Social Identity and Electoral Accountability

Dimitri Landa   New York University
Dominik Duell   New York University

In a laboratory experiment, we explore the effects of group identities on the principal-agent relationship between voters and representatives. In an adverse selection framework with observable effort, voters can choose to condition their reelection choices on representatives' effort alone, beliefs about representatives' competence, or both of those jointly. We show that inducing social identities increases the weight of representatives' effort in voters' reelection decisions. Further, when voters and representatives share a social identity, representatives tend to invest less effort and their effort is independent of their competence. In contrast, "out-group" representatives compensate for lower competence with higher effort and reduce effort when voters are likely to perceive them as competent. Voters often adopt laxer retention standards for representatives who are fellow group members and are responsive to evidence of other-regardingness from out-group representatives, but some voters actively resist treating representatives with shared identity more favorably and "overcorrect" as a consequence.

A key feature of the political process that may be expected to influence electoral accountability is the social identities of voters and politicians. Yet, despite the proliferation of ever more nuanced accounts of electoral control, the political economy of accountability has little systematic to say about the effects of social identities. In this article, we present a laboratory experiment that explores the effects of shared and unshared social identities on representatives' efforts on behalf of voters, and on voters' choices in response to them.

Our point of departure is the recognition of tensions entailed in the distinct and sometimes mutually contradictory effects ascribed to social identities. On one hand, sharing voters' social identity is often understood to give a representative an electoral advantage that can lead her to slacken her efforts on behalf of the voters. By this logic, sometimes invoked on behalf of white candidates

running in majority black districts, a candidate with unshared identity should, all else equal, be preferred because she will work to earn the electoral support that a candidate with a shared identity will take for granted. On the other hand, elected officials are sometimes thought to enjoy a "warm glow" from making efforts on behalf of voters who share their social identities, and so to work harder to deliver more favorable policy outcomes. By this logic, voters should prefer to elect "one of their own." These arguments regularly reappear in electoral campaigns for different levels of government.[1] Assuming that voters are, indeed, affected by them, these arguments about representatives' behavior may be seen as predictive with respect to voters' choices as well.

While both of these arguments may have some resonance, and voters are frequently found to assign an important role to social identities in evaluating

[1]Recent notable showings reaching national headlines include congressional elections in the predominantly black 9th District in Tennessee and the Democratic primary in the mayoral election in Chicago. In Tennessee, explicit arguments about race were a key factor in the 2006, 2008, and 2010 campaigns, in which Steve Cohen, a liberal Democrat and Caucasian, was elected running against African American opponents. In Chicago, race was also the main dimension on which African American opponents of Rahm Emanuel sought to distinguish themselves in 2010.

DOI: 10.1111/ajps.12128

candidates (Chandra 2004; Kaufmann 2004), the recent wave of methodologically sophisticated empirical studies of representatives' behavior is in broad agreement that social identity lacks obvious effects on representatives' performance (Ferreira and Gyourko 2009, 2011; Gerber and Hopkins 2011; Hopkins and McCabe 2012). A key to explaining this puzzling conjunction may be that the focus of the debates on majority-minority representation—representatives' effort—is but one ingredient in the policymaking process. Recent developments in the political economy of electoral control underscore the importance of considering how representatives' quality (typically, their underlying competence) may affect individual behavior in identity-primed settings. Indeed, one should expect representatives' competence to be a key factor determining precisely when and whether arguments in favor of in-group or out-group candidates have bite. Addressing these questions empirically, however, faces significant challenges since it requires accounting for competence and effort separately, even as they interact in generating the incumbent's performance. These challenges underscore the appeal of a controlled laboratory experiment. To grasp the relative influence of competence versus effort in interaction with shared group membership as the determinants of representatives' and voters' choices, it is desirable to focus on an environment in which candidates can vary in quality and exert effort, and voters can account separately for each. With this in mind, our laboratory experiment is based on an adverse selection game between a voter and a representative (described in the third section), in which the representative's effort choice is observable. Effort observability implies that voters can condition their reelection choices on effort alone, ignore effort entirely and condition them on their beliefs about representatives' competence, or condition them on those beliefs and effort jointly. Representatives can, similarly, behave in distinct ways that may rationalize those strategies. To analyze the effects of social identities in a controlled environment, we induce artificial identities rather than priming naturally existing ones; as discussed in detail in the fourth section, which details the experimental setup, we do this in a way that creates relatively weak group identification, biasing our results in the direction of understating the true effects of social identity.

Our analysis of average treatment effects (in the fifth section) focuses on the comparison of subject behavior both between an identity-free baseline treatment and an identity treatment (i.e., the effects of identity priming as such) and between identity-matched (*in-group*) and unmatched (*out-group*) pairings of subjects as voters and as representatives in the identity treatment. We show that the presence of social identities reduces the willingness of individuals to make voting decisions based on the competence of representatives and increases the relative weight of effort in determining reelection. When voter and representative share a social identity, the representative puts in less effort, and that effort is independent of his or her competence. In contrast, representatives who do not share the social identity of the voters tend to choose their effort levels in a competence-dependent way, seeking to make up for lower competence with heavier investment into effort, but reducing effort when they are likely to be perceived as highly competent. In short, we find that social identities have significant effects on subjects' choices in the electoral principal-agent environment, but that these choices are fundamentally mediated by representatives' competence and, as such, display features beyond the scope of the existing accounts of the effects of social identity.

Our analysis of individual-level behavior in the subsequent section provides further insights into the mechanisms by which social identity relationships shape individual choices. We find that a substantial proportion of subjects display significant in-group bias—a systematically greater willingness to retain representatives in in-group than in out-group matches. Our evidence suggests that these differences stem from fundamentally different approaches to evaluating representatives' performance in in-group versus out-group matches. In in-group matches, such voters' behavior appears driven by a version of "warm glow" toward fellow in-group members, manifesting in a general tendency to forgive both low effort and low competence. In out-group matches, the same voters adopt a more consequentialist, expected-utility-driven approach to evaluating their (out-group) representatives, consistent with seeking a threshold level of utility or searching for costly signals of out-group members' willingness to invest in effort when electoral oversight is slackened. However, we also find that some of the more cognitively sophisticated voters resist in-group favoritism and "overcorrect" for it in adopting higher standards when judging the performance of representatives who belong to their social group.

As we discuss in the last section, the overall pattern of our findings suggests a way of reconciling conflicting claims about the benefits and liabilities of descriptive representation and sheds light on empirical puzzles emerging from observational studies.

Electoral Accountability and Social Identity

Whether and how voters are able to exercise effective electoral control over their representatives is the

subject of considerable debate. According to the *moral hazard* perspective on accountability, voters evaluate past performance retrospectively: setting an ex ante threshold of acceptable performance and conditioning their reelection decision on whether the politician meets it (Ferejohn 1986; Fiorina 1981). However, forward-looking voters may be better off selecting whomever they believe to be a more competent candidate (Fearon 1999), and in the *adverse selection* framework that captures this intuition, voters base their reelection decisions primarily on their beliefs about the representative's underlying outcome-relevant "type" (e.g., competence or personal ideology)— beliefs that are, in turn, informed by the politician's performance in office (Banks and Sundaram 1998). Several recent papers (Ashworth, Bueno de Mesquita, and Friedenberg 2010; Landa 2010; Woon 2012) suggest that the moral hazard incentives to sanction and the adverse selection incentives to retain based on beliefs about the representative's type may coexist in the same strategic environment. The closest to the present study, Landa (2010), shows that, in a laboratory experiment with unambiguous adverse selection incentives, voters sanction low effort and retain low-competence incumbents who exert sufficiently high effort.

Empirical scholars of political behavior view voters as having limited abilities to process abstract and systematic information to make consistent political decisions on their own, and they assign voters' social identity a determinative role in motivating their political choices.[2] Extensions of this view argue that voters may use social identity as a low-cost informational cue (Chandra 2004; Lupia and McCubbins 1998, 2000), more valuable to them when they are otherwise more ignorant (Morton, Williams, and Bassi 2011). Because social identities can play such a role, politicians may have incentives to prime voters to focus on their group identity at the expense of considerations of candidate competence or office performance (Dickson and Scheve 2006; Mendelberg 2001; Metz and Tate 1995). While our work is related to these studies, we seek to explore a distinct question: how does the presence of social identity considerations affect the strategic interplay between representatives' effort choices in office and voters' reelection decisions?

Following the work of social psychologists, "social identity" may be defined as an aspect of a person's sense of self that derives from perceived membership in a social group. Theoretical elaborations on this account typically conceive of group identity as implying psychological

benefits from acting in accordance with group norms (Akerlof and Kranton 2000, 2010). In line with this reasoning, in political settings representatives are said to experience a "warm glow"—a psychic utility boost—when behaving favorably toward in-group voters (Swain 1993). Experimental literature in social psychology and economics finds in-group favoritism driven by warm glow in a broad range of social interactions (Bernhard, Fehr, and Fischbacher 2006; Chen and Li 2009; Goette, Huffman, and Meier 2006; Tajfel 1981; Tajfel and Turner 1986), though its implications for principal-agent settings are controversial. McLeish and Oxoby (2007) find that because individuals expect warm-glow behavior from fellow group members, they are more willing to punish those members when treated unfavorably by them than they would out-group members, whereas Chen and Li (2009) report results conforming to the more conventional view that in-group bias triggers behavior that is simply favorable toward individuals with shared group membership. Further, outside the strategic framework, recent work in social psychology provides evidence of subjects' behavior that is, in fact, favorable toward stigmatized out-groups. In trying to appear egalitarian, such subjects are "overcorrecting," effectively leaning toward preferring out-group individuals (Devine et al. 2002; Harber 1998; Peffley, Hurwitz, and Sniderman 1997; Plant and Devine 1998). The overcorrection weakens when individuals lack cognitive resources necessary to engage in the effort of checking in-group bias (Mendes and Koslov Forthcoming, Pearson, Dovidio, Phills and Onyeador Forthcoming) and strengthens with awareness that actions favoring their own group might be observed by a third party (Green, Pallin, Raymond, Iezzoni, Carney, Ngo and Banaji 2007).

Scholarship on group behavior also finds considerable evidence of other examples of social preferences, driven by concerns with fairness and reciprocity (Bolton and Ockenfels 2000; Engelmann and Strobel 2004; Fehr and Gaechter 2000; Fehr and Schmidt 1999; Kahneman, Knetsch, and Thaler 1986; Rabin 1993). However, unlike warm-glow accounts, those concerns can often be rationalized without assuming a primitive attachment to fellow group members, and the analysis of those preferences has not focused on their connection to the effects of social identity.

A Simple Model of Electoral Politics

We place our analysis in the context of an adverse selection (learning) model of electoral accountability with observable effort. A voter faces a representative with privately

[2]Compare Lazarsfeld, Berelson, and Gaudet (1944); Campbell et al. (1960); Converse (1964); Lipset and Rokkan (1967); and Beck et al. (2002).

known competence, modeled as her *type t,* which is commonly known to be uniformly distributed on $[\underline{t}, \bar{t}]$. Before the election takes place, the representative chooses her effort level, $e_1 \in [\underline{e}, \bar{e}]$, the effort being beneficial to the voter and costly to the representative. The voter then observes the value of e_1 and her own utility $f(e_1) + t + \omega_1$, where $f(\cdot)$ is a commonly known function and ω_1 is a random draw from a commonly known symmetric distribution Ω with mean zero. Having received this information, the voter decides whether to keep the representative in office.

If the voter chooses to reelect the representative, the representative determines his post-election investment into effort, $e_2 \in [\underline{e}, \bar{e}]$, giving the voter an additional payoff of $f(e_2) + t + \omega_2$, where ω_2 is a random draw from Ω . If the voter chooses not to reelect the representative, the voter receives instead an additional payoff of $f(\underline{e}) + E(t) + \omega^*$, where ω^* is a random draw from Ω. If the representative is reelected, she receives a payoff of $2B - e_1 - e_2$, where B is a fixed valuation of being in office; if he is not reelected, she obtains a payoff of $B - e_1$. The game ends and these payoffs are realized after the representative chooses e_2 or after the voter chooses not to reelect the representative.

Experimental Treatments

The experiment instantiates the interaction between voters and elected representatives closely following the model described above, motivating subject reasoning with performance-based payments. The baseline treatment does not include any identity priming. In the identity treatment, we artificially induce group identification for voters and representatives they are matched with. After the end of each session, subjects answer a series of questions about their demographics and about their choices throughout the experiment.[3] In three sessions of the baseline treatment and four sessions of the identity treatment, we collected observations on a total of 122 participating subjects. Each experimental session lasted between 14 and 20 rounds, with 14–20 participating subjects.

[3]In an exit survey, subjects were asked whether they found the instructions clear, whether the tasks they were asked to perform were easy, and whether they thought other subjects found them to be easy. Almost all respondents provide affirmative answers. Subjects' incentivized choices also support the view that they had a good understanding of the key elements of the game: Because representatives' incentives with respect to post-election effort are invariant across rationalizable strategy profiles, they should choose the lowest levels of effort once they no longer face reelection. As the details below make clear, while there are interesting variations, they do so overwhelmingly.

Baseline Treatment

At the beginning of the first round of the voter-representative game, subjects are randomly assigned the roles of *Voter* or *Representative.* These roles remain fixed for the first half of the session; in the second half of each session, roles are reversed. At the beginning of each round, subjects are randomly matched into voter-representative pairs.

Subjects in the role of a representative are randomly assigned a number drawn from a uniform distribution on the set of integers from 20 to 50. In the instructions, this number is referred to as the representative's *true number,* and below, for ease of reference, as her *type.* Representatives then select a number, *choice 1* (below sometimes referred to as their *pre-election effort*), from the set of integers from 5 to 20. Following that selection, *choice 1* and *choice 1 consequence,* defined as the sum of the pre-election effort, the representative's type, and a realized noise draw (described as a "random bump") from a uniform distribution on $[-15, 15]$, are made public to the voter and the representative within that pair. The voter then chooses whether to reelect the representative. If reelected, the representative then selects a value of *choice 2* (below sometimes referred to as his or her *post-election effort*), again from the set of integers from 5 to 20.

If a voter decides to retain the incumbent representative, her payoff is calculated as the sum of *choice 1 consequence* and *choice 2 consequence,* which is computed similarly, but based on the value of *choice 2* and a separate noise draw. If a voter decides not to retain the incumbent representative, her payoff is calculated as the sum of *choice 1 consequence,* the (prior) expected value of representative's type (35), the equilibrium expected *choice 2* (5), and a new realized noise draw from the uniform distribution on $[-15, 15]$.

Representatives' payoffs in each round are $(80 - choice\ 1 + 80 - choice\ 2)$ if retained and $(80 - choice\ 1)$ if not retained, where 80 can be interpreted as the value of being in office.

To avoid making additional assumptions about the quality of updating following the receipt of the noisy signal of representative's type, $choice\ 1\ consequence - choice\ 1 = type + noise$, we report results in relation to $type + noise$, which, for convenience, we refer to as $type_\omega$ throughout the article.

Identity Treatment

At the beginning of each session of the identity experiment, subjects are shown five pairs of paintings, with each

pair consisting of one painting by Paul Klee and one by Wassily Kandinsky. Subjects are then asked to choose their preferred painting in each pair. Based on which painter's work a subject prefers in a majority of the five pairs, he or she is assigned to be a *Klee* or a *Kandinsky*.[4] Once identities are assigned, subjects participate in a quiz in which they are asked to identify the painter (Klee or Kandinsky) of three further paintings. In answering the question about each of those paintings, subjects give initial guesses that are made available to other subjects in the same identity group before everyone is asked for their final answer. Subjects within a group receive 100 tokens if the majority of their group members name the correct painter in the final answer.

In the subsequent voter-representative game part of the experimental session, the group identity of each subject within a matched pair, Klee or Kandinsky, is displayed for both of them on the screen. Otherwise, the treatment proceeds identically to the baseline. Note that in the realized matching draws in the experiment, subjects were assigned to in-group matches about as often as to out-group matches (see Table A2 in Appendix A).

As implemented, the identity treatment is related to the *minimal group paradigm* (Tajfel and Turner 1986),— an approach to inducing a (weak) notion of identity that is seemingly unrelated to the behavior of interest. A key argument for that approach is that it allows the experimenter to avoid uncontrolled associations of identity with the particular choices available to the subjects. Moreover, insofar as weak identities are also priming residual out-of-the-lab group experiences, they do so in a way that allows subjects to self-select from such experiences those that seem to them most relevant to the group experience modeled in the lab—something the researcher cannot herself do for the subjects.

Considerable experimental literature using the minimal group paradigm has shown its effectiveness in inducing patterns of responses to identity, including in-group favoritism and inter-group competition, that resemble those usually observed outside the laboratory with naturally occurring group identities.[5] In particular, utilizing the minimal group paradigm approach, our ex-

periment allows for subjects' inclinations to behave in a group-biased way to reveal themselves but also makes it possible for subjects consciously to control and correct their biases—both phenomena characteristic of outside-world identities.[6] As such, this approach opens the door for a range of inferences about the shape of these effects in a relationship between voters and representatives. The effects of artificially induced weak identities increase with the salience of identities (Charness, Rigotti, and Rustichini 2007; Chen and Chen 2011; Eckel and Grossman 2005); operationally, a key factor that raises such salience is interactions with fellow group members in performing joint tasks, such as the group quizzes we administer as part of our identity treatment.

Note, importantly, that our treatment instantiation and, in particular, the explicit communication of induced social identities to subjects within the matched voter-representative pairs mean that subjects' behavior is likely to reflect both the automatic responses to identity as well as subjects' perceptions about what the proper response to social identities should be (some of which may, of course, be themselves internalized and automatized to various degrees). The identity-related behavior we characterize should, thus, be interpreted as reflecting the subject's bundle of attitudes to social identity. In the world outside the lab, on top of those attitudinal effects of identity, there are, surely, others, including effects driven by institutional incentives of electoral rules, divided government, and so forth; the attitudinal effects may be thought of as a kind of baseline for those effects: It is a measure of how subjects perceive identity without additional explicit payoff pressures.

Predictions

Social identities enter our subjects' reasoning in voter-representative matches in a way that is not driven by the explicit payoffs assigned to them. As indicated above, this allows us to elicit effects of identity, including subjects' responses to identity, without imposing them on

---

[4] See Tajfel et al. (1971) and Chen and Li (2009) for the use of painter preferences to induce identities.

[5] Behavior driven by weak identities can, in this sense, be seen to approximate behavior ascribed to the effects of strong, more contextualized, or previously existing group identifications. Thus, the experimental literature provides evidence that "weak" induced identities significantly affect subject behavior with respect to individual shirking and free riding (Eckel and Grossman 2005), cooperation, and willingness to reward or punish (Bernhard, Fehr, and Fischbacher 2006; Chen and Li 2009; Goette, Huffman, and Meier 2006; McLeish and Oxoby 2007). Eckel and Grossman (2005) and

Goette, Huffman, and Meier (2012) provide evidence that the effects of identity being induced are monotone in the strength of that identity (i.e., the weakness of identity inducement does not bias results in the wrong direction).

[6] Eliciting such responses to social identities with more subtle priming in settings with artificially induced identities would be more difficult, and null results on the effects of identity would be explainable by the weakness of the artificially induced identities, rather than by the irrelevance of identities as such. In contrast, when the research question is about the implication of the specific, well-defined identities, a subtle priming of those preexisting identities (e.g., Habyarimana et al. 2007) would be more appropriate.

the subjects by directly and exogenously tying identity biases to their payoffs. With this in mind, we discuss below two sets of behavioral expectations: one coming out of the Nash-based equilibrium predictions for our underlying electoral politics game, and the other implied by the empirical, including social psychology, work on the behavioral implications of social identity.

Equilibrium Predictions. In any perfect Bayesian equilibrium of this game, the voter retains the representative if and only if her posterior on the representative's *type* is greater than or equal to $(\bar{t} + \underline{t})/2$. The intuition is as follows. The representative's optimal action after the election is to choose $\underline{e}$, independent of his *type*. Anticipating this, the voter should condition her reelection decision only on the expectation of the representative's type, retaining the representative if and only if she believes his type to be at least $(\bar{t} + \underline{t})/2$.

Because the representative's effort is perfectly observable, by subtracting the observed pre-election effort from her realized pre-election utility, the voter can arrive at the value of $t + \omega_1$, and from that value, given the distribution of ω_1, at the value of the incumbent's expected type. If the voter conditions the reelection decision on that value alone, the representative has no incentive to invest more than the minimal amount into e_1. This corresponds to an equilibrium in which effort is irrelevant for reelection, and voters effectively ignore it in their electoral decisions. We refer to this (type-pooling) equilibrium as the *no-effort equilibrium.*

There is another perfect Bayesian equilibrium in which representatives can send costly signals of their type by choosing higher values of effort. In this equilibrium, the voters condition their retention choices on both the observed effort and the observed type. For this equilibrium, there exists a subset of the type space such that within that subset, higher types of representatives choose higher effort than lower types. For the lower types, the probability of their receiving a noise draw that leads to an outcome high enough to cause the voter to retain, conditional on $e_1 = \bar{e}$, is too low to justify the additional effort. We refer to this semi-separating equilibrium as the *type-signaling equilibrium.*

The environment described in our baseline game is "identity-free." When we prime and reveal to group members their social identities in the identity treatment environment, we do so without altering this payoff structure. One equilibrium behavioral expectation (denoted below as *BE1*) is, thus, that identity has no effect on behavior. However, because players observe social identity matches and there are multiple identity-free equilibria, the game with the identity treatment also admits "meta" equilibria

in which both equilibrium profiles from the no-identity environment are played, with the profile choice determined by the identity match (e.g., no-effort profile in in-group matches and type-signaling profile in out-group matches). In this way, identity matches could matter as "switchers" between type-signaling equilibrium and no-effort equilibrium play (*BE2*).

Behavioral Expectations Associated with Social Identity outside Equilibrium Play. As indicated above, the predictions for representatives' behavior arising from the debates about the benefits of majority-minority representation run in opposite directions. The expectations driven by the strategic responses to in-group favoritism suggest that *in-group representatives will choose lower effort than out-group representatives (BE3.a)*, and expectations driven by a warm-glow payoff from higher effort in in-group matches imply that *in-group representatives will choose higher effort than out-group representatives* (BE3.b).

Our design separates competence and effort and allows for a more precise statement of how group identities affect behavior. In particular, we can condition our expectations of that behavior on the behavior in the identity-free baseline, in which low types choose high effort to compensate for their competence deficit and high types choose low effort because they expect to sail through on competence alone (Landa 2010). If there is a variation across in-group and out-group matches, it must be that, expecting in-group favoritism from the voters, it is the low-competence representatives who would drive the result consistent with *BE3.a* and, on the basis of a warm-glow attitude toward the voters, it is the high-competence representatives who would drive *BE3.b*. Note that, circumscribed in that fashion, these expectations—call them *BE3.a′* and *BE3.b′*—are no longer conflicting.

In line with a broad range of results from other studies, we expect *in-group favoritism on the part of the voters, manifesting in higher retention rates in in-group matches than in out-group matches (BE4).* Part of our goal in evaluating this expectation is to ascertain its scope in our experiment. One specific point of interest is what happens at low levels of effort: in line with results from Chen and Li (2009), one may expect a lower willingness to punish in-group representatives and thus higher rates of retention than for out-group representatives at low effort. In contrast, consistent with McLeish and Oxoby (2007), we may not see higher retention rates for in-group representatives at low effort, given the higher willingness to remove in-group representatives from office if they do not deliver high enough effort.

A different instantiation of in-group bias, reflecting an implicit willingness to accept a trade-off between

effort and competence and a more conditional retention strategy for in-group representatives, is the expectation that *voters set outcome thresholds for retaining representatives and choose a lower threshold in in-group matches than in out-group matches (BE5).* Such a threshold may result from a reciprocal agreement (exchange) between a representative offering a mix of effort and competence and a voter offering retention.

A different line of reasoning gives rise to expectations of the consequences of identity that invoke previous work on other-regarding preferences. As indicated above, other-regardingness may come in the form of voters' distributional preferences (such as when voters may seek to prevent effort-shirking representatives from enjoying utility levels that are too high relative to their own). But other-regardingness may also be interpreted as a distinct unobservable aspect of representatives' types (e.g., the aspect of type that prefers a higher investment into effort in the interest of increasing the welfare of the voters). A costly signal of such type in our setting is high effort when the representative's expected competence level is high (a high-competence representative may otherwise have expected to be retained for that reason alone). Voters may reward representatives demonstrating such behavior with higher retention rates not merely because they prefer to reward other-regarding representatives, but because higher-than-minimal investment into *choice 2* is not incentivized, and evidence of a representative's other-regardingness is a signal that that representative would choose higher-than-minimal values of *choice 2* (see Gneezy et al. 2012 for an account of costly signaling of prosocial behavior). This brings us to the implications of in-group bias. If voters are anticipating favored treatment from in-group representatives or are feeling more generous toward them, they may be less demanding of such signals in in-group than in out-group matches. Put differently, we may expect that *in-group-biased voters may reward effort more highly in out-group matches than in in-group matches* (BE6).

Finally, as indicated above, cognitively sophisticated individuals may not only intentionally ignore identities, but also may, perhaps subconsciously, overcorrect against the in-group favoritism associated with in-group bias. Thus, we may expect that, *holding fixed levels of type and effort, voters retain out-group representatives more often than in-group representatives when those voters are more cognitively sophisticated (BE7).*

Average Treatment Effects

Because our primary interest is in how the introduction of identities alters subject behavior, the focus of our analysis is on the treatment effects, and we provide the evaluation of the equilibrium predictions in that context.

Retention Decision

The first treatment effect we report concerns the relative importance of representatives' competence in voters' retention decisions. Recall that in the no-effort equilibrium, retention decisions must turn completely on the voters' beliefs about representatives' competence, with only expected high-competence types being retained. Here we find that while competence clearly matters for retention across treatments, subjects in the identity treatment generally pay significantly less attention to competence when making their retention decisions than subjects in the baseline treatment.

Figure 1 A gives the treatment-specific difference between proportion of representatives retained with high $type_\omega$ ($t_\omega \geq 35$) and the proportion of representatives retained with low $type_\omega$ ($t_\omega < 35$). Both in the baseline as well as in the identity treatment, the retention rate is higher for higher $type_\omega$, for both in-group and out-group matches, and for low (*choice 1* < 13) as well as high values of *choice 1* (*choice 1* ≥ 13). This underscores the importance of voters' beliefs about representatives' competence (i.e., adverse selection concerns) in determining retention decisions. However, the drop in proportion retained from high $type_\omega$ to low $type_\omega$ is, on average, 14 to 16 points smaller in the identity treatment than in the baseline.[7]

Figure 1 B also shows that, contrary to the prediction in the no-effort equilibrium, effort matters for reelection in most cases—except for high $type_\omega$ representatives in the baseline. The difference in proportion retained for representatives with high and low effort choices runs in the identity treatment from about 13 percentage points for the in-group matches to about 25 percentage points for the out-group matches, and in the baseline from about 5 percentage points for high $type_\omega$ representatives to about 23 points for the low. Thus, similar to the baseline, *choice 1* systematically determines retention for low $type_\omega$ in the identity treatment, but unlike in the baseline, this strong *choice 1* dependency is also true in the identity treatment for high $type_\omega$. Moreover, while average differences in retention rates between low and high $type_\omega$

[7]Section D in the online supporting information provides evidence of the significance of this effect for the full range of $type_\omega$ values. To focus the analysis, in what follows, we often present results for the effects of *type*, $type_\omega$, and *choice 1* aggregating up to binary values, with the cut-points between low and high *type*/$type_\omega$ at 35 and between low and high *choice 1* at 13. Wherever we do so, we always run appropriate tests utilizing the full range of values of those variables and present only robust results.

FIGURE 1 **Effects of type$_\omega$ and *choice 1* on Retention Decision. Figure 1A: Differences in proportions of representatives retained between high ($\geq$ 35) and low ($<$ 35) type$_\omega$, given choice 1. Figure 1B: Differences in proportions of representatives retained between high ($\geq$ 13) and low ($<$ 13) values of *choice 1*, given type$_\omega$.**
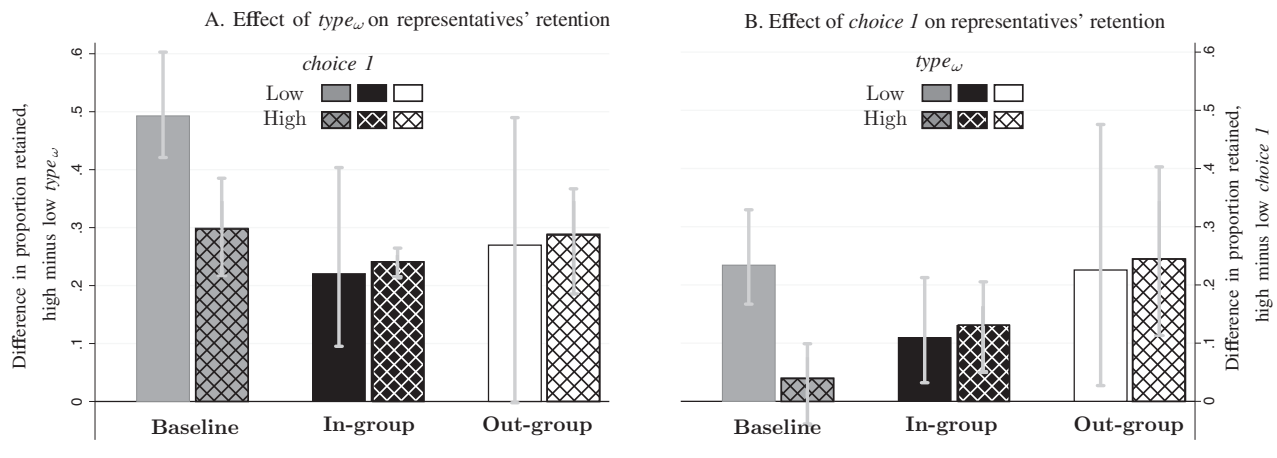
are essentially indistinguishable across in-group and out-group matches, *choice 1 is a systematically more important predictor of retention in out-group matches.*[8] We will see below that this result is driven by a subset of subjects for whom the induced in-group bias is greatest.

Investment into Effort Pre- and Post-Election

The first point to note about the pre-election (*choice 1*) effort levels is that they are systematically higher than minimal. The mean of *choice 1* in the baseline treatment is 13.21, compared to 12.27 in the in-group matches and 12.33 in the out-group matches of the identity treatment (the corresponding median values are 14, 12, and 12, respectively)—that is, about halfway up the given effort range. The expected value of *choice 1* in the baseline is systematically higher than in either in-group or out-group matches in the identity treatment; the absolute value

of the difference ranges up to 13% of the full *choice 1* range.[9]

The statistical significance of these differences notwithstanding, the absolute values of the summary statistics appear largely of a piece across treatments. As Figure 2 indicates, however, these apparent similarities conceal substantial differences between distributions, clearly contradicting BE1's prediction of a null effect of inducing identities.
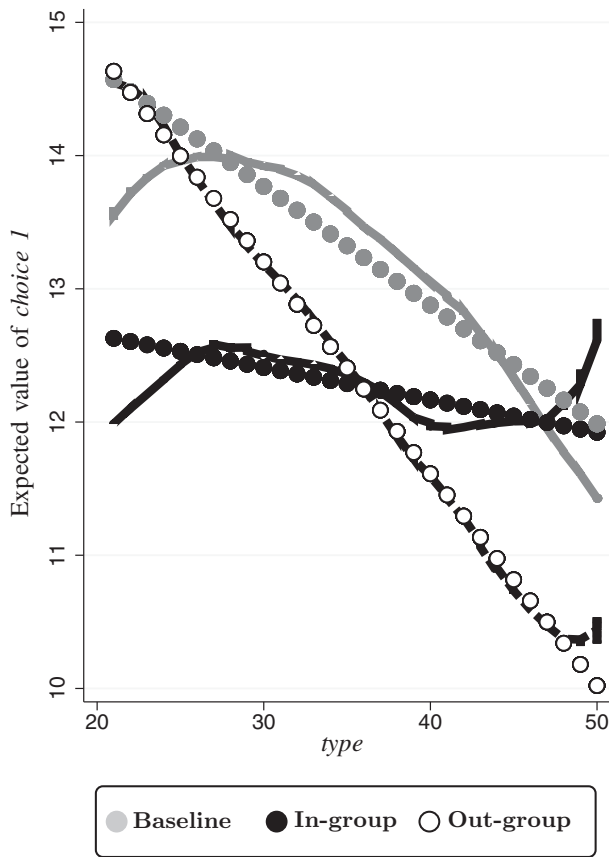
The figure provides a graphic display of the treatment effect with respect to representatives' choices. We find general type-independence of pre-election investment into effort in in-group matches, in contrast with weak type-dependence in the baseline treatment and strong type-dependence in out-group matches (see Figure 2).[10] The difference in expected values of *choice 1* in in-group and out-group matches ranges up to 14% of the *choice 1 range* (in fact, as we indicate below, for some categories of subjects, this difference is over twice that). Of special note is the observation, corroborating behavioral expectations *BE3.a′* and *BE3.b′*, that low-competence representatives in in-group matches invest less into effort than those in out-group matches, but high-competence ones invest more. Further, compared to the effort levels in the baseline, we see in the identity treatment lower effort from low-competence in-group and from high-competence out-group representatives. Antic-

---

[8]While the difference-in-means test comparing the effect of *choice 1* in in-group and out-group matches suggests that the differences fall short of significance (difference = .09 (−.04, .22), p = .25), difference-in-distribution tests give strong evidence for effort as a better predictor of retention in out-group matches than in in-group matches. The corresponding Fligner-Policello test, which relaxes both equal variance and approximately normal distribution, delivers nearly identical p-values. As a further robustness check, we simulated permutations of retention decisions and ran the test on each generated sample. The created distribution of test statistics, again, yields nearly identical results. In what follows, whenever we compare distributions, we run all three tests. Throughout, we report p-values from a two-tailed test.

[9]The Wilcoxon test confirms the significance of the decline (comparison of baseline versus in-group matches: $z = -2.48$, p = .01 baseline versus out-group matches: $z = -2.16$, p = .03).

[10]Further analysis of robustness in Appendix B confirms these findings.

Note: Estimates are based on a linear least squares regression of choice 1 on type and session round for the baseline, in-group match, and out–group match subject subsets. The nonlinear curves are the correspondent lowess curves fitted through the (type, choice 1) space.

ipating an individual-level analysis below, the explanatory intuition here is likely a version of a warm-glow effect: In-group representatives expect good will from in-group voters and so perceive no added need to raise the effort beyond a level reflecting their own warm glow, and high-competence out-group representatives likely feel no need to do anything beyond showing their competence for retention and may even experience a psychic disutility from incurring effort on behalf of the out-group voters.

Figure 2 shows also that behavior in none of the three treatment conditions is consistent with the prediction of the type-signaling equilibrium. In that equilibrium, representatives' investment into effort should be (weakly) increasing in *type* for an interior subset around the expected value of the *type* range. Figure 2, however, shows

that there is no such subset, thus undermining the behavioral expectation *BE2*.[11]

What about the representatives' post-election effort choices? Those choices are significantly lower than pre-election effort choices across treatments and identity matches and considerably closer to the universal equilibrium prediction of minimal effort: Mean values are 6.57 in the baseline and 8.39 and 7.96 in the in-group and out-group, respectively (over 50% of subjects chose 5 in all three cases), though the identity treatment comes with a statistically significant increase in post-election effort.[12]

Individual-Level Strategies of Play

In this section, we move from the analysis of the aggregate-level effects to the characterization of individual-level elements of subject behavior by defining a measure of group bias and using it to account for individual subject behavior. The measure also allows us to evaluate experimental evidence related to the rest of our behavioral expectations.

Measuring Group Biases in Retention Decisions

The basic notion of retention bias that underlies our measure is that of retaining representatives from one identity group in a wider range of circumstances than representatives from the other group. Since, for each voter, our sample contains choices for a relatively small subset of the range of possible circumstances—here, ($type_\omega$, *choice 1*) pairs—our first step in constructing our measure of subject bias must be to fill out the range of the ($type_\omega$, *choice 1*) pairs for which the subject may be expected to retain or not retain the representative, given the match between the voter's and the representative's identities and the voter's other observed election decisions. We do this by constructing, for each subject, two sets of monotone (incomplete) rules for each given group g she faces, $g \in \{in\text{-}group, out\text{-}group\}$. The first set, denoted by $\mathcal{R}_i^g$, contains all rules defined as sufficient conditions for retention such that the retention decisions these rules yield are consistent with observed retention choices made by a given subject i. The second set, denoted by $\mathcal{N}_i^g$, is similarly defined but now with respect to that subject's

[11] In fact, there is no such subset in any of the four possible match profiles: Klee-Klee pairings, Klee-Kandinsky pairings, Kandinsky-Klee pairings, or Kandinsky-Kandinsky pairings.

[12] As will be seen below, that increase is attributable entirely to a subset of the subjects displaying the in-group bias.

observed non-retention choices when facing representatives from that group.[13] Monotonicity in voting choices requires that whenever a voter chooses to retain (not to retain) a representative at a given pair of $type_\omega$ and *choice 1* values, she will also do so at any pair with weakly larger (smaller) values of both elements of the pair. Assuming that a voter's utility is always non-decreasing in $type_\omega$ and *choice 1*, asking whether a given subject's record of choices can be induced by a monotone decision rule is equivalent to asking whether those choices may be consistent with rational agency.

Suppose, then, that we are considering the set of monotone retention rules for in-group matches. Choose a rule $\bar{r}_i \in \mathcal{R}_i^g$ such that it is the most demanding rule in $\mathcal{R}_i^g$ with respect to the reelection requirements; that is, there exists no pair of $type_\omega$, *choice 1* values such that, for that pair, rule $\bar{r}_i$ induces reelection, but some other rule in $\mathcal{R}_i^g$ does not. Define i's *in-group retention set* to be the set of all ($type_\omega$, *choice 1*) pairs that rule $\bar{r}_i$ determines to be above the bar for reelection. Similarly, define i's *in-group non-retention set* to be the set of the ($type_\omega$, *choice 1*) pairs that rule $\underline{n}_i$ determines to be below the bar for reelection, where $\underline{n}_i \in \mathcal{N}_i^g$ is the least demanding rule in $\mathcal{N}_i^g$ with respect to the reelection requirements; that is, there exists no pair of $type_\omega$, *choice 1* values such that for that pair, some other rule in $\mathcal{N}_i^g$ induces reelection, but the rule $\underline{n}_i$ does not. We can similarly define subject i's retention and non-retention sets for out-group matches.

If there is exactly one and the same rule in both $\mathcal{R}_i^g$ and $\mathcal{N}_i^g$, then it must be that $\bar{r}_i^g = \underline{n}_i^g$. This can occur only if the voter's behavior is fully consistent with rational agency *and* the observational data are sufficiently complete and so fortuitously situated that the voter's behavior can be extended to cover the entire ($type_\omega$, *choice 1*)space. In practice, because of the limited nature of the data, there may be ($type_\omega$, *choice 1*) pairs that are in neither the (conservatively defined) retention set nor the non-retention set, and because of the vagaries of subject choices, there may also be pairs that are in both. We account for both of these possibilities by incorporating information about both the non-retention sets and the retention sets into the measure of bias.

Let R_i^g be the area of i's retention set for g-group representatives and let N_i^g be the area of i's non-retention set for g-group representatives. Then i's group bias is the difference-in-differences between the areas of reten-

tion and non-retention sets in in-group versus out-group matches:

$$D_i = (R_i^{in} - N_i^{in}) - (R_i^{out} - N_i^{out}),$$

Where R_i^{in} and N_i^{in} are the sizes of subject i's retained and non-retained sets for in-group matches, respectively, and R_i^{out} and N_i^{out} denote similar quantities for out-group matches. Whenever D_i is positive, we classify i as *in-group biased*, and whenever it is negative, as *out-group biased*. In a nutshell, an in-group-biased subject is willing to retain representatives who are, on average, assigned lower values of $type_\omega$ and who choose, on average, lower values of *choice 1* when those representatives are from his or her own identity group (i.e., are in-group matches) than when they are from the other identity group (i.e., are out-group matches).

Using our group bias measure, we behaviorally classify 37 out of the 70 subjects as *voters with in-group bias* and 33 subjects as *voters with out-group bias*. Seventy percent of voters in in-group matches and 73% in out-group matches do not intersect and so imply fully monotonic retention behavior. While there are non-empty intersections for the remaining subjects, these intersections span more than 10% of the ($type_\omega$, *choice 1*)space for only 11% of subjects in in-group matches and 7% of subjects in out-group matches, suggesting that our classification is relatively successful in making sense of subject behavior.[14] Behavioral prediction *BE4* anticipates the existence of subjects whose behavior can be characterized as in-group biased, and the above discussion provides evidence consistent with it.

What Drives the Behavioral Group Biases?

Having described the observed subject behavior using our measure of group bias, we next turn to the analysis of what motivates the biases it helps uncover. We proceed in two steps: first, considering more precisely how and where biased behavior manifests itself, and second, weighing

---

[13]These rules are incomplete insofar as they are silent on the ($type_\omega$, *choice 1*) pairs that do not meet the relevant sufficient condition.

[14]In a series of tests detailed in Section E.2 of the supporting information, we demonstrate the general robustness of our measure of group bias. In particular, we show that (1) the structure in retention choices it describes is, indeed, related to the treatment of experiencing an in-group or out-group match and is not an artifact of the particular ($type_\omega$, *choice 1*) pairs a voter encounters; and (2) that structure is not artificially imposed on behavior that is, in fact, random (randomizing retention choices produces systematically higher non-monotonicities than our data exhibit). Further, we show that (3) differences in retention rates between matches yield estimates consistent with the direction of subjects' group biases: Voters with in-group bias retain representatives in in-group matches more often than those in out-group matches, and bias-resisting voters retain representatives in out-group matches more often than those in in-group matches.

different pieces of evidence regarding the psychological, epistemic, or other possible causes of the observed behavioral patterns.

Manifestations of Behavioral Biases. By construction of our bias measure, in-group-biased voters retain in-group representatives more frequently than out-group representatives, with the reverse holding for out-group-biased voters. However, because that construction does not determine how group biases manifest themselves beyond that, detailed analysis of behavioral variation within and across the two bias groups can reveal critical elements of the nature of the entailed group favoritism.

Averaging across values of *choice 1*, in-group-biased voters retain representatives in in-group matches at higher rates than those in out-group matches: .59 for low $type_\omega$ and .79 for high $type_\omega$ in the in-group matches, compared to .43 for low $type_\omega$ and .59 for high $type_\omega$ in out-group matches. Further, compared to out-group-biased voters, in-group-biased voters are more lenient toward low $type_\omega$ in in-group matches but more demanding with respect to *choice 1* from high $type_\omega$ in out-group matches. Figure 3 demonstrates the clearly positive average treatment effect of .17 of being in the in-group match on retention rates by in-group-biased voters for low $type_\omega$ representatives and systematic negative average treatment effect of .36 of being in the out-group for high $type_\omega$ representatives who exert low effort.

In contrast, voters classified as out-group biased behave quite similarly to voters in the baseline, retaining low $type_\omega$ at lower rates than high $type_\omega$ (.33 for low $type_\omega$ in in-group matches and .49 in out-group matches compared to .62 for high $type_\omega$ in in-group matches and .82 in out-group matches) but accepting higher *choice 1* as a substitute for low $type_\omega$; the rate of retention in the top right panel of Figure 3 is significantly increasing in $type_\omega$ and in *choice 1* for low values of $type_\omega$, with the increase flattening out for high $type_\omega$. The key difference between the behavior of out-group-biased voters and the voters in the baseline is that for the latter, the retention rates in in-group matches are higher, even though the marginal effects of $type_\omega$ and *choice 1* are quite similar.

Motivational Rationales for Behavioral Biases. *In-group-biased voters.* The two starkest features of the behavioral profile of in-group-biased voters are their leniency toward both low type and low effort for in-group representatives and their increased demand for effort from high-type out-group representatives. What explains them?

Prior psychological research on social identity associates it with attachment-guided behavior, channeled as

an automatic preference for the in-group and relative apprehensiveness about the out-group.[15] The evidence from our experiment suggests that the in-group bias on the part of our subjects is more likely to be driven by the warm glow from benefiting members of their own group, which is consistent with this account (though see below), than by considerations of explicit reciprocity, either (1) individual level or (2) generalized. Speaking to (1), our data show a weak or no effect of $type_\omega$ and *choice 1* on retention choices by in-group-biased voters within in-group matches (see Figure 3). Regarding (2), we find no evidence that in-group bias classification is driven by the history of experiencing favorable retention choice as low $type_\omega$ in in-group matches, that is, that in-group bias is induced by generalized exchange among subjects (i's change from "not retained most of the time" to "retained most of the time" when having a low expected $type_\omega$ in in-group matches increases D_i by a mere .1%). The warm-glow account is also consistent with the effort choices by the same subjects in the role of a representative: Their in-group effort is larger than minimal and *type*-independent.

There are, however, important caveats that shed further light on the nature of the in-group bias we observe. First, in the exit survey, 28% of subjects we classified as voters with in-group bias, compared to only 4% of those classified as out-group biased, said that they were more likely to retain an in-group representative than an out-group representative independent of observed $type_\omega$ and *choice 1*. The self-awareness implicit in such an acknowledgment of one's own in-group bias may indicate that the responder's bias is, perhaps, less the unaware, purely subconscious attachment that some interpretations of warm glow describe and more a product of an explicit acceptance, albeit perhaps an unquestioned one, of a norm or habit to be nice to in-group members.[16]

The second caveat concerns another aspect of the behavior of in-group-biased subjects in our sample: retaining high types in out-group matches at a rate almost

---

[15] See Brewer (1979); Perdue, Dovidio, Gurtman and Tyler (1990); Greenwald and Banaji (1995); Brewer and Brown (1998); Ashburn-Nardo, Voils, and Monteith (2001) and DeSteno, Dasgupta, Bartlett and Cajdric (2004) for research in psychology and cognitive science confirming such findings.

[16] In response to another question, seeking to elicit the relevance of induced identities for the respondent's own decisions, 60% of in-group-biased voters claimed that identities did not matter, in contrast to 80% of the out-group-biased voters. Given considerable evidence of self-censoring in explicit expression of bias (Harber 1998; Nosek et al. 2007), it remains an open question whether these last survey responses implicitly demonstrate some subconscious bias or the usual response-bias-driven misreporting.

FIGURE 3 Proportions of representatives Retained by representatives' type$_\omega$ and *choice 1* and by Voters' Bias.
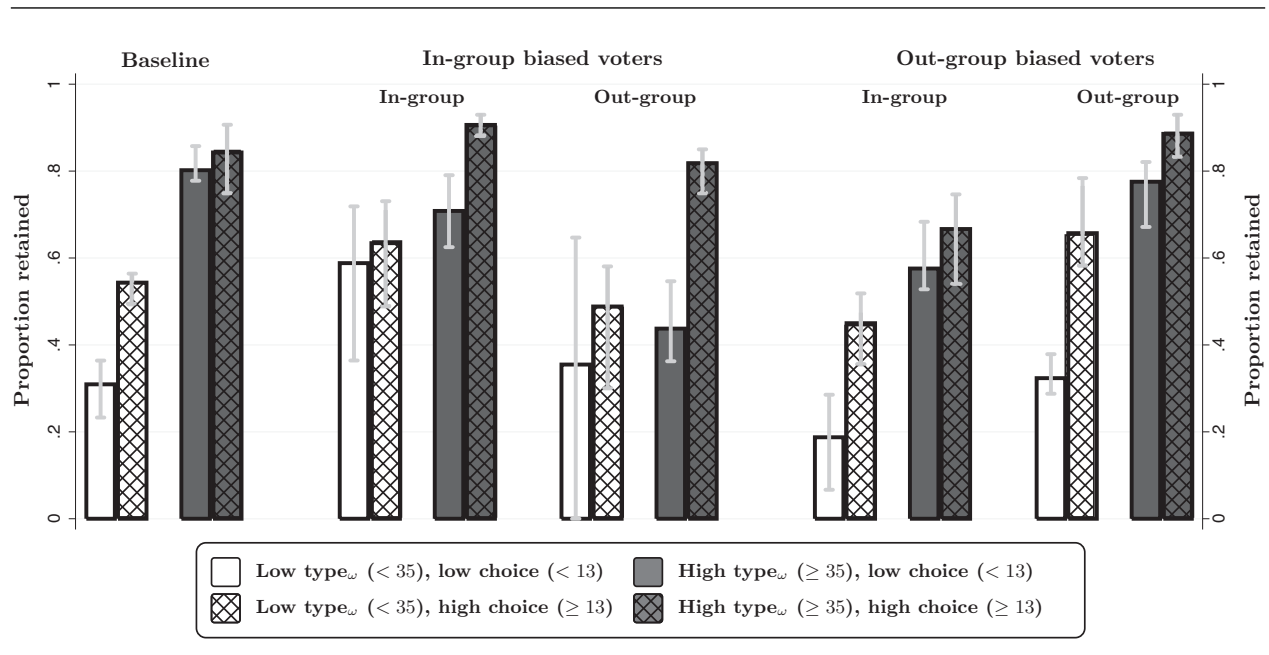
40% higher when they choose high rather than low levels of effort. This is about twice the size of the effect for the in-group matches and about 10 times that for the baseline. One possible explanation for this finding is the implicit presence, for the in-group-biased subjects, of a performance threshold that representatives must meet to be retained, with higher threshold values set for out-group than for in-group representatives (behavioral expectation *BE5*). Comparing low and high effort and competence quadrants in our sample, it is clear that out-group representatives are expected, on average, to show both high type$_\omega$ and high effort to have better than even odds of retention, whereas for in-group representatives, the expected probability of retention is already better than half for low type$_\omega$ and low effort, and the value added from increasing effort is much smaller. In-group-biased voters may be responding to high effort choices inducing outcomes that clear such performance thresholds as evidence of having other-regarding preferences—here, giving sufficient weight to their welfare; pushing the logic of warm glow further, it seems reasonable to suppose that the demand for such costly signals is higher in out-group matches both because voters themselves feel less of a fellowship with the out-group representatives and because they expect those representatives to feel similarly toward them.

An alternative explanation consistent with *BE6* may, however, be more powerful. A high investment into effort as a high type in out-group matches may signal that the representative is the kind of person who would exert

high effort even after the election—perhaps because the representative has an other-regarding preference or, more cynically, has an insufficient ability to perceive the benefits of post-election shirking. [17] In in-group matches, the appeal of such a type is also positive, but the mutual warm glow may account for the lower marginal effect of *choice 1* on retention. Indeed, in-group-biased voters, in contrast to out-group-biased voters, may be aware of and follow the behavioral pattern implied by this explanation. As representatives, they tend to spend more on effort before and after the election and, further, demonstrate a strong correlation between the two when matched with in-group voters ($\rho = .29$, $p = .08$) and especially when matched with out-group voters ($\rho = .49$, $p = .00$).[18]

Out-group-biased voters. Recall that like the voters in the baseline treatment, out-group-biased voters reward effort from low types and essentially ignore it for high types. The key behavioral distinction between them is that

---

[17]We thank an anonymous reviewer for pointing out this relationship to us. Note that because higher-than-minimal *choice 2* is never a best response, such signaling cannot be sustained as equilibrium in our strategic model, but insofar as there are subjects who behave that way, we can interpret that behavior as simply another unobserved dimension of their type that could be incorporated into a model with a richer specification of their preferences (Levine and Zheng 2010).

[18]These behavioral patterns hold true for low and high type$_\omega$, but high effort as a low type does not have the same informational content because the signal of being a high post-election chooser is confounded there with compensating for the low type.

out-group-biased voters retain in-group representatives systematically less frequently across all levels of type$_\omega$ and *choice 1*. In an exit survey, those voters were about 30% more likely than in-group-biased voters to choose "identities did not matter for my choices" in describing their behavior as voters (80% versus 60%) and about 50% more likely as representatives (40% versus 20%). We interpret the conjunction of these observations as suggesting that out-group bias may be brought about by a kind of resistance to identity-based favoritism, with the strongest resistance manifesting in what is, essentially, an "overcorrection" (in this sense, a more apt term for this group of subjects may be "bias-resisting" rather than "out-group-biased").

The resistance to the (implied) expectation that voters favor a candidate just because of their shared group membership may be prior to any aspect of our experimental treatment, reflecting education and upbringing that sensitize the subject to inequities. But it may also be reinforced by a history of play in which the relevant subject, in the role of a representative, is treated in a way that may be perceived as dissonant with in-group-biased behavior. What may be the most striking such instance is being retained in out-group matches as a representative with a low type$_\omega$, which should have the lowest expected probability of retention. Indeed, we find that having a history of such retention as a representative decreases the subject's in-group favoring bias as a voter by 25% (p = .08 in the difference-in-means test).[19]

The subject reasoning entailed in bias-resisting motivation for out-group bias suggests a considerable degree of self-awareness and sophistication, as indeed anticipated by the last behavioral expectation we articulated above (*BE7*). In fact, our data provide independent support for this expectation. First, if out-group-biased voters are more sophisticated, we should expect them to choose values of *choice 2* closer to the minimum (*choice 2* = 5) than voters with in-group bias. Indeed, mean values of *choice 2* are 9.67 for representatives who as voters show in-group bias but only 6.47 for those showing out-group bias[20]—a significant difference of 21% of the *choice 2* range that also pinpoints the culprit behind the differences between average *choice 2* values in the baseline and the identity treatments (6.57 versus 8.18). We find, further, that this difference in *choice 2* is entirely explainable by reelection as a low type$_\omega$ in out-group

matches (i.e., precisely by the aspect of the history of play that is a significant predictor of out-group bias). In contrast, neither the experiences within the in-group matches nor the marginal effects of histories on representatives who become in-group-biased voters are significant or robust.

Second, if out-group-biased voters are more cognitively sophisticated, we may expect them to be better at spotting the differences between Klee's and Kandinsky's styles in the pairs of paintings shown at the beginning of the identity treatment, and so then better able to identify the artist's hand in the quiz that followed it. Indeed, the difference-in-means shows that out-group-biased voters gave a somewhat larger proportion of correct answers than voters with in-group bias did (.22 versus .15, p = .12).

Taken as a whole, this evidence suggests that out-group bias does, indeed, go with greater cognitive sophistication, bolstering the plausibility of the resisting-the-group-favoritism motive behind out-group bias. Moreover, the contrast between more automatic in-group preference of in-group-biased voters and the more intentional calculus of the out-group-biased voters can be seen to parallel the distinction between the system 1 and system 2 cognitive processes made by cognitive psychologists (Stanovich 1999; Stanovich and West Forthcoming). That said, the overcorrection as such is unlikely to be intentional: Out-group-biased voters overwhelmingly reject the influence of group identities on their behavior as recorded in our exit survey and are themselves more likely to expect neutrality or a residual in-group favoritism from other voters.[21]

Evolution of Play

As we emphasized above, in our experimental design, identity is *not* reinforced by the assigned game payoffs. It is payoff relevant only insofar as subjects choose to behave in a way that makes it relevant. Because of this lack of direct payoff dependence and the relative weakness of the group identification we induce, it is reasonable to suppose that the subjects' initial behavior in the lab, as opposed to their behavior after reinforcement learning, is the best proxy for their preexisting perceptions and mental models. Consistent with this view, it would be unsurprising to observe identity attachment

[19]Other histories of reelection as either high or low type in in-group and out-group matches have no significant effect on one's identity bias.

[20]Those values are not significantly different when we control for high versus low type of the representative, and for in-group versus out-group matches.

[21]In further analysis in Section F of the supporting information, we consider and reject other potential motivations driving out-group bias, including the possibility that it could stem from what may be a version of in-group favoritism manifesting in higher expectations for fellow group members.

decreasing at the halfway point of the sessions, when subjects change their initially assigned roles and mutual expectations are updated and clarified: Once subjects realize that payoffs are dependent on identities only through mutual expectations, identities may lose their influence.

In fact, we find essentially no change in the distribution of bias values and in the ratios of subjects with in-group versus out-group bias, comparing subsamples before and after the midsession role switch. As we document in Section G of the supporting information, the basic patterns of subjects' retention decisions in both halves of the sessions also remain the same. The analysis of representatives' *choice 1* selections across the midsession role switch presents a somewhat more dynamic picture. Subjects who are, on the basis of their voting behavior, categorized as having in-group bias systematically chose, as representatives, values of *choice 1* type-dependently when matched with out-group voters but type-independently when matched with in-group voters, regardless of whether they played the role of representative in the first or in the second half of the experimental session. Subjects categorized as having out-group bias who were representatives in the first half of the session behaved similarly. In contrast to this continuity, out-group-biased voters who became representatives in the second half made *choice 1* decisions independent of type regardless of whether they were in an in-group or out-group match. Given the relative sophistication of these subjects and the above-noted history dependence of their play, the most likely explanation for this change is strategic learning, driven in part by what they take away from their experiences as voters in the first half of the session. It seems likely that, as voters, these subjects are learning about the correlation between *choice 1* and *choice 2*, and then as representatives selecting *choice 1*, seeking to pool with subjects for whom such correlation is, in fact, descriptive of the aspect of their underlying true type. With such subjects, then, higher levels of *choice 1* may be expected independent of their assigned type, whereas *choice 2* remains low. Both of these expectations are borne out in our data.

From the Experimental Evidence to Observational Studies of Accountability

Observational studies of the effects of social identity have focused on how shared social identity affects the elected officials' performance in office and on whether shared

social identity matters to the voters when they make election decisions. As noted in the introduction, their findings with respect to these two questions appear somewhat at odds: While the consensus of recent studies finds at best a very weak positive effect of shared group membership on favorable office performance,[22] studies focusing on voters' choices, in contrast, find that shared identity often matters.[23]

In allowing the sharp identification of the effects of effort and competence, our experimental design helps uncover the conditional effect of identity that suggests a way of reconciling these sets of results. Averaging across representatives' competence levels, we find, in line with the observational studies, no systematic effect of identities on representatives' effort; however, controlling for competence, we see fundamental identity-contingent differences. Voters who receive a signal that their representative is of high competence should be more enthusiastic about retaining him if he shares their group membership than if he does not because in the former case, representatives will be less likely to slacken their effort. At the same time, a signal of low competence should be more damning for representatives who share the social identity than for those who do not because the latter will do more to compensate for low competence with higher effort.[24]

The conjunction of the voters' identity biases and the conditional effect of shared membership on representatives' choices that we identify suggests how shared identity might matter to the voters. It offers a way of making sense of the seemingly disparate parts of the observational evidence while underscoring the value of a controlled experimental analysis for doing so. This point brings us to our final comment. In order to allow voters to condition their retention choices separately on effort or on their beliefs about the level of competence of their matched representative, effort and competence need to be at least partially observable. To understand this conditionality, we made effort observability a key element of our experimen-

---

[22]While some of the earlier studies have reported stronger effects, a recent wave of observational studies (Ferreira and Gyourko 2009; Gerber and Hopkins 2011; Hopkins and McCabe 2012 uses improved research designs and leads to far more skeptical conclusions.

[23]In mayoral elections in big urban centers like New York, Chicago, and Los Angeles, voters systematically show preference for candidates who share their racial background (Abrajano, Alvarez, and Nagler 2005; Kaufmann 2004).

[24]The examples of the incumbent mayors of New York and Los Angeles in the 1990s, Rudy Giuliani and Richard Riordan—by all counts, high-competence public officials who made few attempts to ingratiate themselves with out-group constituents (African Americans)— are, arguably, cases in point. See Kaufmann (2004), whose detailed account is consistent with our mechanism.

tal design. In the politics of actual electoral representation, effort is, indeed, sometimes observable (e.g., successful bill sponsorships, constituency service) and sometimes not (e.g., backroom deals, work to prevent the introduction or passage of bills that might hurt the constituency). The two primary motivations we identify behind representatives' choices in our experiment—the competence-independent warm glow for the in-group and the competence/effort substitutability for the out-group members—are not likely to decrease in their strength if effort is unobservable. Affective identification, which underlies the first, does not prima facie depend on observability. The absence of such identification, which is a key element behind the second, is, presumably, only reinforced when the choice is obscured. Thus, while unobservable effort may create incentives for other kinds of motivations (e.g., competence signaling) that are weak in the present experiment, one should expect the key motivations we discuss to remain.

Conclusion

The question of how shared social identity affects the nature of electoral accountability taps into long-standing debates in both academic and policy communities. The stakes in those debates appear high: They concern both the quality of democratic representation as well as racial/ethnic relations between representatives of different identity groups. Still, the present study is, to our knowledge, the first one to consider how shared or unshared social identity affects voters' responsiveness to distinct determinants of representatives' performance in a fully strategic environment.

Our key findings concern the behavior of both voters and representatives. On the representatives' side, we find, most strikingly, that whether the identity is shared affects whether representatives feel compelled to invest into effort by way of compensating for type deficit and, similarly, whether they feel comfortable slackening the effort when their type is high. This finding underscores one of our primary substantive conclusions: that the determination of whether to expect higher or lower effort from a representative with shared identity—the goal motivating much of the empirical discussion on the effects of candidates' race in U.S. elections—cannot be made independent of that representative's underlying competence. Assuming shared identity, representatives may, indeed, take the electoral support for granted, but they will, nonetheless, offer a steady higher-than-minimal level of effort. Representatives with unshared social identity will

work harder than fellow group members when their type is low, but, convinced of their reelection and lacking the warm glow, they may work less hard than fellow group members when their underlying competence is high.

On the voters' side, the presence of social identities raises the profile of representatives' effort in voters' retention decisions. This effect appears driven most strongly by the high demands on out-group representatives from the in-group-biased voters—the same voters who, in deciding whether to retain in-group representatives, are more apt to ignore both effort and competence. However, the evidence that more cognitively sophisticated voters tend to overcorrect for in-group favoritism and apply higher standards when judging the performance of representatives who belong to their social group further reinforces the view that a more complex picture of voter behavior than previous studies have allowed may be in order.

Appendix A

Summary Statistics

TABLE A1 Number of Subjects, Klees, and Kandinskys by Experimental Session

	Baseline			Identity				
Session	**1**	**2**	**3**	**1**	**2**	**3**	**4**	**Total**
Subjects	20	14	18	20	16	18	16	122
Klees				10	6	8	9	33
Kandinskys				10	10	10	7	37

TABLE A2 Rate of Being Matched with In-Group and Out-Group by Group Identity

	Matched with	
	Klee	**Kandinsky**
Klees	53%	47%
Kandinskys	45%	55%

Appendix B: Robustness of the choice 1 Model

The most comprehensive model of investment into effort (*choice 1*) would include current and past choices by the same player, her experienced

TABLE A3 Summary Statistics of Key Variables by Treatment

	Baseline (n=460, N=52)		Identity (n=618, N=70)		In-Group (n=308, N=70)		Out-Group (n=310, N=70)	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD
retention choices	.63	.48	.59	.49	.59	.49	.59	.49
type	36.25	8.72	35.50	8.57	35.57	8.32	35.44	8.83
type$_\omega$	35.62	11.92	34.90	12.17	34.50	11.72	35.31	12.61
choice 1	13.21	5.09	12.30	5.29	12.27	5.24	12.33	5.36
choice 2	6.57	3.59	8.18	5.20	8.39	5.26	7.96	5.14

retention decisions, as well as her history of being matched with in-group or out-group voters; it would allow for temporal effects, individual-level effects, and censoring. Table B1 demonstrates that the significant downward trend of the effect in *choice 1* over

type in out-group matches and in the baseline but not in-group matches we presented in Section "Investment into Effort Pre- and Post-Election" remains largely the same across estimators, error structures, and model specifications.

TABLE B1 Marginal Effects of Type on *choice 1* from a Regression of *choice 1* on Type and Treatment Dummies

Model	1	2	3	4	5	6	7	8
Baseline	−.089	−.089	−.137	−.137	−.102	−.100	−.149	−.086
	(.027)	(.044)	(.048)	(.047)	(.047)	(.029)	(.035)	(.039)
In-group match	−.023	−.023	−.038	−.038	−.013	−.015	−.033	−.075
	(.035)	(.044)	(.069)	.(.067)	(.037)	(.035)	(.052)	(.194)
Out-group match	−.159	−.159	−.240	−.242	−.152	−.152	−.226	−.129
	(.033)	(.041)	(.066)	(.065)	(.035)	(.033)	(.051)	(.066)
Estimator								
Least squares	x	x						
Censored mle			x	x				
Fixed effects					x			
Random effects						x		
Censored random effects							x	
First differenced								x
Allowed error structure								
iid	x	x	x	x	x	x	x	x
Heteroskedastic		x	x	x	x	x	x	x
Serially correlated		x	x	x		x	x	
Correlated with individual-level effect					x			x
Model specification								
Temporal effect				x	x	x	x	x
Individual effect					x	x	x	x
Choice history								x
Matching history								x
Retention history								x
Earnings history								x

(continued)

TABLE 5 (Continued)

Model	Estimator	Notes
1	Pooled OLS	Coefficient estimates are consistent when the pooled idiosyncratic errors are uncorrelated with the regressors (iid).
2	Pooled OLS with clustered errors	Errors may be correlated over time for a given subject so adjustment for clustering is appropriate.
3	Pooled tobit	*choice 1* may be censored from below at 5 and from above at 20.
4	Pooled tobit with round dummies	Particular rounds of play may have an effect on behavior.
5	Fixed effects	{ Individual-level effects may exist; a robust Hausman test suggests that both fixed and random effects are appropriate when regressors are strictly exogenous.
6	Random effects	
7	Tobit random effects	
8	First-differenced	Regressors may not be strictly exogenous so that a dynamic model may be necessary; including one lag of dependent variables and regressors as well as including experienced retention decision and accumulated payments lagged by one period eliminates autocorrelation. Consistent estimation is achieved by combining first–difference estimator and using further lags as instruments.

References

Abrajano, Marisa, Michael Alvarez, and Jonathan Nagler. 2005. "A Natural Experiment of Race-Based and Issue Voting: The 2001 City of Los Angeles Elections." *Political Research Quarterly* 58(2): 203–18.

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

Akerlof, George, and Rachel Kranton. 2010. *Identity Economics.* Princeton, NJ: Princeton University Press.

Ashburn-Nardo, Leslie, Corrine I. Voils, and Margo J. Monteith. 2001. "Implicit Associations as the Seeds of Intergroup Bias: How Easily Do They Take Root?" *Journal of Personal and Social Psychology* 81: 789–99.

Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg. 2010. "Creating Incentives and Selecting Good Types Revisited."

Banks, Jeffrey, and Rangarajan Sundaram. 1998. "Optional Retention in Agency Problems." *Journal of Economic Theory* 82(2): 293–323.

Beck, Paul, Russel Dalton, Steven Greene, and Robert Huckfeld. 2002. "The Social Calculus of Voting: Interpersonal, Media, and Organizational Influence on Presidential Choice." *American Political Science Review* 96(1): 57–73.

Bernhard, Helen, Ernst Fehr, and Urs Fischbacher. 2006. "Group Affiliation and Altruistic Norm Enforcement." *American Economic Review* 96(2): 217–21.

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90: 166–93.

Brewer, Marilynn. 1979. "In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis." *Psychological Bulletin* 86: 307–24.

Brewer, M., and R. Brown. 1998. "Intergroup Relations." In *The Handbook of Social Psychology* (Vol. 2), ed. D. Gilbert and S. Fiske. New York: McGraw-Hill, 554–94.

Campbell, Angus, Philip Converse, Warren Miller, and Daniel Stokes. 1960. *The American Voter.* New York: Wiley.

Chandra, Kanchan. 2004. *Why Ethnic Parties Succeed.* New York: Cambridge University Press.

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

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.

Converse, Philip. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David Apter. New York: Free Press, 206–61.

DeSteno, David, Nilanjana Dasgupta, Monica Bartlett, and Aida Cajdric. 2004. "Prejudice from Thin Air: The Effect of Emotion on Automatic Intergroup Attitudes." *Psychological Science* 15(5): 319–24.

Devine, P., A. Plant, D. Amodio, E. Harmon-Jones, and S. Vance. 2002. "The Regulation of Implicit Race Bias: The Role of Motivations to Respond without Prejudice." *Journal of Personality and Social Psychology* 82(5): 835–48.

Dickson, Eric, and Kenneth Scheve. 2006. "Social Identity, Political Speech, and Electoral Competition." *Journal of Theoretical Politics* 18(1): 5–39.

Eckel, Catherine, and Philip Grossman. 2005. "Managing Diversity by Creating Team Identity." *Journal of Economic Behavior & Organization* 58: 371–92.

Engelmann, Dirka, and Martin Strobel. 2004. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." *American Economic Review* 94(4): 857–69.

Fearon, James. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation.* Cambridge: Cambridge University Press, 55–97.

Fehr, Ernst, and Simon Gaechter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives* 14(3): 159–81.

Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114(3): 817–68.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50: 5–25.

Ferreira, Fernando, and Joseph Gyourko. 2009. "Do Political Parties Matter? Evidence from U.S. Cities." *Quarterly Journal of Economics* 124(1): 399–422.

Ferreira, Fernando, and Joseph Gyourko. 2011. "Does Gender Matter for Political Leadership? The Case of U.S. Mayors." NBER Working Paper 17671.

Fiorina, Morris. 1981. *Retrospective Voting in American National Elections.* New Haven, CT: Yale University Press.

Gerber, Elisabeth, and Daniel Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55(2): 326–39.

Gneezy, Ayelet, Alex Imas, Amber Brown, Leif Nelson, and Michael Norton. 2012. "Paying to Be Nice: Consistency and Costly Prosocial Behavior." *Management Science* 58(1): 179–87.

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): 212–16.

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, Alexander, Daniel Pallin, Kristal Raymond, Lisa Iezzoni, Dana Carney, Long Ngo, and Mahzarin Banaji. 2007. "Implicit Bias among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients." *Journal of General Internal Medicine* 22(9): 1231–38.

Greenwald, Anthony, and Mahzarin Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102(1): 4–27.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101(4): 709–25.

Harber, Kent. 1998. "Feedback to Minorities: Evidence of Positive Bias." *Journal of Personality and Social Psychology* 74(3): 622–28.

Hopkins, Daniel, and Katherine McCabe. 2012. "After It's Too Late: Estimating the Policy Impacts of Black Mayoralties in U.S. Cities." *American Politics Research* 40(4): 665–700.

Kahneman, Daniel, Jack Knetsch, and Richard Thaler. 1986. "Fairness as a Constraint on Profit Seeking." *American Economic Review* 76(4): 728–41.

Kaufmann, Karen. 2004. *Urban Voter: Group Conflict and Mayoral Voting Behavior in American Cities.* Ann Arbor: University of Michigan Press.

Landa, Dimitri. 2010. "Selection Incentives and Accountability Traps: A Laboratory Experiment." Paper prepared for presentation at the 2010 Annual Meeting of the American Political Science Association, Washington, DC.

Lazarsfeld, Paul, Bernard Berelson, and Hazel Gaudet. 1944. *The People's Choice.* New York: Columbia University Press.

Levine, David, and Jie Zheng. 2010. "The Relationship of Economic Theory to Experiments." In *The Methods of Modern Experimental Economics*, ed. Guillame Frechette and Andrew Schotter. Oxford: Oxford University Press, forthcoming.

Lipset, Seymour, and Stein Rokkan. 1967. *Party Systems and Voter Alignment: Cross-National Perspectives.* New York: Free Press.

Lupia, Arthur, and Mathew McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

Lupia, Arthur, and Mathew McCubbins. 2000. "The Institutional Foundations of Political Competence: How Citizens Learn What They Need to Know. cäIn *Elements of Reason*, ed. Arthur Lupia, Mathew McCubbins, and Samuel Popkin. New York: Cambridge University Press, 47–66.

McLeish, Kendra, and Robert Oxoby. 2007. "Identity, Cooperation, and Punishment." IZA Discussion Paper No. 2572.

Mendelberg, Tali. 2001. *The Race Card.* Princeton, NJ: Princeton University Press.

Mendes, Wendy, and Katrina Koslov. Forthcoming. "Brittle Smiles: Positive Biases towards Stigmatized and Outgroup Targets." *Journal of Experimental Psychology.*

Metz, David, and Katherine Tate. 1995. "The Color of Urban Campaigns." In *Classifying by Race*, ed. Paul E. Peterson. Princeton, NJ: Princeton University Press, pp. 262–277.

Morton, Rebecca, Kenneth Williams, and Anna Bassi. 2011. "The Effect of Identities, Incentives, and Information on Voting." *Journal of Politics.*

Nosek, Brian, Frederick Smyth, Jeffrey Hansen, Thierry Devos, Nicole Lindner, Kate Ranganath, Colin Smith, Kristina Olson, Dolly Chugh, Anthony Greenwald, et al. 2007. "Pervasiveness and Correlates of Implicit Attitudes and Stereotypes." *European Review of Social Psychology* 18(1): 36–88.

Pearson, Adam, John Dovidio, Curtis Phills, and Ivuoma Onyeador. Forthcoming. "Attitude-Goal Correspondence and Interracial Interaction: Implications for Executive Function and Impression Formation." *Journal of Experimental Social Psychology.*

Peffley, Mark, Jon Hurwitz, and Paul Sniderman. 1997. "Racial Stereotypes and Whites' Political Views of Blacks in the Context of Welfare and Crime." *American Journal of Political Science* 41(1): 30–60.

Perdue, Charles, John Dovidio, Michael Gurtman, and Richard Tyler. 1990. "Us and Them: Social Categorization and the Process of Intergroup Bias." *Journal of Personality and Social Psychology* 59(3): 475–86.

Plant, E., and P. Devine. 1998. "Internal and External Sources of Motivation to Respond without Prejudice." *Journal of Personality and Social Psychology* 75(3): 811–32.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83(5): 1281–1302.

Stanovich, Keith E. 1999. *Who Is Rational? Studies of Individual Differences in Reasoning.* New York: Psychology Press.

Stanovich, Keith, and Richard West. 2000. "Individual Differences in Reasoning: Implications for the Rationality Debate." *Behavioral and Brain Sciences* 23(5): 645–726.

Swain, Carol. 1993. *Black Faces, Black Interests: The Representation of African Americans in Congress.* Cambridge, MA: Harvard University Press.

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

Tajfel, Henri, Michael Billig, R. Bundy, and Claude Flament. 1971. "Social Categorization and Inter-Group Behavior." *European Journal of Social Psychology* 1: 149–77.

Tajfel, Henri, and John Turner. 1986. "The Social Identity Theory of Intergroup Behavior." In *The Psychology of Intergroup Relations*, ed. Stephen Worchel and William Austin. Chicago: Nelson- Hall, 7–24.

Woon, Jonathan. 2012. "Democratic Accountability and Retrospective Voting: A Laboratory Experiment." *American Journal of Political Science* 56(4): 913–30.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

C Model of in-group biased behavior
D Robustness of average treatment effects
- ROC-analysis of retention decisions of type ω.
E Examples and robustness of the group bias measure
- Examples of four subjects illustrating the construction of the group bias measure
- Evidence of robustness of the bias measure to idiosyncrasies of particular observations, non-randomness of observed choices, and consistency of the bias measure with observed retention choices.
- ROC-analysis of retention choices as a function of type ω and choice 1.
- Estimation of predicted probabilities of retention by bias as a function of type ω and choice 1.
F Robustness of findings: motivational rationales for behavioral bias
- Further evidence ruling out alternative explanations for the observed biases
G Robustness of bias measure to role-switch
- Further evidence from voters' retention decisions that behavior before and after the role–switch remains the same
H Exit-survey
- Subjects self-reports on the strategies they employed
I Experimental design
- Basic information on laboratory and software–Instructions