of social polarization on voters' voting decisions. As outlined in our theory section, the identity model predicts that in a setting with both policy polarization and social polarization, voters will anticipate the in-group bias of the

[20] See Figure D.1 in the appendix for more detail regarding the distributions of Δ^x across treatments.

candidates, and hence be willing to vote for in-group candidates with lower relative valence (partisan voting). Moreover, partisan voting is predicted to increase with the degree of policy polarization.

First we look at the summary statistics. Under perfect polarization ($q = 1$), we find a strong bias among voters to elect a candidate of their group: the average rate of in-group voting is .81 (.76, .85). Also, as to be expected, voters overwhelmingly vote for the in-group candidate whenever the valence of the out-group candidate is lower than the valence of the in-group candidate, for any level of policy polarization (the rate of "anti-partisan" voting is less than five percent across all elections). Figure 5 illustrates the average in-group vote share at different levels of Δ^{α} and q, and shows that the average rate of partisan voting is high for all levels of Δ^{α}.

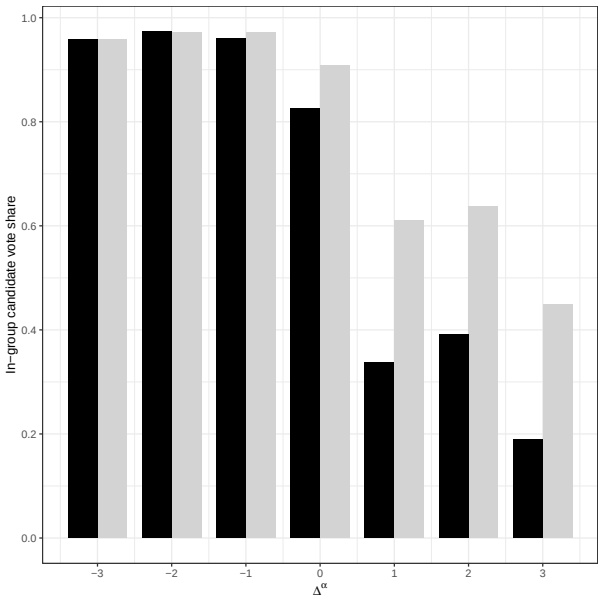


Figure 5: In-group candidate vote share plotted over relative payoff from the ability of the out-group candidate (Δ^{α}) for no ($q = 0$; black) and perfect policy polarization ($q = 1$; grey) in the *A vs B* treatment.

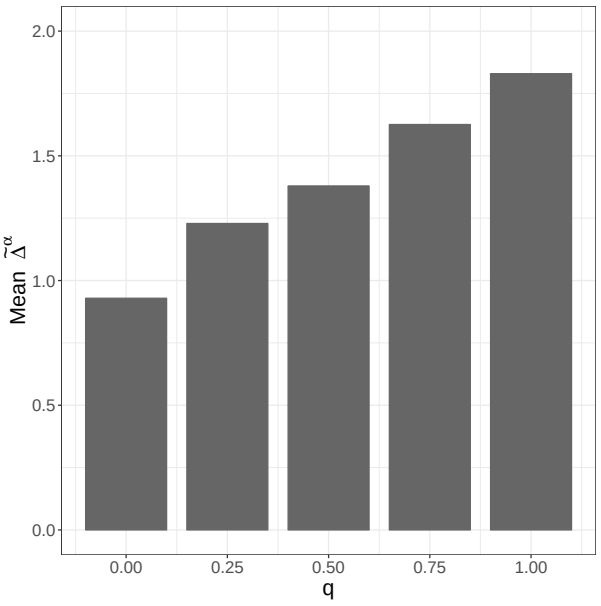


Figure 6: Mean of maximum relative payoff from the valence of the out-group candidate ($\tilde{\Delta}^{\alpha}$) voters accept and still vote for the in-group candidate plotted over policy polarization (q) in the *A vs B* treatment.

4.2.1 The instrumental impact of identity on partisan voting:

While Figure 5 confirms that voters are willing to vote for in-group candidates with lower relative valence, it does not distinguish between partisan voting due to expressive concerns and partisan voting due to an anticipated in-group bias in candidates' policy choices. Figure 6, which illustrates the average individual willingness to vote for the in-group candidate as a function of q, shows that $\tilde{\Delta}^{\alpha}$ is just under 1 when $q = 0$, but rises to approximately 1.8 when $q = 1$. This suggests that partisan voting is driven by both expressive and instrumental concerns.

To test the instrumental impact of social polarization on partisan voting statistically, we estimate our second empirical model and report the results in Table 5 (columns (5) and (6)). Any expressive impact of identity on partisan voting is captured by the constant (see Section 4.4

for details), while the coefficient on q estimates the instrumental impact of identity on partisan voting. We estimate a positive impact of q on $\tilde{\Delta}^{\alpha}$ in all treatments, with point estimates 0.86 and 0.88 for the minimal group treatments in Berlin and Florida, respectively, and all coefficient estimates are statistically significant with $p < .01$, other than the natural identity in Berlin, which is not statistically significant.[21]

Table 5: Maximum difference in the relative valence of the out-group candidate at which voters still vote for the in-group candidate ($\tilde{\Delta}^{\alpha}$) regressed on policy polarization q (columns 1-4) and natural identity treatment (columns 5 and 6); decision order included in all regressions.

	Dependent variable: $\tilde{\Delta}_a$					
	A vs B		*Bike vs Car*	*Dem vs Rep*		
	Berlin	*Florida*	*Berlin*	*Florida*	*Berlin*	*Florida*
	(1)	(2)	(3)	(4)	(5)	(6)
q	0.858***	0.888***	0.243	0.687***	0.867***	0.890**
	(0.185)	(0.330)	(0.166)	(0.176)	(0.184)	(0.331)
Natural identity					0.235	0.024
					(0.239)	(0.434)
$q \times$ Natural identity					-0.620**	-0.305
					(0.248)	(0.374)
Constant	0.913***	1.150***	1.260***	1.110***	0.966***	1.110***
	(0.238)	(0.327)	(0.204)	(0.250)	(0.209)	(0.292)
Observations	240	130	240	240	480	370
Subjects	48	26	48	48	96	74
Adjusted R^2	0.049	0.040	-0.001	0.030	0.022	0.346
F Statistic	10.82***	3.63**	1.50	9.75***	6.16***	6.42***

Standard errors clustered by subject
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The positive findings in both our baseline treatments regarding the instrumental impact of social polarization on partisan voting support our main prediction, Hypothesis 4:

Result 4
In the baseline minimal-group treatment, when social polarization exists, partisan voting increases as policy polarization (q) increases.

Lastly we highlight that, in addition to the aggregate results, subjects respond to instrumental incentives at the individual level, which suggests that partisan voting is not driven by

[21] The null finding in the natural identity treatment is interesting given that candidate policy choices were the most biased in this treatment. Even though it is not statistically significant, Figure D.4 in the Appendix shows a consistent increase in the mean $\tilde{\Delta}^{\alpha}$ as q increases, but at a must lower rate than the other treatments. It could be the case that our experiment is too low-powered to capture a smaller impact of q.

a subset of voters who always vote along partisan lines: only 14 percent of voters vote for the in-group candidate when policy polarization is low and Δ^{α} is high, while 61 percent vote for the in-group candidate when policy polarization is high and Δ^{α} is low. Overall, our result regarding the instrumental impact of identity is increasing in q is robust to subject-level heterogeneity. The distribution of individual-level choices display the same relationship between q and Δ^{x}, and mostly the same relationship between subject-specific $\tilde{\Delta}^{\alpha}$ and q. Moreover, we show that the subject-level heterogeneity in partisan voting can be traced back to subjects best responding to their own allocation decisions.[22]

4.2.2 The impact of natural identity on partisan voting:

Finally, we consider the impact of natural identities on partisan voting (Hypothesis 5). Note that we can only properly test Hypothesis 5 if our experiment is successful in inducing stronger identities in our natural identity treatments than our minimal group treatment (higher λ). Therefore, given our null finding for Hypothesis 3, which tests for a higher λ in the natural identity treatments, we cannot evaluate Hypothesis 5.[23] However, we present the analysis here for completeness.

Our results on voter choices are consistent with the null finding for Hypothesis 3 on candidate allocations: we find no evidence that partisan voting is higher under the natural identity treatments than the minimal-group treatment. On average, the vote share of the the in-group candidate is .77 (.73, .80) in the *A vs B* treatment, .76 (.72, .79) in the *Bike vs Car* treatment, and .80 (.76, .83) in the *Dem vs Rep* treatment. Additionally, the coefficient on the interaction of q and natural identity, presented in columns (5) and (6) of Table 5 is neither positive nor significantly different from zero for either of the natural identity treatments.

Our final result follows:

Result 5

In our main specification, we do not find evidence of an additional instrumental impact of social polarization on partisan voting in the natural identity treatments.

Absent variation in social polarization from the natural identity treatments, we are unable to provide a direct test of Hypothesis 5. However, in the Appendix, we explore two alternative metrics for social polarization by comparing "strong" and "weak" partisans and analyzing the impact of multiple identities. These alternative measures provide suggestive evidence that partisan voting is increasing in social polarization (see Section B in the Appendix).

4.3 Discussion

Overall, with positive findings for Hypotheses 1, 2 and 4, we find strong support for the predictions of the identity model. That is, we find that in both the minimal group and natural

[22] Section D.6 in the appendix details more of the analysis of subject-level heterogeneity.

[23] We do find that subjects in the *Dem vs Rep* treatment self-report a significantly higher feeling of closeness to their identity group in the exit survey than this is the case for subjects in the *A vs B* treatment ($p < .05$). They also offer a larger share of their endowment as first mover in a dictator game to in-group than out-group partners ($p < .05$).

identity treatments, candidates' choices are biased towards the partisan in-group, confirming that subjects' identity biases their distributional preferences towards in-group members.

We also find that voters are biased towards the in-group candidate, in the sense that they are willing to vote for an in-group candidate with lower relative valence (partisan voting). Our test of Hypothesis 3 confirms that partisan voting is more than an expressive phenomenon: the fact that partisan voting is increasing in q confirms that voters anticipate the in-group bias of the candidates, and strategically vote for the in-group candidate in a manner that reflects the changing instrumental incentives.

4.4 Robustness to expressive payoffs

In the model presented above, we do not explicitly model the expressive impact of identity on voting behavior. In an earlier version of the paper (see Duell and Valasek, 2018) we analyze a model where agents receive an expressive payoff for voting for the in-group candidate: there we show that if agents receive expressive payoffs, however, then the voter's decision becomes a simple calculus of comparing the expressive payoff for voting for the in-group candidate and the relative valence of the out-group candidate, weighed by the perceived probability of influencing the election outcome.

Accordingly, we can use the following strategy to identify a lower bound on the instrumental impact of identity on partisan voting.

1. Since both candidates select centrist policies when there is no polarization in voters' policy preference, if $\tilde{\Delta}^{\alpha} > 0$ when $q = 0$ the in-group voting bias is driven by expressive motives.
2. Assuming the probability of being pivotal is not negatively correlated with q, if $\tilde{\Delta}^{\alpha} > 0$ is higher for $q > 0$ relative to $q = 0$, then the increase in the in-group voting bias is driven by instrumental motives.

This identification strategy does depend on the assumption that the probability of being pivotal is not negatively correlated with q – if this assumption is violated, then expressive voting may be increasing in q since voting expressively become less costly (in expectation) as the probability of being pivotal decreases. However, given the structure of our model, the probability of being pivotal is increasing with the rate of in-group voting: since the probability of being pivotal is the highest when a randomly-drawn voter is equally likely to vote for each candidate, the probability of being pivotal is maximized when all agents vote for the in-group candidate. This implies that the probability of being pivotal is increasing in $\tilde{\Delta}^{\alpha}$, and hence increasing in q. Additionally, this assumption is consistent with our empirical findings: as we show below, the average margin of victory is decreasing in q.

To empirically assess the appropriateness of the pivotality assumption, we first check whether the margin of victory (the closeness of the election) is positively correlated with policy polarization in our experimental data. Given the low observations of tied elections, we follow the literature on elections and use the margin of victory as a proxy for the probability of a tied election (a tied election is more likely with a lower margin of victory).

We test our assumption by regressing the margin of victory on q, valence of the in-group candidate, and valence of the out-group candidate. We conclude that there is not a positive

relationship between q and the margin of victory: the coefficient estimate on q is $-.19\,(-.41, .03)$ with $p = .06$ in the *A vs B* treatment, $-.12\,(-.87, .63)$ in the *Bike vs Car* treatment, and $-.11\,(-.70, .48)$ in the *Dem vs Rep* treatment.[24]

Given that we do not find evidence of a a positive relationship between q and the margin of victory—indeed there is weak evidence of a negative relationship, consistent with the theory—we infer that there is no evidence to support the claim that the probability of being pivotal is decreasing with the degree of partisan voting. In fact, our coefficient estimates point to a positive correlation between q and the probability of being pivotal. Therefore, the observed voting behavior is consistent with our interpretation of partisan voting for $q = 0$ as a *lower bound* on the instrumental impact of identity on partisan voting.

Second, we consider the share of votes among all votes in which subjects were actually pivotal: that is, the proportion of votes that resulted in a tie or that were decided by a win margin of two (subjects voting with the majority are pivotal given a win margin of two). We find that subjects were pivotal in a significant proportion of elections, but that most of the pivotal elections occurred when there was no difference in valence between the two candidates (see Figure C.5 in the appendix for a graphical representation of pivotal elections as a function of valence differences and policy polarization). However, we find that for each identity treatment, there was at least one election where the candidates had different valence and voters were pivotal, showing that the probability of being pivotal was non-zero.

Given the robustness of our experimental strategy of accounting for expressive payoffs, we are also able to say something about the relative magnitude of each phenomenon: when the political game is zero-sum ($q = 0$) we find that expressive voting accounts for approximately 45 percent of partisan voting, while instrumental voting accounts for the remaining 55 percent. This suggests that both expressive and instrumental concerns are important in accounting for the voting choices of subjects in our experiment.

5 Conclusion

In this paper we characterize and measure the impact of salient identities on candidate and voter behavior. In particular, we demonstrate theoretically that identity can have an instrumental effect on partisan voting by influencing voters beliefs regarding the ex post decisions of political representatives. Our laboratory experiment largely confirms the predictions of our formal model, and shows that identity impacts voter behavior due to both instrumental and expressive reasons. Moreover, we show that the impact of social polarization on voting behavior is strongest when there is large underlying polarization in voters' policy preferences.

Our findings suggest a society divided into two groups with starkly divergent policy preferences but with low levels of social polarization may still support a relatively consensual, non-partisan and efficient political system. Instead, socially-costly partisan behavior will occur when both policy and social polarization are high. In this case, as a society becomes polarized

[24] Standard errors are clustered at the session-level. The win margin is computed for each session, treatment, and (q,in-group valence, out-group valence)-triple. Figures C.3 and C.4 in the appendix further illustrates the distribution of win margins and the relationship of win margins and q.

on a social dimension that is correlated with political preferences, voters come to expect a higher degree of partisanship from elected officials, which causes them to rationally respond by voting along partisan lines, leading to the selection of candidates with lower average levels of quality.

This narrative is particularly problematic when applied to the US context, where evidence suggests that social polarization occurs along an explicitly political dimension (Republican/Democrat). In this case, a self-reinforcing cycle may arise: Social polarization leads to the perception of increasing divergence in the policy platforms of the two parties—in turn, this may cause partisan identities to strengthen, leading to an even greater degree of social polarization and a further increase in partisan behavior.

References

Akerlof, George and Rachel Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115(3):715–53.

Andreoni, James, Deniz Aydin, Blake Barton, B Douglas Bernheim and Jeffrey Naecker. 2018. When Fair Isn't Fair: Understanding Choice Reversals Involving Social Preferences. Technical report National Bureau of Economic Research.

Bassi, Anna, Rebecca B Morton and Kenneth C Williams. 2011. "The effects of identities, incentives, and information on voting." *The Journal of Politics* 73(2):558–571.

Charness, Gary, Luca Rigotti and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review* 97(4):1340–52.

Charness, Gary and Matthew Rabin. 2002. "Understanding social preferences with simple tests." *The Quarterly Journal of Economics* 117(3):817–869.

Chen, Roy and Yan Chen. 2011. "The Potential of Social Identity for Equilibrium Selection." *American Economic Review* 101(6):2562–89.

Chen, Yan and Sherry Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 99(1):431–57.

Cox, James C., Daniel Friedman and Steven Gjerstad. 2007. "A tractable model of reciprocity and fairness." *Games and Economic Behavior* 59(1):17 – 45.
URL: *http://www.sciencedirect.com/science/article/pii/S0899825606000662*

Duell, Dominik and Justin Valasek. 2018. "Social Polarization and Partisan Voting in Representative Democracies." *CESifo Working Paper Series* (7040).

Eckel, Catherine and Philip Grossman. 2005. "Managing Diversity by Creating Team Identity." *Journal of Economic Behavior & Organization* 58:371–392.

Goette, Lorenz, David Huffman and Stephan Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review* 96(2):212–6.

Goette, Lorenz, David Huffman and Stephan Meier. 2012. "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups." *American Economic Journal: Microeconomics* 4(1):101–15.

Green, Donald, Bradley Palmquist and Eric Schickler. 2002. *Partisan Hearts and Minds*. New Haven: Yale University Press.

Hamlin, Alan and Colin Jennings. 2011. "Expressive political behaviour: Foundations, scope and implications." *British Journal of Political Science* 41(03):645–670.

Huddy, Leonie, Alexa Bankert and Caitlin L Davies. 2018. "Expressive vs. Instrumental Partisanship in Multi-Party European Systems." *Advances in Political Psychology* .

Huddy, Leonie, Lilliana Mason and Lene Aarøe. 2015. "Expressive partisanship: Campaign involvement, political emotion, and partisan identity." *American Political Science Review* 109(01):1–17.

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

Klor, Esteban and Moses Shayo. 2010. "Social Identity and Preferences over Redistribution." *Journal of Public Economics* 94(3):269–78.

Landa, Dimitri and Dominik Duell. 2015. "Social Identity and Electoral Accountability." *American Journal of Political Science* 59(3):671–89.

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

Shayo, Moses. 2009. "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution." *American Political Science Review* 103(2):147–74.

Tajfel, Henri. 1981. *Human Groups and Social Categories*. Cambridge: Cambridge University Press.

Tajfel, Henri and Michael Billig. 1974. "Familiarity and Categorization in Intergroup Behavior." *Journal of Experimental Social Psychology* 10:159–70.

Turner, John C and Rupert Brown. 1978. "Social status, cognitive alternatives and intergroup relations." *Differentiation between social groups: Studies in the social psychology of intergroup relations* pp. 201–234.

Tyran, Jean-Robert. 2004. "Voting when money and morals conflict: an experimental test of expressive voting." *Journal of Public Economics* 88(7):1645–1664.

Appendix

A Proofs for Section 2

Proof of Lemma 1:
First, note that the candidates' aggregate payoffs are constant and equal to $x^l + x^w$. Next, note that $\max_{\mathbf{p}} \sum^N E[v(\mathbf{p}, p_i)] = \{0, 1, 0\}$ since a "policy unit" placed at p^m generates aggregate expected payoffs of $n(1-q)+nq$, while a policy unit placed at p^l or p^r generates aggregate payoffs of $1/2(n(1-q)+nq) \leq n(1-q)+nq$. Therefore, since $\mathbf{p} = \{0, 1, 0\}$ maximizes aggregate expected payoffs and results in an equal distribution of expected payoffs, $\mathbf{p} = \{0, 1, 0\}$ maximizes $\sum^N E[v(\mathbf{p}, p_i) + \alpha^w]^{1/m}$ for any $m \geq 1$ and α^w. Lastly, given that valence is a public good, aggregate payoffs are higher when the candidate with $\alpha_k \geq \alpha_{k'}$ wins the election. Together, this implies that maximal aggregate payoffs are achieved when both candidates choose centrist policies, and all voters vote for the highest-valence candidate. ■

Proof of Lemma 2:
Since the equilibrium concept is SPNE, by backward induction, the winning candidate will choose the policy that maximizes their distributional preferences, $g(\mathbf{E}[\mathbf{x_i}]^I, \mathbf{E}[\mathbf{x_i}]^{I^-}) = \sum_N E[x_i]^{1/m}$, which is equivalent to the following maximization problem, $\max_{\mathbf{p}} \sum^N E[v(\mathbf{p}, p_i)+\alpha^I]^{1/m}$, which is equal to $\{0, 1, 0\}$ by the proof of Lemma 1. ■

Proof of Lemma 3:
As in the above proof, the winning candidate will choose the policy that maximizes their social preferences, $g(\mathbf{E}[\mathbf{x_i}]^I, \mathbf{E}[\mathbf{x_i}]^{I^-})$, which results in the following maximization problem:

$$\max_{\mathbf{p}} \left(\lambda \sum_{i \in I} E[x_i]^{1/m} + (1-\lambda) \sum_{i \in I^-} E[x_i]^{1/m} \right).$$

For simplicity, we assume c^A wins the election and sets $\mathbf{p}^w = \mathbf{p}^A$ ($\mathbf{p}^B$ is symmetric).

To prove the result, we focus on the marginal utility that c^A receives from shifting a unit of policy to p^l, *given* $\mathbf{p}^A = \{0, 1, 0\}$. Note that:

$$E[v(\mathbf{p}^w, p_i) + \alpha^A] = q(p^l + 0.5p^m) + (1-q)(0.5p^l + p^m + 0.5p^r) + \alpha^A,$$

for $i \in A$, and

$$E[v(\mathbf{p}^w, p_i) + \alpha^A] = (1-q)(0.5p^l + p^m + 0.5p^r) + q(0.5p^m + p^r) + \alpha^A,$$

for $i \in B$. Therefore, the relative marginal utility of partisan policy at $\mathbf{p}^A = \{0, 1, 0\}$ is equal to:

$$\begin{aligned} &\lambda(n/2-1)\frac{1}{m}E[x_i|i \in A]^{-\frac{m-1}{m}}[\frac{1}{2}q - \frac{1}{2}(1-q)] \\ &+(1-\lambda)(n/2-1)\frac{1}{m}E[x_i|i \in B]^{-\frac{m-1}{m}}[-\frac{1}{2}(1-q) + -\frac{1}{2}q], \end{aligned} \tag{9}$$

which is positive iff:

$$\lambda[\frac{1}{2}q - \frac{1}{2}(1-q)] > (1-\lambda)[\frac{1}{2}(1-q) + \frac{1}{2}q] \Rightarrow \lambda(2q-1) > 1-\lambda. \tag{10}$$

(Note that the same calculation for p^r yields the equation $(1-\lambda)(2q-1) > \lambda$, which implies that $p^r = 0$, since the marginal relative utility of placing a unit of policy at the outgroup extreme is always negative.) Therefore, $\mathbf{p}^A = \{0, 1, 0\}$ if Equation 10 does not hold, and $\mathbf{p}^A = \{p^l, p^m, 0\}$

with $p^l > 0$ if Equation 10 holds. ■

Proof of Proposition 1:
We prove this result separately for $m = 1$ and $m > 1$. The proof for $m > 1$ follows from Expression 9 and Equation 10 in the proof of Lemma 3 above. First, fixing λ, there exists q^* such that $\lambda(2q-1) = 1-\lambda$, since $\lambda > 0.5$. Specifically:

$$q^* = \frac{1}{2\lambda}.$$

For, $q < q^*$, Equation 10 does not hold, and as shown in the proof of Lemma 3, the unique equilibrium policies for both candidates set $\mathbf{p}^I = \{0, 1, 0\}$; for $q > q^*$, however, both candidates will allocate a strictly positive amount to ingroup extreme. For convenience, assume c^A wins the election. In this case, c^A will set p^l, p^m such that the relative marginal utility of partisan policy (Expression 9) is equal to zero. That is:

$$\frac{E[x_i|i \in A]}{E[x_i|i \in B]} = \left(\frac{\lambda(2q-1)}{(1-\lambda)}\right)^{\frac{m}{m-1}} \tag{11}$$

Both the RHS and LHS of Equation 11 are continuous in q, λ and p^l. Moreover, for $q > q^*$, the LHS is increasing in p^l since $E[x_i|i \in A]$ is increasing in p^l and $E[x_i|i \in B]$ is decreasing in p^l, which implies a unique crossing.

For $\partial\Delta^x/\partial\lambda$, note that the RHS of Equation 11 is increasing in λ while the LHS is constant. Therefore, the equilibrium value of p^l is increasing in λ, which implies that $E[x_i|i \in A] - E[x_i|i \in B]$ is increasing in λ as well. Second, the case of $\partial\Delta^x/\partial q$ is a bit more complex since both the RHS and LHS of Equation 11 are functions of q. However, the result can be proved directly. First, we introduce the following simplified notation: $X = (\lambda(2q-1)/(1-\lambda))^{m/(m-1)}$.

Consider a discrete increase in q to $q' = q + \Delta q$. We wish to show that $E[x_i|i \in A]' - E[x_i|i \in B]' > E[x_i|i \in A] - E[x_i|i \in B]$ or, equivalently, that $\Delta E[x_i|i \in A] > \Delta E[x_i|i \in B]$. Using this notation and rearranging Equation 11, we get:

$$E[x_i|i \in A]' + (X + \Delta X)E[x_i|i \in B]' = 0 = E[x_i|i \in A] + XE[x_i|i \in B].$$

And since X is increasing in q, which implies that $\Delta X > 0$, we get:

$$E[x_i|i \in A]' + XE[x_i|i \in B]' > E[x_i|i \in A] + XE[x_i|i \in B].$$

Rearranging this equation gives:

$$\Delta E[x_i|i \in A] + X\Delta E[x_i|i \in B] > 0 \Rightarrow \Delta E[x_i|i \in A] > X\Delta E[x_i|i \in B].$$

Note that since $X > 1$ this expression implies that $\Delta E[x_i|i \in A] > \Delta E[x_i|i \in B]$, which proves the result for a discrete change. Lastly, since all expressions are continuous in q, this result holds as $\Delta q \to 0$.

For $m = 1$, the proof stems from the fact that with linear utility, candidates will either allocate all policy to the center, or all policy to the partisan extreme. Too see this, note that by Expression 9 the marginal relative return for allocating a unit of policy to the partisan extreme is equal to:

$$(n/2 - 1)[\lambda(q - 1/2) - 1/2(1-\lambda)],$$

which is positive if $q > 1/(2\lambda)$. First, this shows that the threshold at which candidates switch from allocating all policy to the center to allocating all policy to the partisan extreme, which

results in an increase in Δ^x, is decreasing in q and λ. Second, Δ^x is increasing in q when all policy is allocated to the partisan extreme, and constant when all policy is allocated to the center. Together, these prove the result that Δ^x is increasing in q and λ for $m = 1$. ■

Proof of Proposition 2:
Lemma 2 shows that under the Benchmark model, both candidates will select $\mathbf{p}^w = \{0, 1, 0\}$. Since both candidates select the same policy, voter i receives the following relative expected utility for voting for the ingroup candidate:

$$\bar{p}(\alpha^I - \alpha^{i^-}),$$

which is positive iff $0 \geq \bar{p}\Delta^\alpha$. ■

Proof of Proposition 3:
Lemma 3 shows that under the Identity model, both candidates may select policies that favor their partisan ingroup. Therefore, voter i receives the following relative expected utility for voting for the ingroup candidate:

$$\bar{p}\big(E[x_i|\mathbf{p}^I] - E[x_i|\mathbf{p}^{I^-}] + \alpha^I - \alpha^{i^-}\big),$$

which is positive iff $\bar{p}\Delta^x \geq \bar{p}\Delta^\alpha$. ■

Proof of Corollary 1:
(i) Follows directly from Proposition 2: Under the Benchmark model the candidates choose the same policy for all q, which implies that voters have a constant incentive to vote for the ingroup candidate.
(ii) In contrast to (i), Proposition 1 shows that the incentive to vote for the ingroup candidate is increasing in q, since candidates choose more partisan policy for higher q.
(iii) Proposition 1 shows that when $q = 0$, there is no instrumental incentive to vote for the ingroup candidate under the Identity model since $\Delta^x = 0$. ■

B Suggestive evidence on the impact of higher social polarization

While our natural identity treatments seemingly do not implement a higher level of social polarization than our minimal-group group treatment, we are able to report on an alternative measure of social polarization. Namely, in the *Dem vs Rep* treatment in Florida, subjects were also asked to self-report the strength of their party-identification in an post-experiment questionnaire. We classify subjects into "strong partisans" and "weak partisans:" subjects who answered that they are either "Strong Democrats" or "Strong Republicans" are classified as strong partisans, while subjects who answered that they are "Independents", "Not Strong Democrats", or "Not Strong Republicans" are classified as weak partisans (this question and classification are standard in election studies). We find that 25% of our subjects self-report as strong partisans.

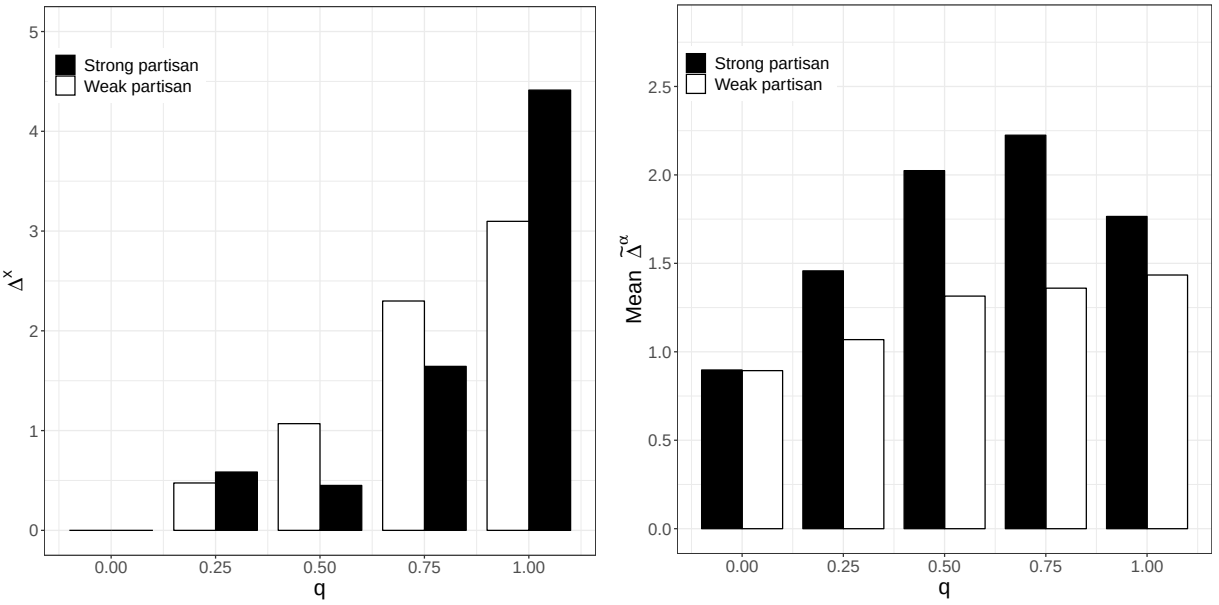
This alternative measure of social polarization allows us to examine whether subjects who self-report a stronger party identity exhibit a greater in-group bias as candidates, and a higher degree of partisan voting as subjects. Our treatments were not explicitly designed to test the impact of higher affect using within-session variation – thus, this ex-post analysis should be considered as exploratory. However, since our natural identity treatments did not induce a higher degree of social polarization (λ), this alternative metric for λ allows us to examine whether the division between strong and weak partisans provides suggestive evidence to support the comparative static predicted by Hypothesis 5.

First, comparing the candidate choices of strong and weak partisans, we find essentially no difference in the point estimate of the average in-group payoff bias, Δ^x, from 1.40 (.92, 1.89) for weak partisans to 1.41 (.68, 2.21) for strong partisans. For voters choices, however, we see a large increase in the willingness to vote for the in-group candidate. Moreover, this increase appears to be solely due to an increase in instrumental partisan voting: there is virtually no difference in expressive partisan voting from strong and weak partisans when $q = 0$, but $\tilde{\Delta}^\alpha$ is on average .57 (.11, 1.08) higher for strong partisans when $q > 0$. These data are displayed in Figures B.1 and B.2.

Given that we find an increase in instrumental partisan voting for strong partisans, individuals who self-report a stronger identification with the in-group identity may expect candidates to display a higher degree of in-group bias.

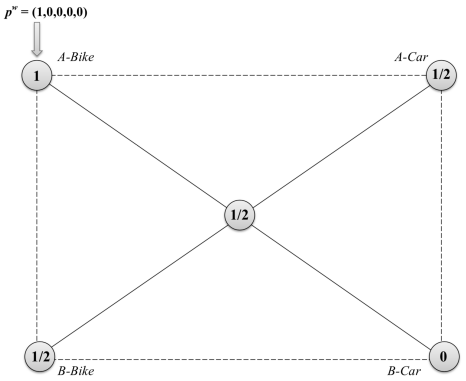
Figure B.1: Difference in average token payoffs from the allocations by in-group and out-group candidates (Δ^x) by policy polarization (q) and self-reported strength of party identification in the *Dem vs Rep* treatment.

Figure B.2: Mean of maximum relative payoff from the valence of the out-group candidate ($\tilde{\Delta}^\alpha$) voters accept and still vote for the in-group candidate by policy polarization (q) and self-reported strength of party identification in the *Dem vs Rep* treatment.



We also find additional suggestive evidence regarding the impact of stronger social polarization from an additional treatment where subjects' group identities were two-dimensional—one exogenous, *A vs B*, and one endogenous, *Bike vs Car*. In this treatment, candidates allocated policy in a five-point policy space consisting of a center point and an extreme point for each two dimensional identity; e.g. a voter with identity *A-Bike* has an ideal at the upper-left with probability q and an ideal point at the center with probability $(1 - q)$ (see Figure B.3 for an illustration of voters' payoffs). We report on this treatment in detail in Section D.5.

Figure B.3: Payoffs to each voter as a function of their ideal point, given that the winning policy allocates one unit of policy at the upper-left extreme—with probability q voters have ideal points at the extreme points corresponding to their two-dimensional identity.



Consistent with Hypothesis 3, candidates display a higher bias towards voters with which they share identities on both identity-dimensions, rather than equalizing payoff for all voters with which they share a single identity (Figure B.4). Similarly, consistent with Hypothesis 5, voters are significantly more likely to vote for candidates with which they share both identities even when they have lower relative valence, and the rate of increase in partisan voting with q is also higher for these candidates (Figure B.5).

While we did not intend for either of these alternatives to be our main test of Hypothesis 5, given our inability to confirm or reject Hypothesis 5 in our main treatments due to the lack of an increase in social polarization with natural identities treatments, we find it relevant to report this suggestive evidence on the impact of increased social polarization on partisan voting.

Figure B.4: Average token payoff from candidate allocation (Δ^x) by policy polarization (q) and target of allocation in the multi-identity treatment: exact identity, A or B, or Bike or Car.

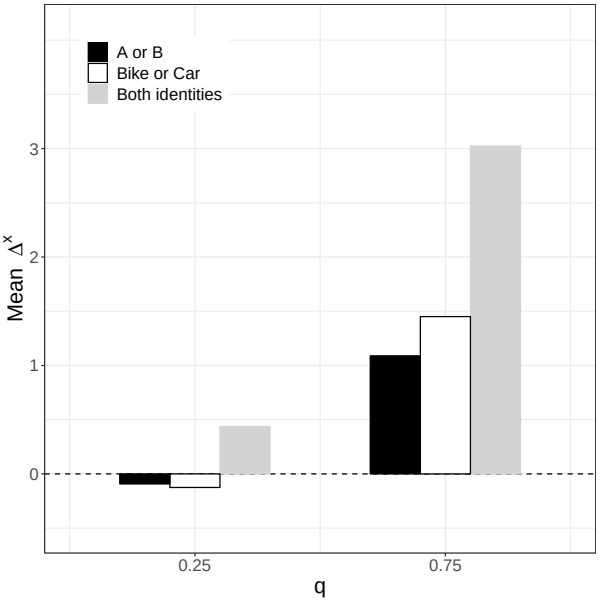
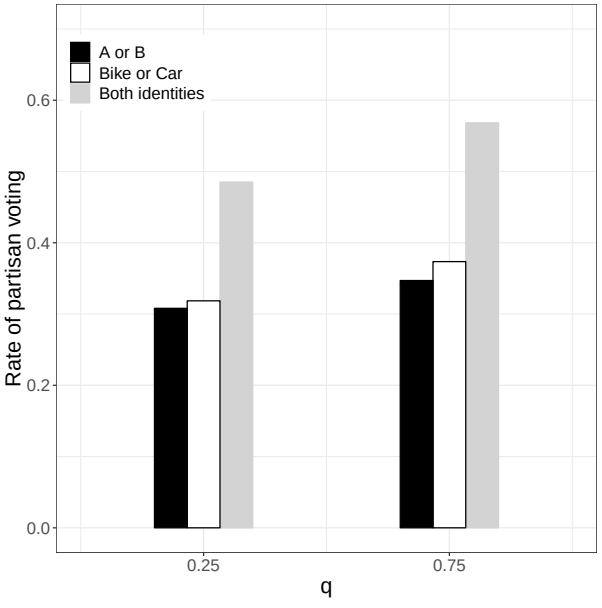


Figure B.5: Partisan voting by policy polarization (q) and target of the partisan vote in the multi-identity treatment: exact identity, A or B, or Bike or Car.



C Experimental design

C.1 Decision environment by round

The placeholder in the table, e.g. the probabilities "1" to "5" in column 4 were filled with the different values of assigned policy polarization (q), that is either 0, .25, .5, .75, or 1. The random order of assignment of q (without replacement) varies by subject.

Table C.1: Sequence of choices by subjects in the role of voters and candidates in the voting game. Policy polarization q (0, .25, .5, .75, .1) is randomly assigned to each subject by block of 9 rounds each

Round	Voters	Candidates	q-Probability Block
1	Vote 1	Allocation 1	1
2	Vote 2	–	
⋮	⋮	–	
9	Vote 9	–	
10	Vote 10	Allocation 2	2
11	Vote 11	–	
⋮	⋮	–	
18	Vote 18	–	
19	Vote 19	Allocation 3	3
20	Vote 20	–	
⋮	⋮	–	
27	Vote 27	–	
28	Vote 28	Allocation 4	4
29	Vote 29	–	
⋮	⋮	–	
36	Vote 36	–	
37	Vote 37	Allocation 5	5
38	Vote 36	–	
⋮	⋮	–	
45	Vote 45	–	

C.2 Instructions: *A vs B* treatment

Introduction
In this experiment you will make a series of choices. At the end of the experiment, you will be paid according to your choices and the choices of other participants. Pay close attention to the instructions because each of your decisions potentially affects your payoff from this experiment. This experiment has two parts. Your total earnings will consist of a show-up fee of 7 Dollars and your earnings from each of the two parts of the experiment. During the course of the experiment you will earn tokens, which will be exchanged into Dollars at the end of the experiment at a rate of

1 Token = 60 Cent.

We will start with a brief instruction period and Part 1 of the experiment. You will then receive instructions for Part 2 of the experiment and finish that part accordingly. Should you have questions while I read out these instructions, please raise your hand and after I have finished

reading the instructions, I will come and assist you. Should you have questions during the experiment, please raise your hand at any time.

Part 1
In part 1 of the experiment, you will make **50 decisions**: 45 in the role of a **Voter** and 5 in the role of a **Candidate**.

Assignment to Group A and Group B
At the beginning of the experiment you will be randomly assigned to either **Group A** or **Group B**. You will remain a member of this group until the end of the experiment; that is, until you have made all 50 decisions.

Decisions as Voter
As a voter you will make 45 decisions. In each decision, you will be asked whether you prefer **Candidate A, who is a member of Group A, or Candidate B, who is a member of Group B**.

While you are making your decisions, you will see the following information on the screen:

1. The level of **Ability** of Candidate A and Candidate B;
2. The probability with which your **Position** is either **Left**, **Center**, or **Right**.

The level of ability of Candidate A and Candidate B is either **Low**, **Average**, or, **High**.

Importantly, if you are assigned to Group A, your position can only be **Left** or **Center** and if you are assigned to Group B, it can only be **Center** or **Right**.

Additionally, when you are assigned to Group A, the probability that your Position is Left, and not Center, in the voter decisions is either 100%, 75%, 50%, 25% or 0%. When you are assigned to Group B, the probability that your Position is Right, and not Center, in the voter decisions is either 100%, 75%, 50%, 25% or 0%.

To assist you, you will see those probabilities on your screen while you make your decision.

Decisions as Candidate
As a Candidate you will make 5 decisions. In each decision, you will be ask how you want **to allocate up to 10 Tokens to the positions Left, Center, or Right**. You may allocate all 10 Tokens to one position, allocate them in any way over two or three positions, allocate less than 10 Tokens, or do not allocate tokens at all.

While you are making your decisions as a Candidate, you will see on the screen **the probabilities with which the Voters are distributed over the three positions Left, Center, and Right**.

That is, for each candidate decision, you will learn whether the Voters in **Group A** are always on Position **Left** (100%) and never on **Center** (0%), mostly Left (75%) and rarely Center (25%), equally likely Left (50%) and Center (50%), rarely Left (25%) and mostly Center (75%), or never Left (0%) and always Center (100%). You will receive similar information about the probability distribution of Voters in **Group B**, differing only in the fact that those Voters are either on Position **Right** or Position **Center**.

To illustrate the distribution of probabilities of Voters in each decision, you will see them on your screen while making your decisions as a Candidate.

Earnings in part 1 of the experiment
After you have made 45 decisions as a Voter and 5 decisions as a Candidate, your earnings will be calculated as follows:

1. One participant in Group A and one participant in Group B is randomly selected as Candidate A and Candidate B.
2. One of the distribution of probabilities of the Voters Positions is randomly chosen – every distribution is equally likely to be picked.
3. Either a low, average, or high ability is randomly assigned to the two participants who are selected as candidates – each level of ability is equally likely to be picked.
4. The choices of voters in the decision situation that corresponds to the randomly chosen levels of ability of the two candidates, and the randomly picked distribution of probabilities of the Voters Positions, is used to determine whether Candidate A or Candidate B has won the election.
5. The candidate who received a majority of votes of the participants who were not picked to be the candidates is the Winner of the Election; should both candidates receive the same number of votes, a fair coin is tossed to determine the winner.
6. The assigned ability and the allocation of tokens of the **Winner of the Election** now serves as basis for the earnings of the remaining participants. The two participants who were selected to be the candidates are paid according to the outcome of the election.

Should you be picked to be one of the Candidates for the purpose of determining earnings, your earnings will be

15 Tokens

if you are the Winner of the Election, but only

5 Tokens

if you are **not** the Winner of the Election.

If you are picked to be one of the Voters for the purpose of determining earnings, your earnings will be

As a Voter in Position Left:
Income based on the ability of the Winner of the Election + Tokens Left + (Tokens Center)/2

As a Voter in Position Center:
Income based on the ability of the Winner of the Election + Tokens Center + (Tokens Left + Tokens Right)/2

As a Voter in Position Right:
Income based on the ability of the Winner of the Election + Tokens Right + (Tokens Center)/2

The **Income based on the ability of the Winner of the Election** is **5 Tokens** when the assigned ability is **High**, is **3 Tokens** when the ability is **Average**, and is **2 Tokens** when the ability is **Low**.

Here is an example of how your earnings are calculated. After all participants have made all of their decisions, two participants, one from Group A and one from Group B, are randomly chosen to be Candidate A and Candidate B, respectively. Now, suppose Candidate A from Group A is randomly assigned a low ability and Candidate B from Group B is randomly assigned a high ability. Further, suppose the randomly chosen probability distribution specifies that the Position of Voters in Group A is Left with a probability of 75% and Center with a probability of 25%, and that the Position of Voters in Group B is Right with a probability of 75% and Center with a probability of 25%.
Additionally, assume that in the decision situation with this probability distribution, a Candidate A with low ability, and a Candidate B with high ability, a majority of Voters prefer Candidate B. Further suppose that the participant chosen to be Candidate B allocated 7 Tokens to Position Left, 2 to Position Center, and 1 to Position Right. In this way, the participant who was chosen to be Candidate B receives 15 Tokens as Winner of the Election, and the participant who was chosen to be Candidate A receives 5 Tokens for losing the election. Moreover, of the remaining participants, those chosen to be Voters in Position Left receive $5 + 7 + 2/2 = 13$ Tokens, in Position Center receive $5 + 2 + (7 + 1)/2 = 11$ Tokens, and in Position Right receive $5 + 1 + 2/2 = 7$ Tokens.

You will receive instructions for part 2 on your screen at the beginning of that part.

Again, your total earnings in this experiment will consist of a show-up fee of 7 Dollars and your earnings in part 1 and part 2 of the experiment.

C.3 Screen shots of subjects decisions screens in the *A vs B* treatment

Figure C.1: Voter decision screen

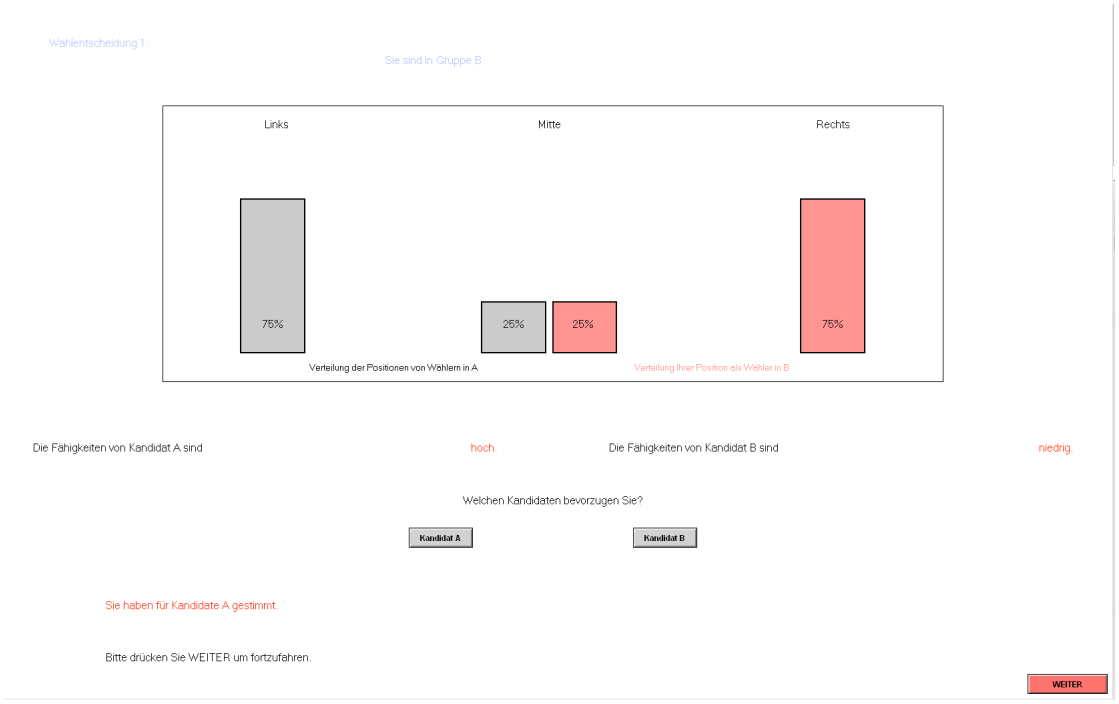


Figure C.2: Candidate decision screen

C.4 Assessing the pivotality assumption

Figure C.3: Distribution of vote margins by identity treatment.

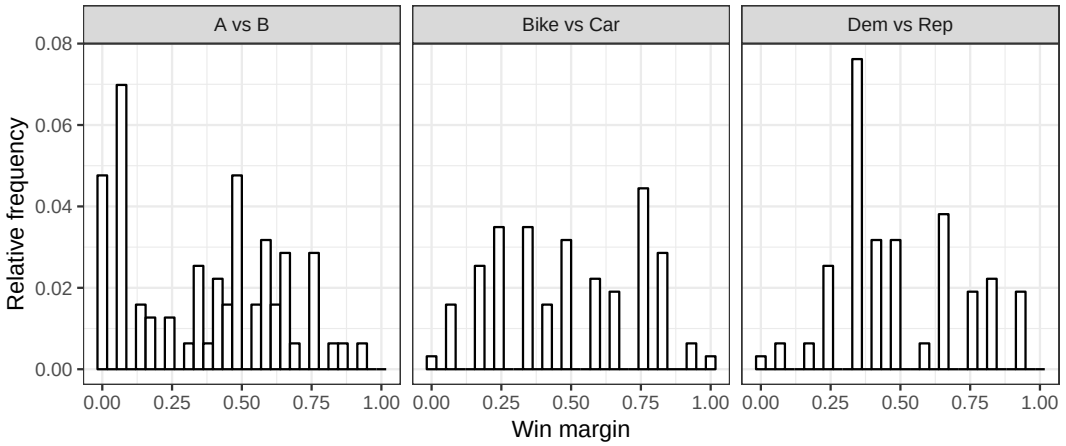


Figure C.4: Vote margin plotted over policy polarization (q) by identity treatment. Blue line indicates linear fit line of vote margins regressed on q.

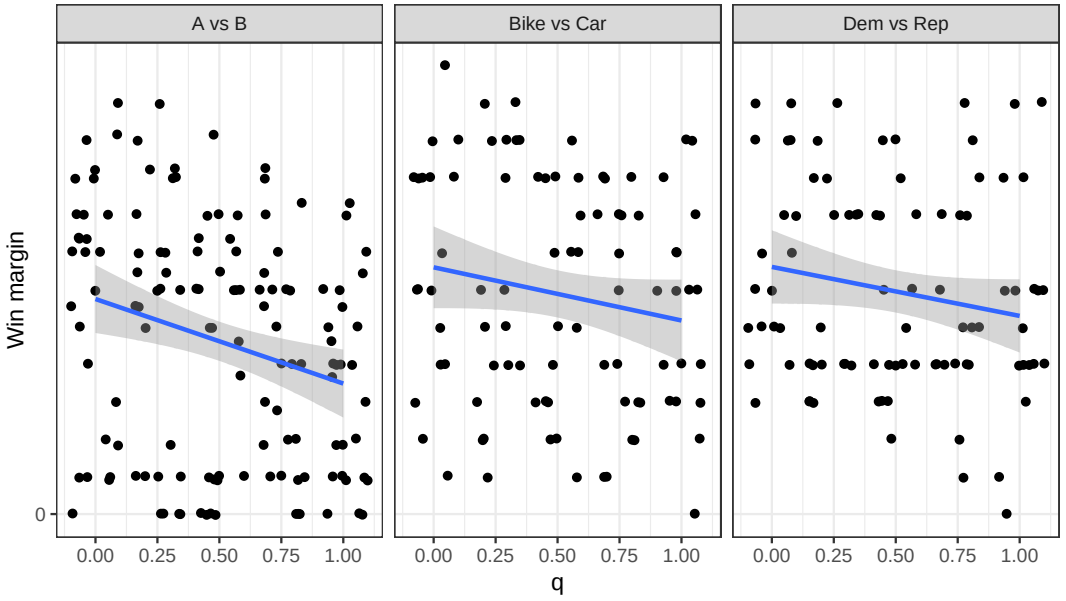
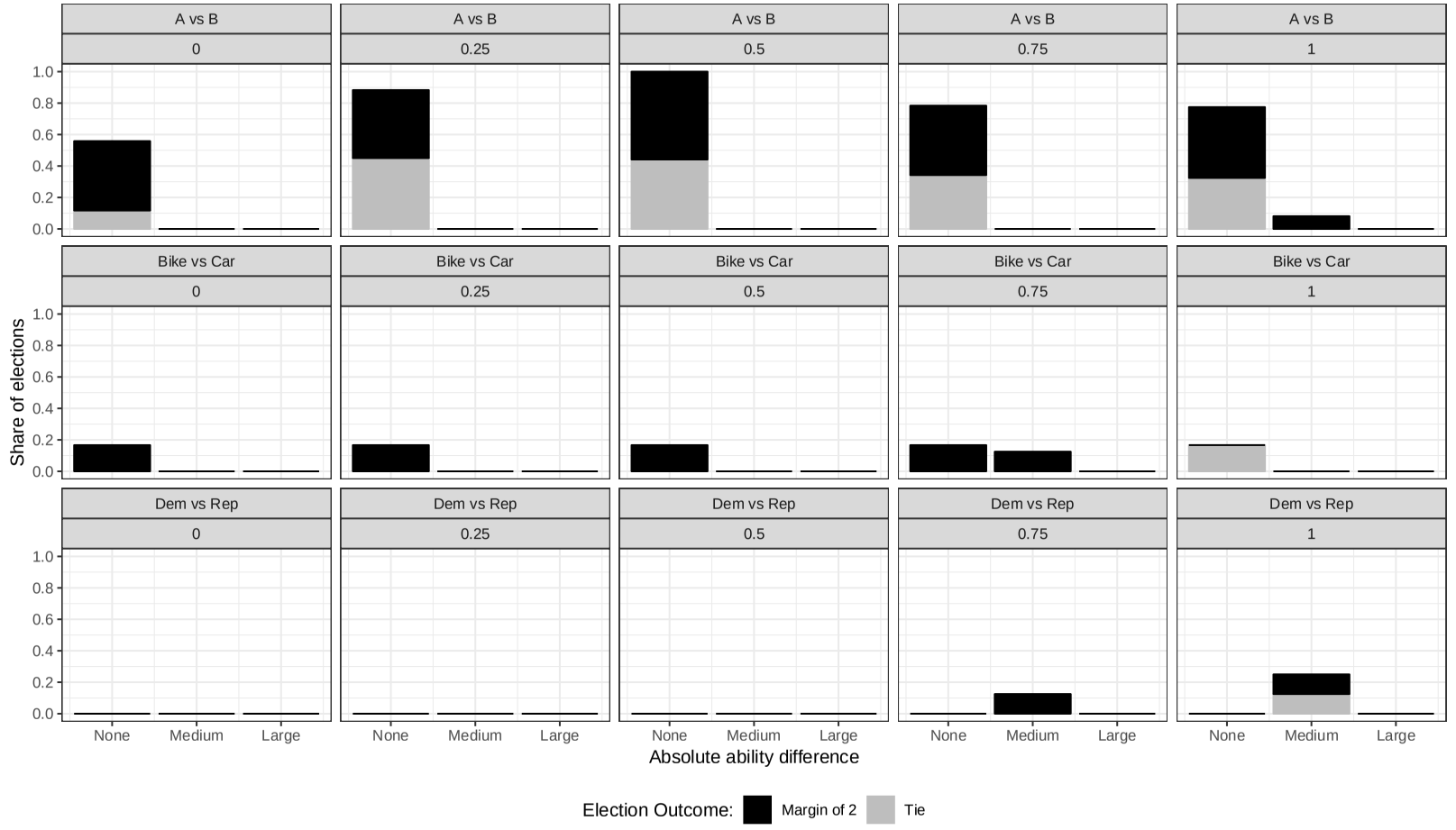


Figure C.5: Frequency of elections in which voters were pivotal, by policy polarization (q) and identity treatment: note that all voters are pivotal given a tie, and voters in the majority are pivotal given a winning margin of 2.



D Statistical appendix

D.1 Session statistics

Table D.1: Session statistics

	Treatment						
	A vs B			**Bike vs Car**		**Dem vs Rep**	
Session	1	2	3	1	2	1	2
Number of subjects	24	24	26	24	24	24	24
A	12	12	13				
B	12	12	13				
Bike				16	9		
Car				8	15		
Democrat						16	17
Republican						8	7
Number of observations							
of voters	1080	1080	1170	1080	1080	1080	1080
of candidates	120	120	130	120	120	120	120
Average earnings (in Euro)	13.9	12.9	13.4	14.5	12.9	13.3	14.3

D.2 Summary statistics

Table D.2: Summary statistics

Variable	**A vs B**	**Bike vs Car**	**Dem vs Rep**
Share of subjects in A	.50 (.50)		
Share of subjects in Bike		.52 (.50)	
Share of subjects in Dem			.69 (.46)
voting decision			
vote for A over B	.52 (.50)		
vote for Bike over Car		.55 (.50)	
vote for Dem vs Rep			.61 (.49)
$\tilde{\Delta}^{\alpha}$			
Overall	.41 (.92)	.78 (.79)	.86 (.82)
at $q = 0$	.54 (.95)	.71 (.77)	.60 (.76)
.25	.73 (.88)	.73 (.68)	.73 (.82)
.50	.84 (.92)	.79 (.85)	.98 (.89)
.75	1.00 (.86)	.81 (.89)	1.00 (.80)
1	1.18 (.87)	.83 (.75)	1.00 (.77)
allocation decision			
Left	2.84 (2.87)	2.89 (2.91)	3.65 (2.87)
Center	4.33 (3.24)	4.40 (3.25)	4.01 (2.92)
Right	2.68 (2.67)	2.67 (3.10)	2.05 (2.02)
Δ^{x}			
Overall	1.86 (3.06)	2.21 (3.20)	1.62 (2.67)
at $q = 0$	0	0	0
.25	.34 (.57)	.55 (.57)	.56 (.75)
.50	1.37 (1.45)	1.59 (1.61)	1.10 (1.59)
.75	2.69 (2.89)	3.30 (3.00)	2.53 (2.83)
1	4.92 (4.50)	5.62 (4.35)	3.90 (3.87)

D.3 Candidate policy choices

Figure D.1: Distribution of difference in token payoffs from the allocations by in-group and out-group candidates (Δ^x) for *A vs B*, *Bike vs Car*, and *Dem vs Rep* treatments.

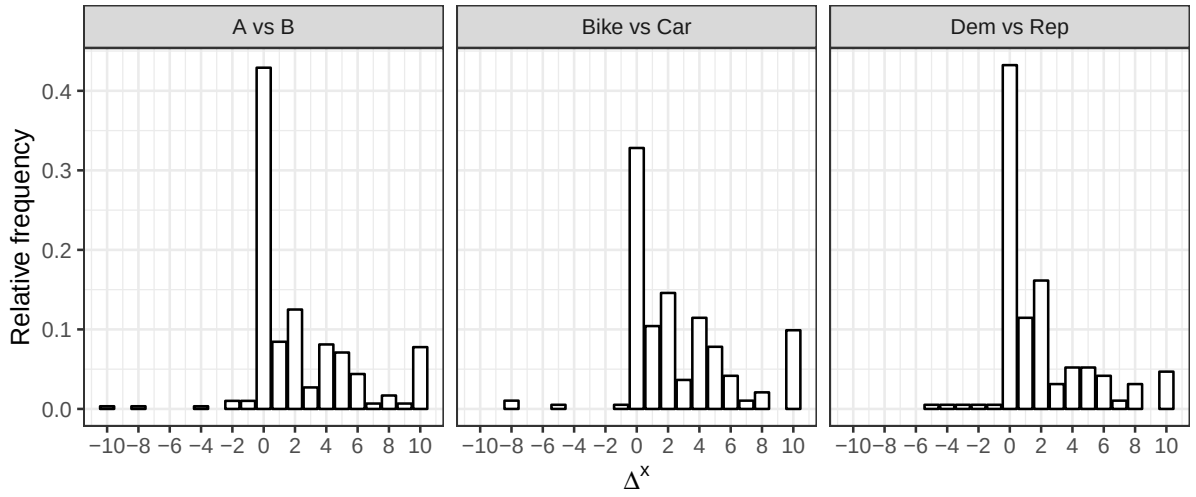
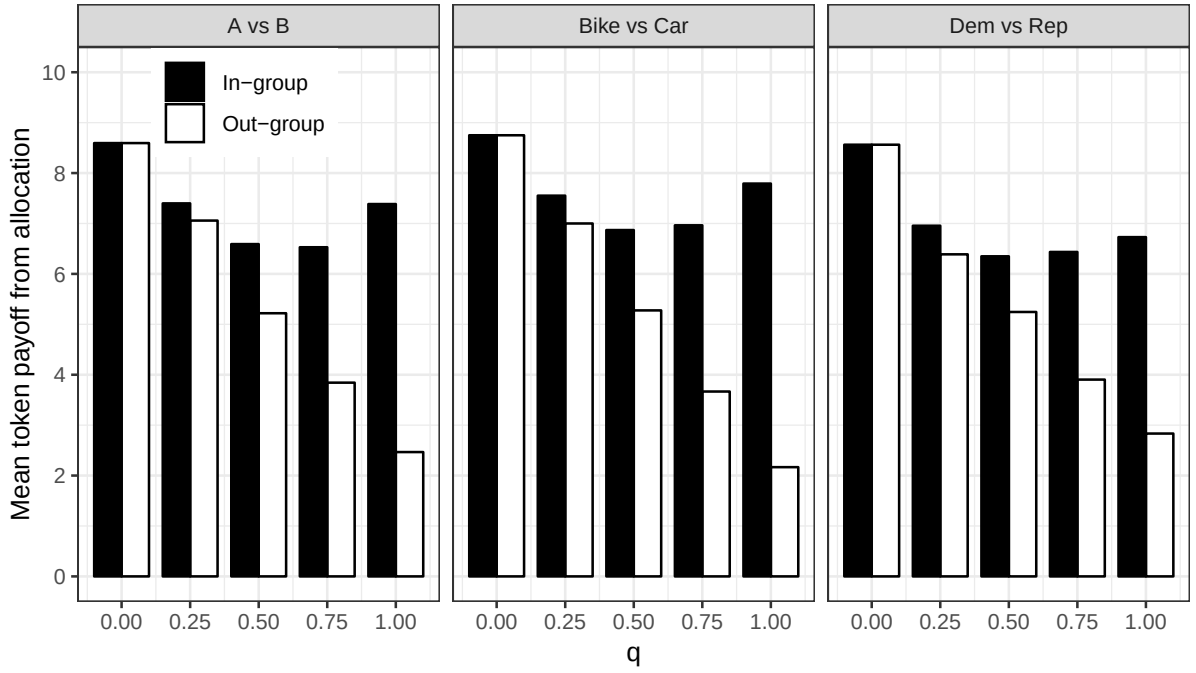


Figure D.2: Average token payoff from the allocations by in-group and out-group candidates (Difference in average token payoffs from the allocations by in-group and out-group candidates is Δ^x) plotted over policy polarization (q) by identity treatment.



D.4 Voter election decisions

Figure D.3: Distribution of the maximum relative payoff from the ability of the out-group candidate ($\tilde{\Delta}^{\alpha}$) by identity treatment.

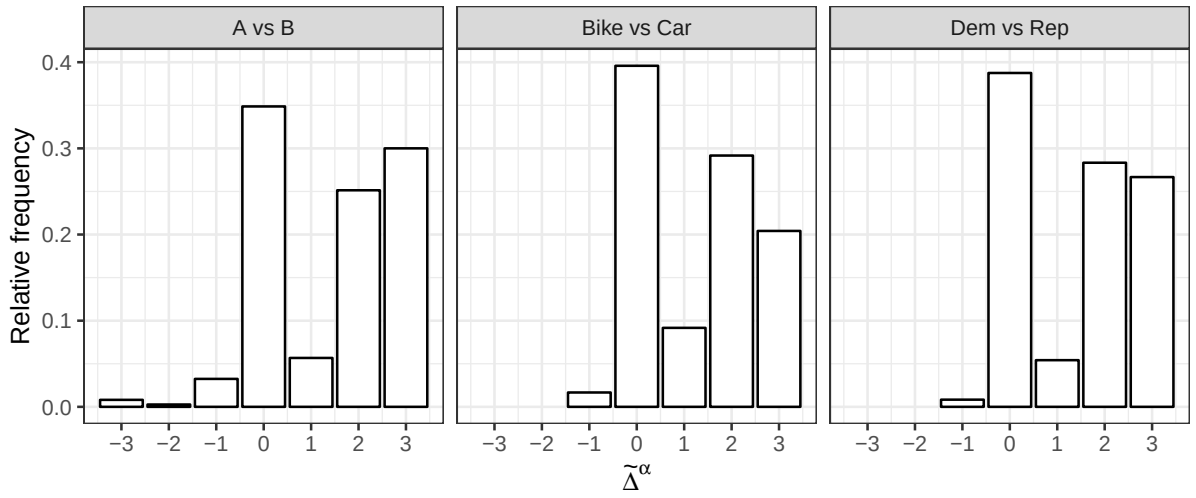


Figure D.4: Mean of the maximum relative payoff from the ability of the out-group candidate ($\tilde{\Delta}^{\alpha}$) voters accept to vote for the in-group candidate plotted over policy polarization (q) by identity treatment.

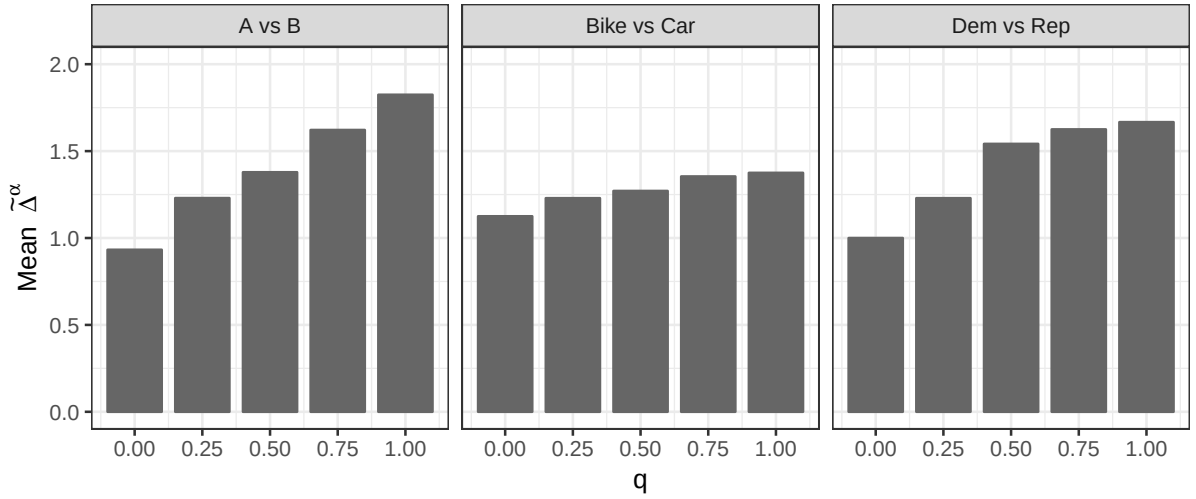


Figure D.5: Distribution of voters' minimum relative payoff from the ability of the out-group candidate ($min(\tilde{\Delta}^{\alpha})$) by identity treatment.

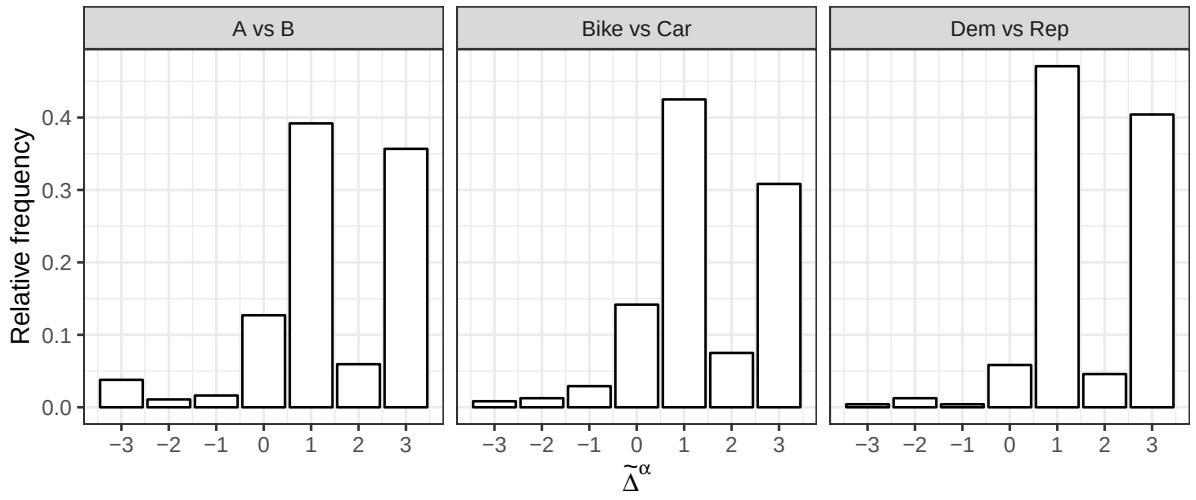
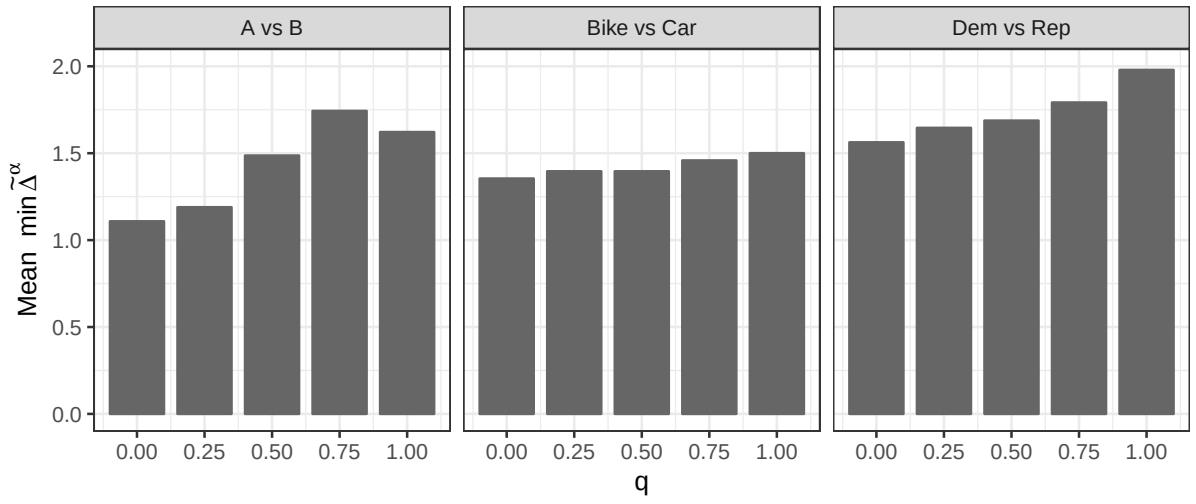


Figure D.6: Mean of the minimum relative payoff from the ability of the out-group candidate ($min(\tilde{\Delta}^{\alpha})$) voters accept to vote for the out-group candidate plotted over policy polarization (q) by identity treatment.



D.5 Multi-identity treatment

Our finding of instrumental motivations for group identity-driven behavior may be undermined by the weakness of identities induced in the laboratory: for any expressive motivation to surface, a stronger, more contextualized identity may be needed. We implement the Bike vs Car-, Dem vs Rep-, and multi-identity-treatments exactly for the purpose of inducing, even if only marginally, stronger identities. Partisan voting and partisan allocation emerges as well, as expected, when inducing stronger identities but we do not see a large increase in the strength of these in-group biases in vote choices and allocation decisions.

A more valid test of whether subjects attach higher salience to a stronger identity may be to put two dimensions of group identities that supposedly differ in their strength side-by-side and let subjects choose which one they emphasis in their decision-making. We do so by letting subjects decide whether they want to put more emphasis on the distinction A vs B or Bike vs Car, as we chose to implement in our multi-identity treatment. In this treatment, subjects are

characterized by a compound identity: A-Bike, B-Bike, A-Car, or B-Car. We find, first, while partisan voting for candidates who share the exact identity with the voter is larger than the one for those who share the A/B- or Bike/Car-identity, in-group allocations are smallest for this candidate. Second, and more importantly, there is, again, a (weak) increase in partisan voting and in partisan allocations with increasing policy preference polarization.

In this multi-identity treatment, voters are labelled A-Bike-, B-Bike-, A-Car-, or B-Car-voters according to their group assignment. Assignment to *Bike* or *Car* takes place before subjects learn the content of the voting game and assignment to A or B occurs after subjects received instructions for the voting game but before the voting game commences. Candidates are asked to allocate up to 16 tokens to the five positions of a preference space: four positions in the extreme corners, *Left-Up*, *Right-Up*, *Right-Down*, or *Left-Down*, and one position at the *Center*. Voters decide whether to vote for one of two candidates in two possible pairings of candidates: *A-Bike* vs *B-Car* or *A-Car* vs *B-Bike*. The probability with which voters are at the extreme of the preference space, that is the degree of policy polarization q, is either .25 or .75. Voters are, again, also told the ability and group membership of both of the candidates. The decision environment is described not only by q but also by which pairing of candidates they can choose from: A-Bike vs B-Car or A-Car vs B-Bike. q is fixed for 18 rounds while the pairing of candidate is the same for 9 rounds (while the level of ability varies over those 9 rounds). For payoffs, a decision situation is picked at random: it is represented by a realisation of q, candidate abilities, and a candidate pairing. If assigned an extreme ideal point, voters receive 1 token for each 1 token allocated to the extreme of the preference space associated with the group of the voter and .5 token for each 1 token allocated to a position in the preference space contiguous to the extreme associated with their group. If assigned an ideal point in the center, voters receive 1 token for each 1 token allocated to the center and .5 token for each 1 token allocated to any of the extreme positions in the preference space. For example, when assigned an extreme ideal point an A-Bike-voter receives 1 token from each 1 token allocated to Left-Up and .5 token from each 1 token allocated to Left-Down, Right-Up, and Center. Similarly, A-Car-, B-Bike-, and B-Car-voters with such an extreme ideal point receive a full 1 token from each 1 token allocated to their extreme corner but only .5 token from each 1 token allocated to the corners contiguous to theirs and to the Center. Note, in the multi-identity treatment, social efficiency is maximized by allocating all tokens to the center. This is most clearly illustrated in the extreme case of $q = 1$. In this case, a token allocated to an extreme point gives a total payoff of 2 tokens (1 to the individuals at the extreme point and 1/2 to each adjacent extreme), while a token allocated to the center gives a total payoff of 2 (1/2 to all individuals).

In 2 sessions with 24 subjects each, we collect, for each subject, 36 observations as voter and 4 observations as candidate. In total, we collect 1728 voter-round observations in the multi-identity treatment. The total number of candidate-round observations is 192. The average earnings in Euro were 18 and 16.6 in the two sessions, respectively.[25] The group identity inducement procedure generated 17 subjects in group A/Bike, 7 in A/Car, 16 in B/Bike, and 10 in B/Car. That is half of the subjects were assigned to Group A and 65% to group Bike.

The vote share of the A-Bike over the B-Car candidate was .55 (.50) and of the A-Car candidate over the B-Bike candidate was .41 (.49). Subjects allocated, on average, 2.76 (2.87) tokens to the Left-Up corner, 2.21 (2.49) to Right-Up, 5.31 (4.35) to the Center, 2.26 (2.42) to Right-Down, and 2.51 (2.71) to Left-Down.

Figure D.7 and D.8 compare in-group voting for and in-group allocations of candidates who

[25] In session 1 of this treatment, voters choices were not recorded correctly in the payment file. We therefore paid the for subjects in the role of voters maximum possible earning of 16 tokens slightly increasing the average payoff of that session relative to the other session of that treatment. Subjects only learned at the moment of payout that this error occurred, in this way, their choices throughout the experiments could not have been affected.

share the exact identity of the voter, those who share their A/B-identity, and those who share their Bike/Car-identity. First, while partisan voting for candidates who share the exact identity with the voter is larger than the one for those who share the A/B- or Bike/Car-identity, in-group allocations are smallest for this candidate.

Figure D.7: Rate of partisan voting by policy preference polarization and type of shared identity – exact, A/B, or Bike/Car-identity – for negative ability difference between in- and out-group candidate

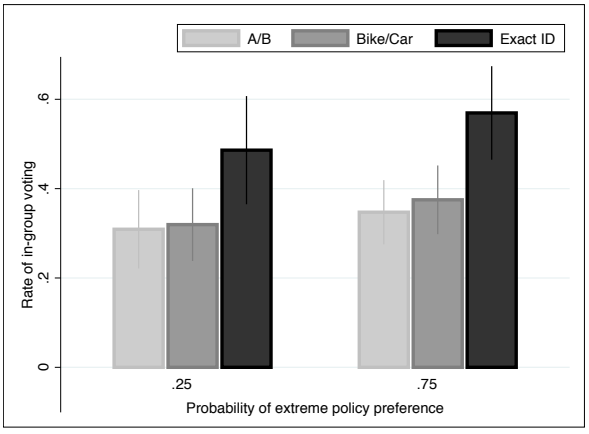
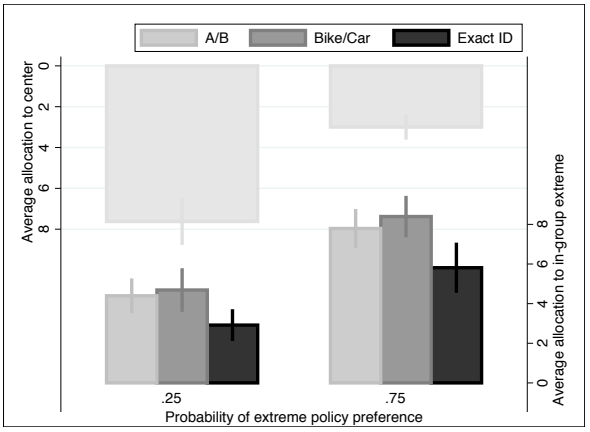


Figure D.8: Rate of allocation to extreme and center by policy preference polarization and type of shared identity – exact, A/B, or Bike/Car-identity



Second, and more importantly, there is, again, a (weak) increase in partisan voting and in partisan allocations with increasing policy preference polarization. The change in partisan voting, however, is not significantly different from zero, i.e., the marginal effect of raising the level of policy preference polarization from .25 to .75 heightens the rate of voting for the in-group candidate who exactly matches the voters identity by .03 % (-.03,.08). In contrast, more polarization in policy preferences systematically increase allocations to all types of in-group voters, i.e., those who share an exact, A/B, or Bike/Car-identity with the candidate. A change in polarization from .25 to .75 raises in-group allocations by 2.53 (1.66,3.41) of the candidates who share the exact identity, 3.11 (2.21,4.00) of those who share the A/B-identity, and by 3.33 (2.38,4.28) tokens of those who share the Bike/Car-identity. More polarization also decreases allocation to the center by 4.43 (3.29,5.57) tokens.

D.6 Subject-level analysis

We now turn to the question whether the observed patterns are robust to subject-level heterogeneity. In other words, are results driven by a small subset of subjects or do we observe a general tendency in behavior. We find, in the aggregate, that in-group favoring allocation decisions by candidates (Δ^x) and partisan voting ($\tilde{\Delta}^\alpha$) is increasing in the level of policy polarization q. To characterize subjects-level behavior, we investigate whether the distribution of individual-level choices follows the overall patterns of Δ^x and $\tilde{\Delta}^\alpha$ in its relationship to q narrowly. We find that it mostly does with respect to candidate allocations but the variation of partisan voting around the treatment mean is less narrow. Here subject-level heterogeneity arises but our overall results are still robust to modeling subject-level idiosyncrasies and can be traced to subjects best responding to their own allocation decisions.

More specifically, in a subject-level regression of Δ^x on q we estimate a significantly positive relationship for 41% and a positive relationship for 41% of subjects. (Recall we collected 5 allocation decisions for each subject, one for each level of policy polarization.) The coefficient estimate on q is negative and not significant for the remaining 18% of the sample, only.

We clearly find more subject-level heterogeneity in the relationship between $\tilde{\Delta}^{\alpha}$ and q. While we estimate a positive subject-level slope in the relationship $\tilde{\Delta}^{\alpha}$ and q for 70% of subjects, only 14% of the coefficient estimates associated with q are significant. On the other hand, of the estimated negative coefficients for the remaining 30%, the subject-level regression of $\tilde{\Delta}^{\alpha}$ on q returns a significantly negative coefficient estimate on q for only 1% of subjects. (Recall we collected 45 voting decisions for each subject, one for each combination of candidate ability pair and level of policy polarization.)

At this point, it is relevant to ask what drives subject-level heterogeneity in the relationship between partisan voting and policy polarization. A reasonable place to start that investigation is the question whether $\tilde{\Delta}^{\alpha}$ varies because of variation in subjects' beliefs about candidate allocations. A good approximation of what voters believe about candidates' behavior is how they themselves chose to allocate. 30-40% of subjects always vote for an in-group (out-group) candidate whenever ability difference and their own allocations would give them higher utility than voting for the out-group (in-group) candidate. About 90% of subjects best respond in at least 2/3 of their choices. This is indicated by the dominance of the light gray area.

Further, these subject-specific best responses are stable across policy polarization and therefore do not bias the overall trend of $\tilde{\Delta}^{\alpha}$ increasing in q. Figure D.9 shows that the profile of strategies within-subject does not change with increasing policy polarization q. For the bulk of subjects, increasing q by .25 increases the share of any of the strategies by less than .01.

Figure D.9: Distribution of estimated change in share of subject choice type (best responding to own allocations, should vote for in-group but votes for out-group, should vote for out-group but votes for in-group) when policy polarization q increases. The reported statistic is the coefficient estimate from a regression of share of subject choice type on$q$..