Respect-Based Reciprocity

Timm Gries*

March 26, 2026

Abstract

This paper proposes respect-based reciprocity: people are more willing to support an actor when they believe that actor holds them in higher regard. I define perceived respect as a second-order belief about types and argue that its behavioral force is strongest when outside respect is scarce. I test these predictions in two settings. First, in a controlled experiment designed to isolate perceived respect from standard alternatives, I find that it substantially increases support. Second, using nationally representative survey data from the 2016 U.S. election, I show that perceived candidate respect strongly predicts vote choice conditional on rich demographics, partisanship, ideology, and policy views, and that this relationship is stronger among voters predicted to receive less outside respect. These findings provide a microfoundation for the importance of recognition gaps and a parsimonious link between social standing and political realignment. The paper then develops implications for communication and signaling: it shows how recognition is often cheap talk, how negative out-group messaging can restore its credibility, and how messages, policies, and rules shape behavior through the respect they convey.

*Bocconi University. Email: timm.gries@phd.unibocconi.it. I am grateful to Luca Braghieri, Leonardo Bursztyn, Stefano Fiorin, Rafael Jiménez-Durán, Salvatore Nunnari, and Guido Tabellini for helpful comments and discussions. I also thank seminar participants at Bocconi and the University of Chicago, as well as participants at the Behavioural Spring School 2026, organized by UC San Diego and FAIR, for valuable feedback. All errors are my own. This research was supported by Bocconi University. The author has no financial relationships or other potential conflicts of interest to disclose.The experiment was pre-registered as AsPredicted 281,027.

1 Introduction

In September 2016, Hillary Clinton described half of Donald Trump's supporters as belonging to a "basket of deplorables" -"the racist, sexist, homophobic, xenophobic, Islamophobic" (Clinton 2016). Trump immediately turned the remark into a political contrast. "My opponent slanders you as deplorable and irredeemable," he told supporters; "I call you hardworking American patriots who love your country" (Trump 2016). What made the exchange politically potent was not only disagreement over policy. It was a contest over regard: over whether voters felt respected or looked down upon. This paper studies that channel.[1]

I propose *respect-based reciprocity* (RBR): people are more willing to support an actor when they believe that actor holds them in higher regard. I define *respect* as an audience's (B's) belief about a person's (A's) type - about whether she is hardworking, competent, decent, or otherwise worthy of esteem. *Perceived respect* is the corresponding second-order belief: what A believes B believes about A's type. The central claim of the paper is that this perceived respect directly affects A's intrinsic valuation of acting in B's favor. The mechanism yields three predictions. First, support should increase in perceived respect. Second, the same increase in perceived respect should matter more when *outside respect* - the regard one perceives from others beyond the focal actor - is scarce. Put differently, respect should have diminishing sensitivity: when it is abundant, an additional increment matters less; when it is scarce, it matters more. Third, support depends not only on the level of perceived respect but also on respect relative to an appropriate benchmark.

This perspective differs from the channels that economists have usually emphasized. A large literature studies *esteem-seeking*: people work, give, or conform in order to influence how others see them (Bernheim 1994; Bénabou and Tirole 2006; Bursztyn and Jensen 2017). Philosophy, political science, and sociology, meanwhile, have long stressed the importance

1. The "deplorables" remark coincided with a visible shift in tracking polls toward Trump (Appendix Figure A1). See also Lamont, Park, and Ayala-Hurtado (2017) for a broader discussion of recognition deficits and the 2016 election. Similar episodes recur across campaigns. In 2008, Barack Obama described some frustrated rural communities as people who "get bitter, they cling to guns or religion"; he later called the remark "my biggest mistake of the campaign" (Obama 2020, p. 143).

of recognition, status, and dignity in political life (Hegel 1830; Fraser 2000; Honneth 2001; Fukuyama 2018; Lamont 2018). RBR complements both: whereas esteem-seeking asks how people act to shape others' beliefs, RBR asks how people respond once they perceive those beliefs to be favorable or unfavorable. Likewise, while the broader social sciences have documented the importance of recognition and dignity, RBR offers a formal channel through which they affect behavior. Since the mechanism runs through second-order beliefs about types, it can operate even when the other side has taken no payoff-relevant action and even when the supportive act itself is fully unobserved. That makes it especially relevant for settings such as voting, anonymous transfers, discretionary effort, and compliance with rules.

The political-economy relevance is immediate. A prominent interpretation of contemporary politics is that large groups of voters feel ignored, disdained, or stripped of social standing (Hochschild 2018; Lamont 2018; Hochschild 2024). But while those ideas are compelling, they are often difficult to map into a sharp behavioral mechanism. RBR provides one. If structural change and social reordering reduce perceived outside respect for some citizens, then those citizens should become especially responsive to candidates who signal regard for "people like them." In that sense, the paper offers a microfoundation for recognition gaps and a parsimonious link between social standing and political realignment (Gidron and Hall 2017; Gethin, Martínez-Toledano, and Piketty 2022; Zollinger and Haeusermann 2024). It also helps explain why symbolic political communication - praise, condemnation, appeals to dignity, and group-directed language more generally - can have electoral consequences even when policy positions change little.

The same logic extends beyond electoral politics. In many environments, actors respond not only to incentives and information, but also to what institutions, rules, and decision makers appear to believe about them.[2] A manager who trusts an employee, a principal

2. Information about B's belief about A's type is of course also information. The distinction from more classical notions of information is that A does not need to learn about any "material" feature of the environment, such as her actual type or the nature of a given task.

who grants discretion, or an institution that relies on fixed pay rather than intrusive monitoring may convey favorable beliefs about the agent's type; suspicion, surveillance, and excessive control may convey the opposite. If people reciprocate perceived respect, then formal and informal institutions interact through a channel that standard incentive models do not fully capture. This broader portability is part of the paper's contribution: respect-based reciprocity is not a mechanism specific to campaigns, but a general social preference with implications for political economy, organizations, and institutional design.

The paper asks three questions. Do people intrinsically reciprocate respect? When is that mechanism strongest? And what follows for political communication and institutional signaling if they do? I address the first question with a controlled experiment designed to isolate perceived respect from leading alternatives. Participants are paired with an anonymous counterpart; the counterpart takes no action toward them; and the participant's support decision is fully unobserved. The treatment varies only what the participant believes the counterpart believes about her type. By construction, the design shuts down intention-based reciprocity, social image, and inference about the counterpart's type, allowing the paper to cleanly test whether perceived respect itself changes support. The experiment finds a large and significant effect: participants in the high-respect condition allocated 35 percent more to their counterpart than those in the low-respect condition ($p < 0.001$).

I then study whether the same logic helps explain political behavior outside the experimental setting. Using nationally representative survey data from the 2016 U.S. election, I find that perceived candidate respect predicts vote choice at a magnitude significantly above almost all individual policy preferences, even after conditioning on a rich set of demographics, partisanship, and ideology. The relationship is also stronger among voters predicted to receive less outside respect, as proxied by occupational prestige - consistent with the diminishing-sensitivity prediction but not easily explained by alternative accounts such as instrumental learning about representation. The value of combining these two pieces of evidence is that they speak to different margins: the experiment isolates the mechanism, while

the survey evidence shows that it organizes politically consequential behavior in the field.

Finally, the paper develops two broader implications. First, in political and other types of communication, RBR implies that recognition is often cheap talk: any candidate has incentives to claim to respect a group. But negative out-group messaging can make recognition credible by showing that praise is selective rather than indiscriminate. Second, in institutional or incentive design, the framework extends the signaling value of messages, policies, and rules. They do not shape behavior only through material incentives or information about task characteristics (Bénabou and Tirole 2003); they also shape behavior through the respect they express. The paper's contribution is therefore not only to introduce a new social preference. It is also to identify it behaviorally, connect it to a central political-economy pattern, and discuss how it changes the way one thinks about communication and institutions.

The rest of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 develops the framework and formalizes the main predictions. Section 4 presents the experiment. Section 5 presents the voting evidence. Section 6 develops the implications for political communication and institutional signaling. Section 7 concludes.

2 Literature

Social preferences and reciprocity. This paper contributes to a large literature that has modeled social preferences as conditional on features of the interaction. Outcome-based models condition on others' payoffs or payoff differences (Fehr and Schmidt 1999; Charness and Rabin 2002); intention-based reciprocity conditions on others' actions and what those actions reveal about kindness or fairness intentions (Rabin 1993; Dufwenberg and Kirchsteiger 2004; Falk and Fischbacher 2006); type-based conditional altruism conditions on others' types (Levine 1998); guilt and disappointment aversion condition on others' expectations about payoffs (Charness and Dufwenberg 2006; Battigalli and Dufwenberg 2007). RBR adds a contingency that is absent from all of these: B's belief about A's type. The relevant

agent is oneself, not the counterpart; and the relevant object is a belief, not a payoff or an action. This makes RBR applicable in settings where standard social preferences are not well defined - large anonymous elections, remote labor relationships, interactions without observable action histories - because it requires neither an action by B nor any observability of A's choice. RBR is naturally cast in the language of psychological game theory (Geanakoplos, Pearce, and Stacchetti 1989; Battigalli and Dufwenberg 2022), in which payoffs can depend on beliefs. It complements earlier work in that tradition by extending the payoff-relevant belief to a second-order belief about types rather than about payoffs or strategies.

Esteem and social image. A significant literature models the intrinsic value of esteem or reputation: individuals care about being held in high regard and adjust behavior to affect others' impressions (Bernheim 1994; Bénabou and Tirole 2006, 2009; Bursztyn and Jensen 2017). In these models, the *target* of the preference is A's reputation in the eyes of B: A acts in order to be seen positively. The direction is reversed in RBR: A acts in B's favor because B already holds a positive belief about A. The closest papers in this tradition are Ellingsen and Johannesson (2007, 2008), who show that the intrinsic value of esteem depends on the audience's type: respect from an admired individual matters more. RBR differs on two dimensions. First, the conditioning variable is different: in Ellingsen and Johannesson, B's *type* affects how much A cares about B's opinion; in RBR, B's *belief about A's type* affects A's utility from supporting B. Second, and more practically, social image concerns, as in Ellingsen and Johannesson, require that A's action be at least partially observable to B. RBR operates even under complete anonymity - a prediction that the experimental design exploits directly, and that makes RBR especially relevant in settings (large elections, anonymous online interactions, unmonitored labor effort) where social image arguments have the least bite (Bursztyn and Jensen 2017).

Recognition in the workplace and the identification gap. Consistent with RBR, a body of field and laboratory evidence documents that expressions of recognition and respect

raise effort and prosocial behavior even when they carry no material content. For example, Kosfeld and Neckermann (2011) show that a certificate of award increases performance; Bradler et al. (2016) document that a short recognition statement increases productivity in a field experiment; Friebel et al. (2025) provide quasi-experimental evidence that managerial respect raises worker effort. These findings are explicable through RBR: a recognition signal conveys that the employer holds positive beliefs about the worker's type, and RBR predicts the resulting behavioral response. However, none of these studies isolate RBR. Awards could trigger career concerns, signal costly investment by the employer, activate social identity, or generate guilt about anticipated expectations. Again, Ellingsen and Johannesson (2008) rationalize some of these findings through esteem concerns, which require observability. The experiment in this paper fills this gap: by holding B's actions, B's type, group cues, and observability all constant while varying only B's stated belief about A's type, I obtain the first clean behavioral identification of RBR.

Intrinsic motivation and policy as a signal. A foundational contribution of Bénabou and Tirole (2003) is the analysis of how extrinsic incentives can crowd out intrinsic motivation because they convey information: offering a reward for behavior that was previously unprompted signals that the principal expects the agent not to engage spontaneously, thereby implicitly attributing a low type. This logic has been extended to laws and norms as signals of social standards (Bénabou and Tirole 2006; Bowles and Polania-Reyes 2012; Ashraf and Bandiera 2018; Benabou and Tirole 2026). RBR provides an additional micro-channel for this signaling effect: the reason crowding-out occurs is that the policy lowers the agent's perceived respect, which reduces the intrinsic value of supporting the principal. Crucially, this channel operates even when the principal has no informational advantage - it suffices that the policy is associated with a belief about the agent's type that the agent perceives. I formalize this in Section 6: fixed wages, broad discretion, and trust signal high beliefs; monitoring, fines, and piece-rate pay signal low beliefs; and the resulting RBR response alters

effort independently of whether the principal is better informed than the agent about the agent's type or the nature of the task.

Political economy of recognition and realignment. A long tradition in political sociology and philosophy documents that demands for recognition - the affirmation of one's dignity and social worth - are a central force in political life (Taylor 1995; Honneth 2001; Fraser 2000; Fukuyama 2018). Empirically, sociologists have linked structural economic changes to "recognition gaps": declines in occupational standing and social visibility that leave large communities feeling dismissed by political and cultural elites (Hochschild 2018; Lamont 2018). These recognition gaps have been proposed as drivers of political realignment toward populist movements (Lamont, Park, and Ayala-Hurtado 2017; Zollinger and Haeusermann 2024). RBR provides the micro-foundation that connects these macro-level observations to individual behavior: perceived disrespect directly reduces the intrinsic value of supporting a political candidate, and H2 in Section 3 implies that the effect is amplified precisely for those with the lowest outside respect. The heterogeneity I document in Section 5 - stronger sensitivity to perceived candidate respect among low-prestige voters - translates this mechanism into a testable and falsifiable prediction. Economic models of identity politics (Shayo 2020; Besley and Persson 2019; Gennaioli and Tabellini 2023) analyze how candidates target group identities to mobilize voters; RBR complements this by identifying an intrinsic channel that increases the importance of social identities. If voters care about being respected, but candidates can not communicate to individual voters, respect may instead be conveyed for social identities. In this sense, social identities function as vehicles for expressing and perceiving respect, and their political salience is in part an expression of voters' sensitivity to the beliefs of those who speak in identity terms.

3 Framework

We propose a social preference called *respect-based reciprocity* (RBR) and derive its behavioral implications. The central idea is that individuals are intrinsically more inclined to support an audience when that audience holds them in higher regard. We formalize this intuition, motivate three hypotheses, and discuss their scope and relation to existing social preferences.

3.1 Setup

Agents and actions. Consider two agents, A and B. Agent A chooses a *support action* $s \in [0, \bar{s}]$: an action A perceives as benefiting B. The support action may be material - allocating money to B, or exerting effort on a task that benefits B - or expressive. In large elections where A is unlikely to be pivotal, voting for B carries little material weight but substantial expressive value: the act of supporting B provides intrinsic utility independent of whether it affects the outcome (Brennan and Buchanan 1984).[3]

Type. Agent A has a type $\theta_A \in [\underline{\theta}, \bar{\theta}]$: a valued characteristic such as ability, honesty, diligence, or prosociality. What constitutes a valued type is context-specific but assumed consensual within a given social setting. We do not require θ_A to be directly observable.

Respect. We define *respect* as an audience's belief about an individual's type. B holds a belief $\mu_B \in \Delta([\underline{\theta}, \bar{\theta}])$ about A's type, summarized by $r_B \equiv \mathbb{E}_{\mu_B}[\theta_A]$. The higher B's expected assessment of A's type, the more B *respects* A. Respect is a belief, not an action: B need not do anything for a belief about A's type to be defined.

Beyond respect from B, two further objects matter. *Outside respect* R summarizes how favorably agents other than B regard A - the aggregate of their beliefs about A's type. The *reference belief* $\bar{r}_B$ is what a neutral observer would believe about A given the information available to B: the default assessment of A's type.

3. In general, s may combine material and expressive components. For the primary applications we have in mind - voting and effort under gift exchange - either the material channel is negligible (large elections) or orthogonal to the respect component.

Perceived respect. A does not directly observe r_B, R, or $\bar{r}_B$, but forms beliefs about them. Let ρ_B denote A's *perceived respect* from B (her second-order belief about r_B), η her *perceived outside respect* (her belief about R), and $\tilde{r}_B$ her *perceived reference belief* (her belief about $\bar{r}_B$).

These second-order beliefs are the relevant state variables: B's actions and messages matter for A's behavior precisely to the extent that they shift ρ_B, η, or $\tilde{r}_B$. This perspective applies broadly. A may perceive respect from B through literal messages, policies, unintended behavior, or signals of any kind. The second-order belief captures the net effect of all these inputs on A's assessment of B's view of A, and it is this assessment - rather than any particular action of B - that enters A's utility.[4]

3.2 Hypotheses

A's utility from taking support action s is

$$U_A(s) \;=\; u(s) \;+\; \varphi(\rho_B,\, \eta,\, \tilde{r}_B) \cdot v(s), \tag{1}$$

where $u(s)$ captures the net utility of the action, the component that is independent of perceived respect; $v(s) > 0$ is its intrinsic value as a support action; and $\varphi(\cdot)$ is a *respect sensitivity function* that scales how much A values supporting B, based on perceived respect. The central departure from standard utility is that the intrinsic weight placed on the support action depends on perceived respect. We propose three properties of φ, each yielding a behavioral hypothesis.

Hypothesis H1 (Respect-Based Reciprocity)**.** A's support for B increases in perceived respect from B: $\partial\varphi/\partial\rho_B > 0$.

4. We treat ρ_B, η, and $\tilde{r}_B$ as point estimates, abstracting from uncertainty over second-order beliefs for tractability. Such uncertainty introduces a further channel: messages and policies can affect behavior by changing A's *confidence* in her perceived respect, not only its level. We return to this briefly in Section 6.

Motivation. Neuroscience evidence establishes that positive social evaluation activates reward circuits with a potency comparable to material gains: "social [...] consequences of performance were [...] of comparable salience to monetary reward" (Zink et al. 2008). If a positive belief about A thus constitutes a gift - valued much like a material transfer - standard gift-exchange logic predicts that A will reciprocate by supporting the giver (Akerlof 1982; Fehr and Gächter 2000). A second, expressive channel reinforces H1: people act partly to confirm and express a positive self-image (Bénabou and Tirole 2003), and supporting B who holds A in high regard is consistent with that image, while supporting someone who disrespects A is not.[5] Consistent with H1, field and laboratory studies find that expressions of recognition raise prosocial behavior and effort, even when recognition carries no material content (Kosfeld and Neckermann 2011; Bradler et al. 2016; Friebel et al. 2025). As we discuss in Section 3.4, none of these studies isolate RBR; our experiment is designed to do so.

Hypothesis H2 (Diminishing Sensitivity)**.** A's sensitivity to perceived respect from B decreases in perceived outside respect: $\partial^2\varphi/(\partial\rho_B\,\partial\eta) < 0$.

Motivation. The marginal value of normal goods diminishes once it is available in abundance; H2 applies this logic to respect. The intuition is captured by the following remark in McMillan (2002): "Respect is like air. As long as it's present, nobody thinks about it. But if you take it away, it's all that people can think about." Converging evidence supports this pattern across disciplines. Social psychology finds that individuals respond to social exclusion with heightened motivation to form new social bonds, actively seeking esteem from alternative sources (Maner et al. 2007). Neuroscience provides a striking parallel: "Midbrain regions showed selective activation to food cues after fasting and to social cues after isolation" (Tomova et al. 2020) - deprivation of social respect creates a drive comparable in neural terms to hunger, heightening the salience and impact of any positive social signal. In economics,

5. H1 does not require that B's beliefs about A be objectively informative. The intrinsic utility effect follows from A's perception of B's belief, not from its accuracy.

Bursztyn et al. (2018) document that social image concerns are particularly pronounced for individuals with a self-image that is reduced through priming, consistent with esteem exhibiting diminishing marginal utility. H2 is important for identification:Individuals who perceive little outside respect should be most responsive to any given source of recognition, allowing us to test RBR in the field via heterogeneity by social standing.

Hypothesis H3 (Reference Beliefs)**.** A's support for B decreases in the perceived reference belief, holding perceived respect fixed: $\partial\varphi/\partial\tilde{r}_B < 0$.

Motivation. What matters for A is not B's respect in absolute terms but relative to what seems appropriate. We define the perceived *respect gap* as $\rho_B - \tilde{r}_B$: the degree to which B's belief about A exceeds the reference standard. When the gap is positive, B is giving A the benefit of the doubt - a flattering departure from the baseline, akin to a gift. When the gap is zero, B holds the default view; there is nothing to reciprocate. When the gap is negative, B departs unfavorably from what is appropriate, a slight that A may penalize. H3 formalizes this: a sufficient condition is that φ depends on excess respect $\rho_B - \tilde{r}_B$ rather than on ρ_B alone, paralleling the role of fairness benchmarks in the reciprocity literature (Rabin 1993; Falk and Fischbacher 2006).We regard H3 as an important complementary prediction. Our experiment is designed primarily for H1 and H2.

3.3 Behavioral Predictions

Under standard regularity conditions (strict concavity of U_A in s), A chooses s^* satisfying the first-order condition $u'(s^*) + \varphi\, v'(s^*) = 0$. The three hypotheses yield immediate behavioral predictions.[6]

Prediction 1 (RBR)**.** A's optimal support for B, s^*, is strictly increasing in perceived respect ρ_B.

6. Differentiating the first-order condition: $\partial s^*/\partial\rho_B = -\varphi_{\rho_B}\, v'(s^*)/[u''(s^*) + \varphi\, v''(s^*)] > 0$ by H1 ($\varphi_{\rho_B} > 0$), $v' > 0$, and strict concavity of the denominator (negative). Prediction 2 follows from H2: $\partial^2 s^*/(\partial\rho_B\,\partial\eta) < 0$. Prediction 3 follows from H3.

Prediction 2 (Diminishing Sensitivity)**.** The marginal effect of perceived respect on support, $\partial s^*/\partial \rho_B$, is strictly decreasing in perceived outside respect η.

Prediction 3 (Reference Beliefs)**.** A's optimal support s^* is strictly decreasing in the perceived reference belief $\tilde{r}_B$, holding ρ_B fixed.

Prediction 1 is the primary target of the experiment in Section 4 and Predictions 1 and 2 are central to the voting evidence in Section 5. Prediction 2 is particularly important: no standard model of reciprocity, altruism, or social image concerns predicts that sensitivity to a given source of positive regard decreases in the outside level of regard received and therefore helps distinguish RBR from alternatives.

3.4 Scope and Distinctiveness

Two features of RBR are worth highlighting before comparing it to existing work. First, *no action by B toward A is required*: respect is a belief, and A's support for B depends on A's second-order beliefs about that belief, not on any observable act of B. Second, *observability of A's action is not required*: A's support for B derives from intrinsic utility, so RBR has bite even when A's action is unobserved by B and has no effect on B's beliefs or material payoffs. We exploit both properties in the experimental design, which implements full anonymity between A and B and ensures A's allocation decision is unobserved. This simultaneously rules out social image concerns and kindness reciprocity as explanations for any observed treatment effect.

Kindness and intention-based reciprocity (Rabin 1993; Dufwenberg and Kirchsteiger 2004; Falk and Fischbacher 2006) make A's behavior respond to B's actions: A rewards kindness and punishes unkindness, defined by B's choice over payoffs. RBR responds to B's *beliefs about A's type*, not to B's actions. This distinction matters practically: kindness reciprocity is well-defined in game forms with clear action sets but becomes difficult to apply when actions are ambiguous, not yet taken, or unattributable - as in elections, where

voters form support decisions based on perceived candidate attitudes rather than observed deeds. RBR does not face this limitation.

Ellingsen and Johannesson (2008). The closest precursor is Ellingsen and Johannesson (2008): A values the esteem of B, and B's *own type* determines how much that esteem is worth. RBR reverses the relevant dimension: it is B's *belief about A's type* that drives A's behavior. The moderation prediction also inverts: while Ellingsen and Johannesson predict that esteem from high-type audiences matters most, RBR predicts that esteem matters most when it is *scarce* - when outside respect is low (Prediction 2), regardless of who provides it. Crucially, Ellingsen and Johannesson require B to observe A's action in order to form esteem; RBR requires neither observability nor any action by B.

Guilt and disappointment aversion (Charness and Dufwenberg 2006; Battigalli and Dufwenberg 2007) also place B's beliefs in A's utility. The key distinction is in *what* B's beliefs concern: in those models, B's beliefs are about A's future actions or payoffs, and A dislikes falling short of B's material expectations. In RBR, B's belief concerns A's type - a fixed characteristic. This difference yields a distinct identification strategy: shift B's beliefs about A's type while holding fixed any information relevant to B's expectations about outcomes, which is precisely what the experiment implements.

Type-based altruism (Levine 1998) has A's prosociality toward B depend on A's belief about B's type. RBR reverses the direction: it is B's belief about A's type that matters. The experimental design holds information about B's type constant across conditions, isolating the role of B's beliefs about A.

Social image concerns (Bernheim 1994; Bénabou and Tirole 2006; Bursztyn and Jensen 2017) motivate A to act in order to *influence* B's beliefs about her type. This channel is orthogonal to RBR: social image concerns address A's effort to earn esteem; RBR addresses A's behavioral response to already-perceived esteem. Social image effects require that A's action be observable by B; RBR does not.

In sum, RBR identifies a channel that is distinct from, and complementary to, all of

the above. The unifying insight is that second-order beliefs about types are a natural state variable: B's actions and messages matter insofar as they shift A's beliefs about what B thinks of A, and these beliefs directly affect A's intrinsic motivation to support B.

3.5 RBR in Voting

To make the framework concrete, consider the example of voting. Candidate B holds a belief about voter A's type - about the positive qualities of people like A. Voter A holds a second-order belief: how much does this candidate respect me and people like me? A's vote for B is the support action s; in large elections, it may be considered that its value is primarily expressive (Brennan and Buchanan 1984).[7] Perceived outside respect η captures how well A believes society at large regards people like her. We proxy this empirically by occupational prestige, a well-established measure of social standing.

Predictions 1 and 2 then yield two testable implications. First, the probability of voting for B increases in A's perceived respect from B. Second, this association is strongest for voters with low outside respect - those who perceive themselves as poorly regarded by society at large.

We establish the behavioral existence of RBR in a controlled experiment (Section 4), then test both predictions in the field (Section 5).

4 Experiment

4.1 The Challenge of Identification

Anecdotal and observational evidence for respect-based reciprocity is suggestive but not decisive. A voter who abandons a candidate after feeling disrespected may be acting on RBR - but she may, for example, also have updated her beliefs about the candidate's policy positions

7. Also, see Feddersen and Sandroni (2006), Callander and Carbajal (2022), Perego and Yuksel (2022), Vaeth (2023), Albuquerque et al. (2025), and Berry, Cox, and Haile (2025) for models of expressive voting.

(if she infers that disrespect signals different values). The awards and recognition literature faces analogous problems: an employee who works harder after receiving a prize may respond to the intrinsic utility of being respected, but the prize may also carry material value, signal a future wage trajectory, activate social image concerns in front of colleagues, or induce guilt at the prospect of disappointing a supervisor who has expressed high expectations (Kosfeld and Neckermann 2011; Bradler et al. 2016; Friebel et al. 2025).

More precisely, I consider five families of confounders that threaten identification of RBR in natural settings. *Outcome- and intention-based reciprocity:* standard models of reciprocity condition on actions or intentions of the counterpart (Rabin 1993; Falk and Fischbacher 2006). If the candidate's message constitutes an action that affects the voter's payoff or reflects the candidate's kindness, behavior will respond even in the absence of any belief-based preference. *Type-based and group-based preferences:* if A's support of B depends on A's assessment of B's type (Levine 1998), or on shared group membership (Chen and Li 2009), then learning that B holds a high belief about A may trigger support because it updates A's assessment of B's type, not because of the belief per se. *Disappointment and guilt aversion:* if A believes B expects a high allocation, A may give more to avoid the disutility of falling below B's expectations (Battigalli and Dufwenberg 2007) - a channel that operates through B's expectations about A's *action*, not about A's *type*. *Social image and reputation concerns:* if A's decision is observable - even partially - to B or to others, A may adjust behavior to manage impressions (Bénabou and Tirole 2006). *Instrumental value:* if B's belief affects some outcome that A cares about materially, the behavioral response requires no intrinsic component at all.

Isolating RBR requires an environment that varies A's perceived respect while holding fixed every element required by each alternative: B's actions, B's type, group membership, expectations about A's action, observability, and material stakes. The following design achieves this.

4.2 Design

Setting and roles. Participants are recruited online via Prolific and assigned to the role of participant A. Each A is paired with a passive participant B, who is fully anonymous: A receives no information about B's identity, demographics, group membership, or type. B takes no action in the experiment that affects A. The sole outcome is A's allocation of a bonus between herself and B.

Treatment: shifting perceived respect without shifting anything else. The treatment varies what A believes B believes about A's contribution to a bonus pool via effortful tasks. The procedure is as follows. First, A chooses how many effortful tasks to complete, with each completed task contributing a small amount to a bonus pool from which A does not benefit. This choice requires effort and benefits others, making it a proxy for A's type with regard to diligence and prosociality. Then, A is assigned a random color label (Purple or Orange) and informed that this label is randomly assigned. Next, A is introduced to B. B is described as a participant from a previous round who was not assigned a label and does not know that labels are randomly assigned. A is informed that B

- was informed that other participants are asked how many effortful tasks they are willing to complete to contribute to a bonus pool.
- observed two example pairs drawn from a large pool of participants. In one pair, the Purple-labeled participant contributed more tasks than the Orange-labeled participant; in the other pair, the Orange-labeled participant contributed more tasks than the Purple-labeled participant.
- was asked to predict whether, in a new pair, a participant with the same color as A would contribute more, the same or less than another randomly drawn participant.

The treatment consists of which B participant A is matched with and the belief that B recorded. In the *high-respect* condition, A learns that B recorded the belief that A "likely

contributes more than another random participant." In the *low-respect* condition, A learns that B recorded the belief that A "likely contributes less than another random participant."

Because the label is assigned at random and A knows this, the treatment conveys no information about A's actual type: the variation in perceived respect is orthogonal to A's own assessment of her ability. Instructions across the two conditions differ only in the description of B's belief; all other wording is identical. Across the entire survey, the treatment conditions differ in precisely two words.[8]

Eliminating the confounders. The design addresses each confounder in turn.

Outcome- and intention-based reciprocity require that B take an action that affects A or that A attribute a kind or unkind intention to B. Neither holds here: B takes no action; A's payoff does not depend on B's behavior; and A is explicitly informed of this. A comprehension check verifies that participants understand that B has not acted toward them and that no other participant can affect their (A's) payoff.

Type-based and group-based preferences require that A have information about B's type or group membership. B is kept fully neutral: A receives no such information. Moreover, B's belief is about relative contributions - neither treatment condition contains a more or less kind belief. From a rational point of view, B's belief is therefore not a diagnostic signal of B's kindness or altruism (Levine 1998): a high belief about A in this setting cannot be read as evidence that B is a generous or kind type.

Disappointment and guilt aversion require that A believe B has formed expectations about A's allocation. A is explicitly informed that B has no information about A's allocation decision and that B will not learn the source of any bonus. This is indicated twice on the allocation screen and tested as a comprehension check.

Social image and reputation concerns and *instrumental value* both require some channel through which A's allocation reaches B or affects A's material outcome. The experiment

8. Instructions were designed to minimize demand effects: participants are told explicitly that "people choose different amounts for different reasons," and comprehension checks are framed symmetrically.

ensures full anonymity: as just discussed, B is explicitly told that she will not learn the source of any bonus; she cannot identify A's decision. Again, participants are comprehension-checked on this anonymity guarantee. Under full anonymity, there is no audience and no material return to generosity: any behavioral response must be intrinsic.

Experimental flow. The survey proceeds in the following order:

1. Instructions and comprehension checks.
2. Contribution phase.
3. Random color label assignment.
4. Information display: A sees the example performance pairs that B observed, establishing the context for B's belief.
5. Treatment: A reads B's stated belief about A's performance (high or low).
6. Allocation decision: A divides a bonus between B and another random participant.
7. Manipulation checks: A reports her perceived respect from B on a 0–100 scale and states an incentivized belief about B's assessment of A's performance.

4.3 Implementation

The experiment is conducted online via Prolific, with gender quotas. The online environment offers two advantages relevant to the present design. First, it permits full control over interactions, which eliminates any channel through which information about B unintentionally leaks to A or information about A's decision leaks to B. Second, it plausibly constitutes a *hard test* for RBR: social image concerns are known to be weaker in anonymous online environments than in lab settings (Henry and Sonntag 2019). If RBR has behavioral bite even when social image concerns are attenuated - that is, when participants likely care less

about what others believe about their types - this strengthens the case for its significance. The experiment was pre-registered as AsPredicted 281,027.

4.4 Results

Sample and comprehension. The results reported here are based on a pre-registered sample of $N = 225$ participants. Of these 225 participants, 92 percent (206/225) passed all comprehension checks within two attempts - indicating that they understood the key design features: B's anonymity, the absence of B's actions toward A, and the full anonymity of the allocation decision - and were also flagged as authentic by Qualtrics's bot detection and Prolific's bot and LLM authenticity checks. As pre-registered, I focus on this subsample in the main analysis.

Manipulation check. The treatment produced a strong shift in perceived respect. In the low-respect condition, participants reported a mean perceived-respect score of 39.3 out of 100; in the high-respect condition, the mean was 75.0 out of 100 - a difference of 35.6 points on a 100-point scale ($p = 0.00$).

Main result. Consistent with Prediction 1, perceived respect increases allocations to B. In the low-respect condition, mean allocation was \$0.47; in the high-respect condition, mean allocation was \$0.63 - an average treatment effect of \$0.16, or approximately 35 percent of the low-condition baseline. This difference is statistically significant ($p = 0.00$) in a two-sample permutation test, which does not rely on distributional assumptions and is robust to outliers. A Wilcoxon rank-sum test yields very similar results ($p = 0.00$). Figure 1 displays the empirical cumulative distribution functions of the allocation by condition. The distribution in the high-respect condition first-order stochastically dominates the low-respect distribution across the entire support, with the largest treatment effects in the interior of the distribution. This distributional shift is confirmed by a two-sample Kolmogorov-Smirnov test ($p = 0.00$).

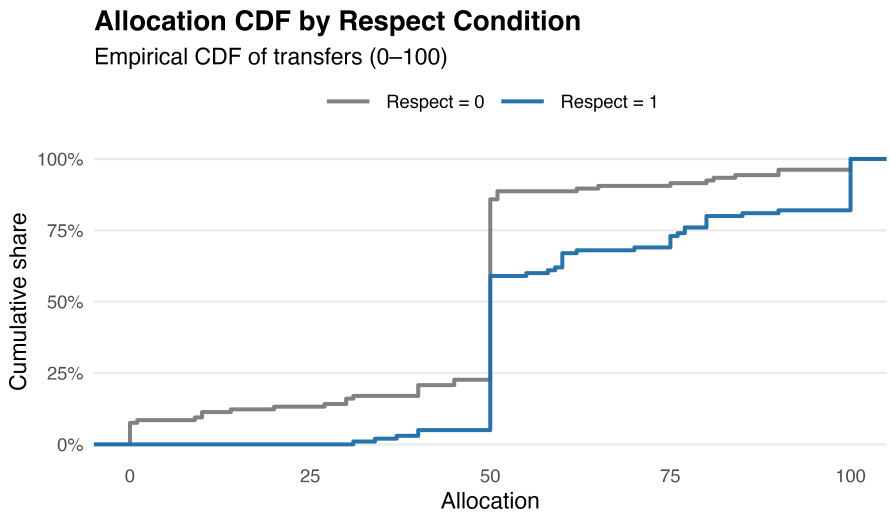


Figure 1: Empirical CDF of allocations to participant B, by treatment condition (low vs. high perceived respect). The high-respect distribution first-order stochastically dominates the low-respect distribution across the entire support ($N = 206$).

Mechanisms. I test whether the treatment effect operates through perceived respect - the channel predicted by H1 - using A's self-reported perceived respect from B (felt respect, 0–100 scale) as the primary pre-registered mediating variable.

Table 1 reports the instrumental-variable analysis. The first stage confirms that the treatment strongly shifts felt respect (first-stage F reported in the table). The 2SLS estimate indicates that a one-standard-deviation increase in felt respect (standardized using the control-group distribution) raises allocations substantially.

Table 1: IV estimates: treatment → felt respect → allocation. Felt respect is standardized using the control-group SD. HC2 standard errors.

	First stage	Reduced form	2SLS
Constant	0.000	46.811***	46.811***
	(0.097)	(2.071)	(2.030)
Respect treatment	1.478***	16.299***	
	(0.123)	(2.905)	
Felt respect (z)			11.025***
			(1.883)
Observations	206	206	206
R^2	0.412	0.133	0.202
Weak-IV F-statistic			142.77

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

Figure 2 provides a complementary visualization of the mediating role of felt respect. The top coefficient shows the reduced-form treatment effect from a bivariate regression. The bottom two coefficients come from a joint regression of allocation on both the treatment indicator and felt respect. Once felt respect is included, the treatment coefficient falls sharply and becomes imprecise, while felt respect remains strongly predictive of allocation. This pattern is consistent with the treatment operating primarily through the perceived-respect channel predicted by H1.

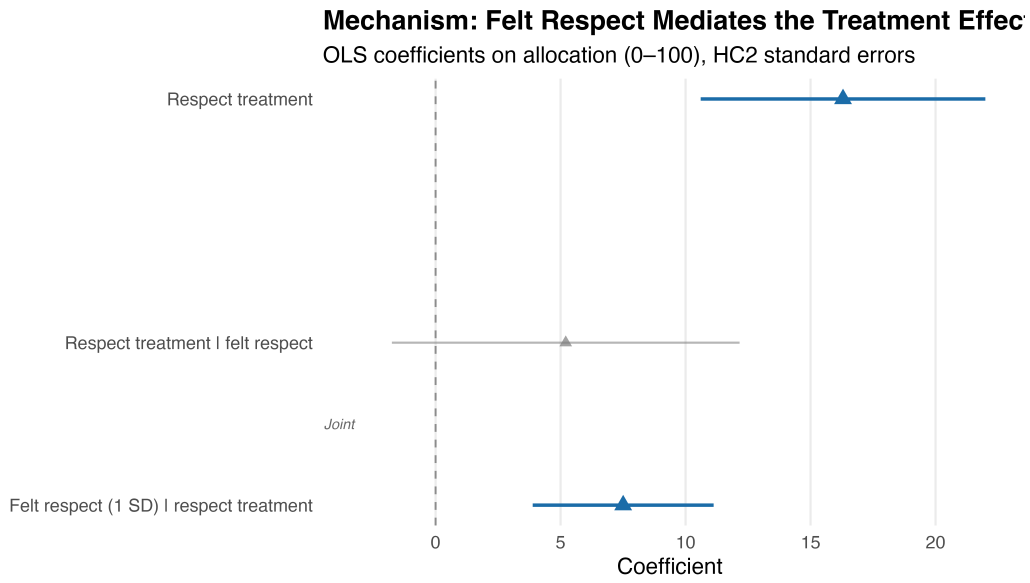


Figure 2: Mediation: the treatment effect on allocation attenuates when controlling for felt respect (z, standardized on control-group SD). Top: bivariate regression. Bottom two: joint regression. HC2 standard errors.

As a secondary mechanism measure, I use A's incentivized belief about B's assessment of A's contribution (B's contribution belief). The IV results are qualitatively similar and mediation results do not replicate the findings with the primary pre-registed mediating variable, which may be because B's contribution belief maps less directly to main concept of interest, percieved respect. These results are reported in Appendix Figure A2 and Table A1.

In summary, the experiment provides clean behavioral evidence for H1: perceived respect causally increases support, and the effect seems to operate through the hypothesized psychological channel.

5 Voting Evidence

RBR makes two predictions about voting. First, vote choice - an expressive support action cast under effective anonymity and without individual pivotality - should increase in perceived candidate respect. The experiment in Section 4 establishes that this preference exists behaviorally in a controlled setting. Second, this association should be strongest for voters

with low outside respect. I now test these predictions in survey data, exploiting a nationally representative survey that contains the rare combination of perceived respect measures and voting behavior. The evidence is observational; I interpret it as broadly consistent with RBR and as field confirmation of the experimental findings.

5.1 Data

The data come from the October 2016 Political Survey conducted by the Pew Research Center (Pew Research Center 2016), which interviewed a nationally representative sample of $N = 2{,}583$ U.S. adults between October 20–25, 2016, shortly before the presidential election.[9] Beyond detailed sociodemographics and policy views, the survey includes two perceived-respect items: "How much respect do you think Donald Trump [Hillary Clinton] has for people like you?" with responses on a four-point scale from *Not at all* (1) to *A great deal* (4). I construct the *perceived respect difference* as Trump's score minus Clinton's score, and standardize it (z-score, $\mu = 0$, $\sigma = 1$) for comparability across specifications. I am not aware of any other nationally representative pre-election survey that jointly measures perceived candidate respect and voting behavior, making this dataset uniquely suited for the analysis.

5.2 Perceived Respect and Vote Choice

Table 2 reports OLS regressions of a binary indicator for voting for Trump on the perceived respect measures. Columns (1) and (2) include no controls; columns (3) and (4) add a saturated set of fixed effects (similar to Ortoleva and Snowberg (2015) and Enke (2020)) for state, age, race, income, education, religious denomination, gender, employment status, and community type, as well as mean perceived respect. Across all specifications, perceived respect is among the strongest individual-level predictors of vote choice. A one-standard-

9. Interviews were conducted in English and Spanish by Princeton Data Source. As many as seven attempts were made to contact every sampled telephone number.

deviation increase in the respect difference is associated with a 34.8 percentage-point increase in the probability of voting for Trump in the fully controlled specification (column 4).

Table 2: Perceived Respect and Vote Choice: OLS with Robust Standard Errors

	Dependent variable: Vote for Trump			
	(1)	(2)	(3)	(4)
Respect Clinton (z)	-0.188^{***}		-0.182^{***}	
	(0.009)		(0.010)	
Respect Trump (z)	0.247^{***}		0.221^{***}	
	(0.009)		(0.009)	
Respect Difference (z)		0.375^{***}		0.348^{***}
		(0.005)		(0.007)
Sociodemographic FE	No	No	Yes	Yes
Mean Respect (z)	No	No	Yes	Yes
Observations	2,070	2,070	1,885	1,885
R^2	0.597	0.593	0.654	0.654

Note: Robust standard errors in parentheses. Sociodemographic FE include state, age, race, income, education, denomination, gender, employment, and community type.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

The most demanding specification conditions on a rich set of policy preferences. Because perceived respect may cause respondents to align with a candidate's policy positions, controlling for policy views may absorb part of the effect of respect; I nonetheless include this specification to assess robustness. Survey subsets differ in the policy questions asked, so I estimate the association separately for two data splits (df1, df2), including all available policy preferences alongside partisanship, ideology, and all socioeconomic controls. Figure 3 compares the coefficient on the respect difference with those on the selected policy controls in data split 1. Perceived respect remains the single strongest predictor of vote choice even in this demanding specification, with a coefficient of 0.156 (SE = 0.020, $p < 0.001$) in df1 and 0.151 (SE = 0.021, $p < 0.001$) in df2.

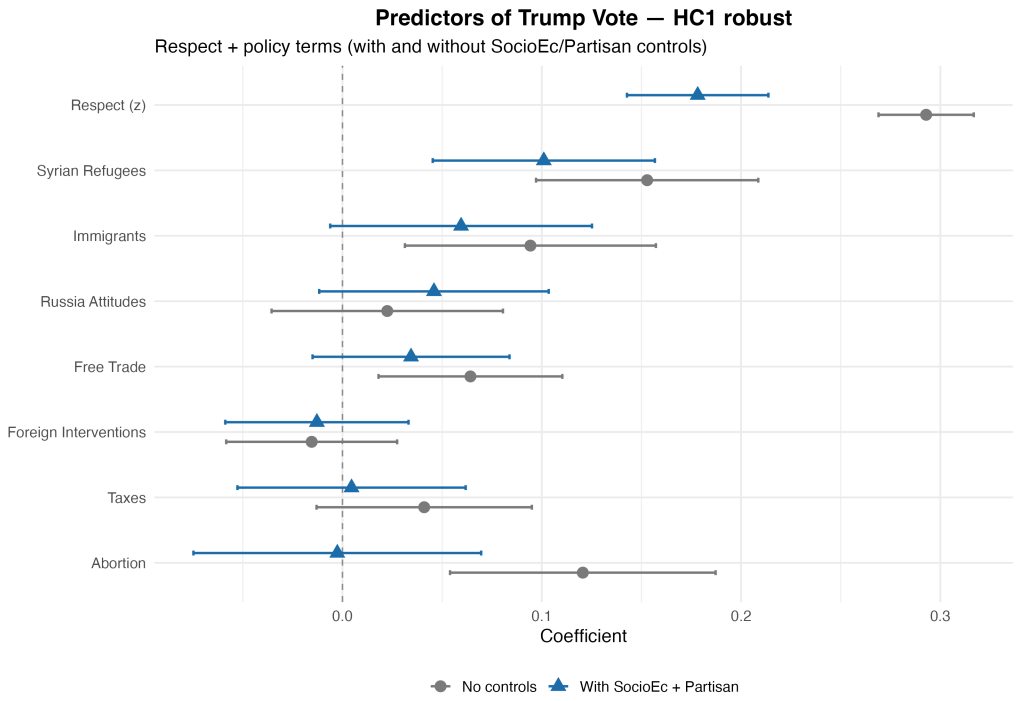


Figure 3: Coefficient on perceived respect difference relative to policy and demographic controls (data split 1). Perceived respect is the strongest predictor of vote choice even when conditioning on a rich set of policy preferences, partisanship, and sociodemographics. Horizontal bars are 95% confidence intervals (HC1 standard errors). See Appendix Figure A3 for data split 2.

5.3 Heterogeneity: Testing Diminishing Sensitivity

A distinctive prediction of the RBR framework concerns heterogeneity. Prediction 2 states that sensitivity to B's respect decreases in outside respect η. Applying this hypothesis to voting, I propose that voters who perceive themselves as poorly regarded by society at large should exhibit a *stronger* behavioral response to perceived candidate respect than voters who already feel well-regarded.

Proxy for outside respect. I construct a voter-level proxy for outside respect using occupational prestige. Occupational prestige captures the degree to which a person's role is positively regarded by society - a natural correspondent to η, the aggregate of others' beliefs about the individual's type. I first estimate the relationship between occupational prestige

and demographics (education, age, gender, employment status, region, and urbanicity) in the General Social Survey (GSS), then use the fitted values to predict expected prestige for each Pew respondent from their observable characteristics. Voters are then sorted into prestige quintiles Q1 (lowest) through Q5 (highest). Q1 voters are those whose predicted occupational standing - as a function of their education, age, gender, employment, region, and urbanicity - is lowest in the distribution; they are most likely to perceive themselves as poorly regarded by society.

Results. Figure 4 plots the estimated interaction coefficient on perceived respect for each prestige quintile (relative to Q5) in the double-selection LASSO specification.

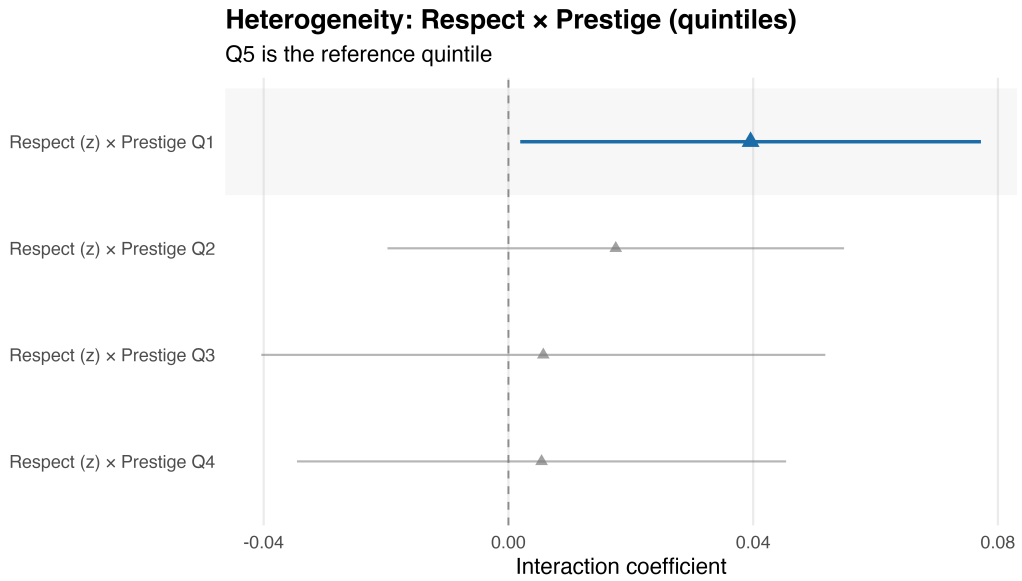


Figure 4: Sensitivity to perceived respect difference by predicted occupational prestige quintile (data split 1; double-selection LASSO, HC3 standard errors). Voters in the lowest prestige quintile (Q1) exhibit the strongest response to perceived candidate respect. Quintiles Q2–Q4 are statistically indistinguishable from Q5. See Appendix Figure A4 for data split 2.

The pattern is consistent with Prediction 2. In data split 1, the interaction coefficient for Q1 is +0.040 (SE = 0.019, $p < 0.05$) relative to Q5, against a main effect of 0.156 - implying that voters in the lowest prestige quintile are approximately 25 percent more sensitive to perceived candidate respect than those in the highest quintile. Quintiles Q2 through Q4

show interaction coefficients near zero and statistically insignificant. The pattern replicates in data split 2: the Q1 interaction is +0.036 (SE = 0.020, $p < 0.10$) against a main effect of 0.151; all other quintile interactions are small and insignificant.

The pattern in the prestige profile - with sensitivity concentrated in the lowest quintile - is consistent with Prediction 2: gains in outside respect matter most when outside respect is already scarce. This heterogeneity pattern is more difficult to rationalize with alternative mechanisms. An instrumental account of perceived respect - where voters use it as a heuristic for anticipated representation - would predict that sensitivity decreases primarily in political information, not in social prestige. I test this directly by interacting the respect difference with self-reported news engagement and election thought. Neither information moderator is significantly different from zero in either data split, while the prestige moderator is significant in both.[10] This asymmetry favors the interpretation that RBR indeed plays a role.

5.4 Group-Level Variation

Perceived respect varies systematically across social groups in the 2016 electorate. Figure 5 plots mean perceived respect by candidate and group with 95 percent confidence intervals. African Americans, Hispanic Americans, Asian Americans, women, college-educated individuals, and urban residents report substantially higher perceived respect from Clinton than from Trump. Rural voters and those without a college degree perceive greater respect from Trump. These patterns are noteworthy for at least two reasons. First, neither candidate ran as an incumbent, so voters were less likely to have inferred respect from past policy actions; the patterns therefore likely reflect, at least in part, the symbolic content of the candidates' rhetoric and public images. This makes the communication channel studied in Section 6.1 especially relevant.

10. In df1, the interaction of respect with "followed not too / not at all closely" is −0.078 (SE = 0.054, $p = 0.15$). In df2, the interaction with "little/no thought about election" is −0.045 (SE = 0.040, $p = 0.26$). In both cases the prestige Q1 interaction is significant while the information interaction is not.

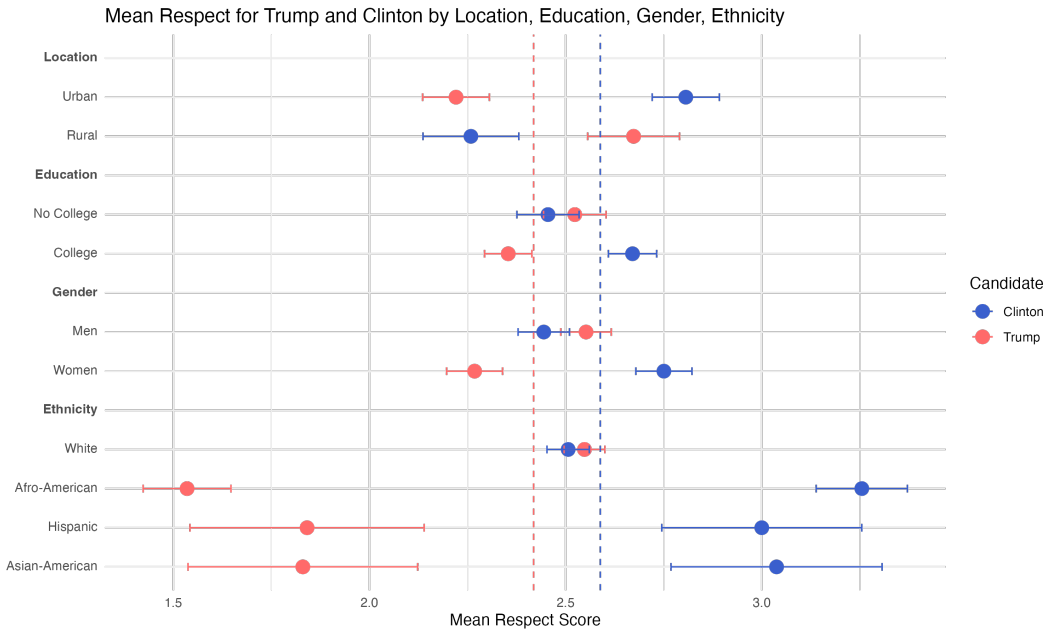


Figure 5: Mean perceived respect ("How much respect does [candidate] have for people like you?") by candidate and social group, with 95% confidence intervals. Pew 2016 ($N = 2{,}583$).

Second, these patterns align with group-level vote shares in 2016 and with much of the broader narrative of political realignment along educational and geographic lines (Calvo, Pons, and Shapiro 2024). RBR offers one micro-foundation for this alignment: if candidates can credibly signal differential respect across groups - through direct messages or through the symbolic content of their rhetoric - the cross-group variation in perceived respect will translate into cross-group variation in political support. The group-level patterns thus extend the individual-level evidence: not only does perceived respect predict voting across individuals, it varies across groups in ways that track the structure of partisan realignment. I provide a preliminary discussion of some supply-side implications - how candidates signal respect to different groups - in Section 6.

5.5 Summary

In summary, the voting evidence complements the experiment along three dimensions. First, perceived candidate respect is a powerful predictor of vote choice, with a magnitude compara-

ble to leading policy preferences even in demanding specifications. Second, this association is strongest among voters with the lowest predicted outside respect - consistent with the diminishing-sensitivity prediction (H2) and difficult to explain through instrumental learning about representation alone. Third, group-level variation in perceived respect tracks the structure of partisan realignment, suggesting that RBR may help connect the macro-level narratives of recognition gaps and status anxiety to individual voting behavior. The evidence is observational and does not establish causality; its value lies in showing that the mechanism identified in the experiment organizes politically consequential behavior in the field.

6 Applications

RBR has general implications beyond the experimental and observational contexts studied above. This section develops two. The first concerns communication: when claims about respect cannot be verified, recognition is cheap talk (Crawford and Sobel 1982). Section 6.1 shows that multi-audience settings can restore credibility - bashing one group can serve as a credible signal of respect toward another - and derives conditions under which bashing becomes more prevalent, consistent with a recent rise in populism. The second concerns policy: I show that RBR extends the conditions under which policies shape agent behavior beyond what standard signaling logic requires, because the signaling value of a policy persists even when the principal holds no private information.

6.1 Communication

Recognition is cheap talk. Because support increases in perceived respect (H1), any sender B who wishes to influence A's behavior has a strong incentive to claim a favorable belief about A's type. Yet if communication is costless and unverifiable, B can always send the most flattering message regardless of her true belief. By the standard argument of

Crawford and Sobel (1982), no information is transmitted in equilibrium: rational receivers anticipate that all senders have incentive to report the highest possible $\hat{\theta}$, so no message is informative. Recognition therefore suffers from a credibility problem that is intrinsic to RBR. The same preference that makes recognition valuable to the receiver makes it generically incredible to convey.

When is communication credible? Several mechanisms can restore partial information transmission. First, if B bears an intrinsic or other cost of misrepresentation - e.g., a discomfort with stating beliefs she does not hold - a sufficiently high lying cost creates incentive compatibility: only senders who genuinely hold a favorable belief are willing to incur the cost of claiming it (Kartik 2009). Second, when B communicates to multiple audiences simultaneously, the comparative structure of her messages can be informative. Claiming to respect group A_1 unconditionally is cheap; doing so while publicly expressing a low view of group A_2 is more costly for an opportunist if A_2's support also has value. Comparative cheap talk logic (Chakraborty and Harbaugh 2010; Lipnowski and Ravid 2020) implies that *relative* recognition - praising A_1 at the expense of A_2 - can convey what unconditional praise cannot. Third, when receivers are not fully Bayesian or fail to account for the sender's incentives to deceive, credulity can sustain partial communication even absent lying costs (Kartik, Ottaviani, and Squintani 2007; Madarász 2014).

Bashing as a credibility device. The multi-audience mechanism has an implication that is, I argue, underappreciated: bashing can be a rational credibility device for expressing respect toward another group. I illustrate this in a simple example and provide a fuller characterization in Appendix C. The setup features a candidate B with a three-dimensional type which is unobserved by two groups of voters, A_1 and A_2, with respective population shares p and $1-p$. B's first two type dimensions correspond to her beliefs about the two voter groups - her respect for A_1 and A_2. Third, B is either honest - bearing a high lying cost λ^H, by assumption always communicating truthfully - or opportunistic, bearing a low

lying cost λ^L and maximizing vote share. Voters have priors about B's respect for each group and a prior about B's lying cost, which I argue corresponds broadly to a notion of trust $\pi = \Pr(\lambda^H)$. Both voter groups update their beliefs about B's respect for them from her messages, taking those messages literally (Kartik, Ottaviani, and Squintani 2007). B can send messages corresponding to claiming high or low respect for each group, yielding four possible message profiles. Voters update as Bayesians, and the vote share in each group depends on the posterior probability that B respects them.

The order is as follows:

1. Nature draws B's type: her respect for A_1, her respect for A_2, and her lying cost.

2. B observes her type and sends a message about her respect for each group.

3. Voters in A_1 and A_2 update their beliefs about B's respect for them and vote accordingly.

In the following I focus on a pooling equilibrium in which all opportunistic candidates claim to respect both groups because its properties and existence conditions are interesting and parsimoniously rationalize a variety of relevant patterns observed in contemporary politics. Three remarks on this equilibrium are worth highlighting; the latter two are comparative statics that follow from the conditions under which the pooling equilibrium unravels:

1. *Bashing signals respect.* In the pooling equilibrium, all opportunistic candidates claim to respect both groups, and voters cannot distinguish them from honest types who genuinely do. A candidate B who respects A_1 but bashes A_2 is, on the contrary, certain to authentically respect A_1 (Remark 1, Appendix C).

2. *Low trust induces bashing.* The pooling equilibrium requires trust π to exceed a threshold. As π falls - as voters grow more skeptical of politicians' sincerity - deviations to contrasting messages become more attractive, and opportunistic candidates may deviate to bash one group to signal credible respect for the other (Remark 2, Appendix C).

3. *Low prior respect induces bashing.* Similarly, when voters' prior belief that politicians genuinely hold them in high regard, θ, is low, universal recognition is generally less credible. This strengthens the incentive to differentiate through bashing, making negative out-group messaging more prevalent precisely when groups feel most disrespected (Remark 3, Appendix C).

The mechanism provides a parsimonious rationalization for several observed patterns. First, at least some voters appear to perceive uncivil behavior - likely akin to bashing - as a positive signal about a candidate, consistent with the first remark.

Second, trust in politicians has declined substantially in most advanced democracies over recent decades, and this decline is associated with a rise in populism (Algan et al. 2017; Bellodi et al. 2023). Populism is commonly characterized in part by rhetorical bashing of elites or minorities - a pattern that maps naturally to disrespect in the model. The second remark provides a micro-foundation: as voters grow more skeptical of politicians' sincerity, bashing becomes a more attractive strategy for signaling credible respect to a target group, increasing the prevalence of negative messaging in equilibrium. Finally, when groups anticipate being less respected by political elites, another notion that one may consider broadly captured by some survey measures of trust, bashing becomes more attractive as a signal of respect, thus providing a further micro-foundation for the observed association between low trust and negative out-group messaging in electoral politics. Both forces, according to the model, push toward more bashing in equilibrium - not because candidates prefer conflict, but because universal recognition is not the obvious equilibrium strategy for opportunistic candidates anymore.

6.2 Policy Signaling

A rich literature documents that policies carry signaling value beyond their material effects (Sunstein 1996; Bénabou and Tirole 2003; Benabou and Tirole 2026): a rule, wage, or fine can crowd in or crowd out intrinsic motivation depending on what it signals to the agent

about the principal's information (see Bowles and Polania-Reyes 2012, for a review). The canonical formalization, due to Bénabou and Tirole (2003), requires that the *principal have private information* about the agent's type or the nature of the task. A high fixed wage signals that the employer believes the worker is intrinsically motivated or "trustworthy"; performance pay signals uncertainty or pessimism about effort absent additional incentives; monitoring signals suspicion. The agent updates her self-image or her beliefs about the task from the policy and adjusts her intrinsic motivation accordingly. Crucially, the authors note: "when concerned about a potential negative impact of rewards, one should first check whether the reward provider has private information about the task or the agent's talent" (492). This condition has been extended by Sliwka (2007) and others who propose that the principal's private information can also pertain to a norm in the social environment and that a policy can signal the principal's information about the norm, which in turn affects the agent's motivation. In either case, the principal's informational advantage is key to the signaling mechanism: if the principal is no better informed than the agent, her policy choice does not convey any new information and thus, cannot affect the agent's perceived respect or motivation.

RBR changes this condition critically. What drives A's behavior is A's *belief about* B's belief about A's type - perceived respect. For a policy to affect perceived respect, it is sufficient that (i) the principal's optimal policy depends on her belief about the agent's type, and (ii) the agent draws inferences about the principal's belief from the policy chosen. Private information on the principal's side is *not* required. Even a principal who is no better informed than the agent - indeed, one whom the agent *knows* to be uninformed - affects perceived respect by virtue of the policy she selects, because the policy choice reveals the principal's prior about the agent's type.

Three canonical classes of policies illustrate the mechanism, each formalized in Appendix D.

Gift exchange. An unconditional transfer from principal to agent - an above-market wage, an unprompted bonus, an advance of resources, or an award - can signal that the principal regards the agent as productive. Assume it is common knowledge that the agent reciprocates respect, e.g., corresponding to the principal's belief about the agent's ability and thus productivity. Now, the value of the agent reciprocating positively depends on the agent's ability, and thus, a gift is only worth it to an agent who the principal expects to be sufficiently able. Thus, the transfer is informative about the principal's belief about the agent's type: making a costly gift is rational only when the expected benefit of the effort it elicits is large enough, which requires the principal to hold a sufficiently favorable prior.

This raises the agent's perceived respect and, by H1, her intrinsic motivation to exert effort. In a separating equilibrium, principals who have received a good signal about the agent's type make gifts; those who have received a bad signal do not. This induces effort precisely from gifted agents, generating the gift-exchange pattern - higher unconditional pay leads to higher effort (Akerlof 1982) - without requiring the principal to be better informed overall.

Unlike accounts based on intentions-based reciprocity (Rabin 1993), this mechanism does not require the principal to have benevolent motives. A principal who gifts strategically - purely to elicit effort - still conveys relevant information under RBR. The informational content of the gift is independent of the principal's intentions. This resolves the concern raised by Koszegi (2014) and Netzer and Schmutzler (2014) that gift exchange cannot be explained by Rabin-type reciprocity when principals are self-interested, since in those models a strategic gift is not "kind." Appendix D.1 formalizes the separating equilibrium and conditions.

Incentive schemes. When a principal ties compensation to output - a performance bonus or a fine for misbehavior - she reveals that she anticipated the agent might not act desirably absent external incentives. This signals a low prior on the agent's intrinsic type, reducing

perceived respect. By H1, intrinsic motivation falls, partially offsetting the direct effect of the incentive. The minimum bonus needed to induce effort may therefore be *larger* than it would be if the incentive scheme did not shift the agent's perceived respect: the principal must overcome both the gap between cost and benefit *and* the RBR-induced reduction in intrinsic motivation that the scheme itself creates. Similarly, a fine for misbehavior signals an expectation of noncompliance, which reduces perceived respect and partly erodes the intrinsic compliance motivation the fine was intended to preserve - a mechanism complementary to the "fine as a price" arguments of Gneezy and Rustichini (2000). Critically, again, these effects persist under RBR even when the agent knows the principal to be uninformed, since what matters is the inference about the principal's prior, not the principal's informational advantage. Appendix D.2 formalizes the intuition and derives the optimal bonus and fine.

Control and delegation. Restricting the agent's action set - through monitoring, supervision, or mandatory approval - signals that the principal believes the agent would choose poorly if given autonomy: a low prior on cooperative or diligent type. The agent observes the degree of control, updates perceived respect downward, and reduces effort through the RBR channel in addition to the direct constraint. Conversely, extending autonomy signals a high type prior, raising perceived respect and intrinsic effort. This generates the "hidden cost of control" (Falk and Kosfeld 2006): removing the agent's option to cooperate voluntarily reduces cooperation, because the act of restricting reveals a low prior regardless of whether the restriction binds. Fehr and Rockenbach (2003) show that the mere availability of a sanction reduces back-transfers in trust games even when the sanction is not invoked, consistent with the RBR channel operating through the revealed distrust rather than through the sanction's material threat. More broadly, the notion that trust breeds trustworthiness, as observed in various forms of the trust game since Berg, Dickhaut, and McCabe (1995), is readily rationalized by RBR: sending a high amount signals a favorable belief about the partner's type, raising perceived respect and thus the partner's motivation to reciprocate. Friebel

et al. (2025) provide a field example in which managers who reduce controls observe greater productivity, consistent with the RBR mechanism. Appendix D.3 formalizes the intuition and characterizes when full delegation is optimal even when monitoring is costless.

Broader scope. The RBR mechanism is especially relevant in settings where agents know they have more information about their own type than the principal does - precisely the case where the Bénabou and Tirole (2003) condition fails and standard signaling logic predicts no crowding-out. Under RBR, crowding-out persists because the principal's prior, revealed through the policy, matters for perceived respect regardless of informational asymmetries. This is consistent with evidence from professional and public sector environments where performance pay and monitoring reduce intrinsic motivation even when workers know themselves to be well-qualified (see Ashraf and Bandiera 2018). The logic extends to civic contexts: a fine signals that the policymaker expects noncompliance, eroding the civic duty motivation that supplements the fine's deterrence effect; paternalistic regulations signal doubt about citizens' judgment, reducing the sense of civic agency through RBR.

7 Conclusion

This paper introduces *respect-based reciprocity*: people are more willing to support an actor when they believe that actor holds them in higher regard. It shows that this mechanism can be isolated experimentally in a controlled setting, and that it helps explain meaningful variation in political choice in survey data. The contribution is therefore not only to document that respect matters, but to show *how* and *for whom* it matters. In particular, the evidence suggests that perceived respect is most consequential where outside respect is scarcest, providing a microfoundation for the political importance of recognition gaps and a parsimonious link between social standing and political realignment.

More broadly, the paper shows how political economy can study a neglected but tractable channel of behavior: people respond not only to material promises or policy positions, but

also to the regard conveyed by political actors' messages, policies, and actions. This helps illuminate when negative out-group messaging can serve as a signal of regard, and it suggests more generally that institutions shape behavior not only by changing incentives, but also by expressing beliefs about the people they address.

References

Akerlof, George A. 1982. "Labor contracts as partial gift exchange." *The quarterly journal of economics* 97 (4): 543–569.

Albuquerque, Amanda de, Frederico Finan, Anubhav Jha, Laura Karpuska, and Francesco Trebbi. 2025. *Decoupling Taste-Based versus Statistical Discrimination in Elections.* Technical report. National Bureau of Economic Research.

Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari. 2017. "The European trust crisis and the rise of populism." *Brookings papers on economic activity* 2017 (2): 309–400.

Ashraf, Nava, and Oriana Bandiera. 2018. "Social incentives in organizations." *Annual Review of Economics* 10 (1): 439–463.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. "Guilt in games." *American Economic Review* 97 (2): 170–176.

———. 2022. "Belief-dependent motivations and psychological game theory." *Journal of Economic Literature* 60 (3): 833–882.

Bellodi, Luca, Massimo Morelli, Antonio Nicolo, and Paolo Roberti. 2023. *The shift to commitment politics and populism: Theory and evidence.* BAFFI CAREFIN, Centre for Applied Research on International Markets Banking ...

Benabou, Roland, and Jean Tirole. 2026. "Laws and norms." *Journal of Political Economy* 134 (2): 000–000.

Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and extrinsic motivation." *The review of economic studies* 70 (3): 489–520.

———. 2006. "Incentives and prosocial behavior." *American economic review* 96 (5): 1652–1678.

———. 2009. "Over my dead body: Bargaining and the price of dignity." *American Economic Review* 99 (2): 459–465.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, reciprocity, and social history." *Games and economic behavior* 10 (1): 122–142.

Bernheim, B Douglas. 1994. "A theory of conformity." *Journal of political Economy* 102 (5): 841–877.

Berry, Steven T, Christian Cox, and Philip Haile. 2025. *Selective turnout, voting policy, and partisan bias: Evidence from multi-level data.* Technical report. National Bureau of Economic Research.

Besley, Timothy, and Torsten Persson. 2019. "The rise of identity politics." *LSE Documents.*

Bowles, Samuel, and Sandra Polania-Reyes. 2012. "Economic incentives and social preferences: substitutes or complements?" *Journal of economic literature* 50 (2): 368–425.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non. 2016. "Employee recognition and performance: A field experiment." *Management Science* 62 (11): 3085–3099.

Braghieri, Luca. 2023. "Biased decoding and the foundations of communication."

Brennan, Geoffrey, and James Buchanan. 1984. "Voter choice: Evaluating political alternatives." *American Behavioral Scientist* 28 (2): 185–201.

Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao. 2018. "Status goods: experimental evidence from platinum credit cards." *The Quarterly Journal of Economics* 133 (3): 1561–1595.

Bursztyn, Leonardo, and Robert Jensen. 2017. "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure." *Annual Review of Economics* 9:131–153.

Callander, Steven, and Juan Carlos Carbajal. 2022. "Cause and effect in political polarization: A dynamic analysis." *Journal of Political Economy* 130 (4): 825–880.

Calvo, Richard, Vincent Pons, and Jesse M Shapiro. 2024. *Pitfalls of demographic forecasts of us elections.* Technical report. National Bureau of Economic Research.

Chakraborty, Archishman, and Rick Harbaugh. 2010. "Persuasion by cheap talk." *American Economic Review* 100 (5): 2361–2382.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and partnership." *Econometrica* 74 (6): 1579–1601.

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

Chen, Yan, and Sherry Xin Li. 2009. "Group identity and social preferences." *American Economic Review* 99 (1): 431–457.

Clinton, Hilary. 2016. Basket of deplorables. Https://web.archive.org/web/20210218215534/https://www.y [Accessed: 2024-07-30].

Crawford, Vincent P, and Joel Sobel. 1982. "Strategic information transmission." Https://web.archive.org/v [Accessed: 2024-07-30], *Econometrica: Journal of the Econometric Society,* 1431–1451.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9trSM [Accessed: 2024-07-30], *Games and economic behavior* 47 (2): 268–298.

Ellingsen, Tore, and Magnus Johannesson. 2007. "Paying respect." Https://web.archive.org/web/202102182 [Accessed: 2024-07-30], *Journal of Economic Perspectives* 21 (4): 135–149.

———. 2008. "Pride and prejudice: The human side of incentive theory." Https://web.archive.org/web/2021 [Accessed: 2024-07-30], *American economic review* 98 (3): 990–1008.

Enke, Benjamin. 2020. "Moral values and voting." Https://web.archive.org/web/20210218215534/https://v [Accessed: 2024-07-30], *Journal of Political Economy* 128 (10): 3679–3729.

Falk, Armin, and Urs Fischbacher. 2006. "A theory of reciprocity." Https://web.archive.org/web/202102182 [Accessed: 2024-07-30], *Games and economic behavior* 54 (2): 293–315.

Falk, Armin, and Michael Kosfeld. 2006. "The hidden costs of control." Https://web.archive.org/web/20210 [Accessed: 2024-07-30], *American Economic Review* 96 (5): 1611–1630.

Feddersen, Timothy, and Alvaro Sandroni. 2006. "A theory of participation in elections." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9trSM [Accessed: 2024-07-30], *American Economic Review* 96 (4): 1271–1282.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and retaliation: The economics of reciprocity." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJV [Accessed: 2024-07-30], *Journal of economic perspectives* 14 (3): 159–182.

Fehr, Ernst, and Bettina Rockenbach. 2003. "Detrimental effects of sanctions on human altruism." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9 [Accessed: 2024-07-30], *Nature* 422 (6928): 137–140.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A theory of fairness, competition, and cooperation." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9 [Accessed: 2024-07-30], *The quarterly journal of economics* 114 (3): 817–868.

Fraser, Nancy. 2000. "Rethinking recognition." Https://web.archive.org/web/20210218215534/https://ww [Accessed: 2024-07-30], *New left review* 3 (3): 107.

Friebel, Guido, Matthias Heinz, Mitchell Hoffman, Tobias Kretschmer, and Nick Zubanov. 2025. *Is this really kneaded? Identifying and eliminating potentially harmful forms of workplace control.* Technical report 3. Https://web.archive.org/web/20210218215534/https://www.you [Accessed: 2024-07-30]. National Bureau of Economic Research.

Fukuyama, Francis. 2018. *Identity: The demand for dignity and the politics of resentment.* 3:107. 3. Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJV [Accessed: 2024-07-30]. Farrar, Straus / Giroux. Techreport.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological games and sequential rationality." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch [Accessed: 2024-07-30], *Games and economic Behavior* 1 (1): 60–79.

Gennaioli, Nicola, and Guido Tabellini. 2023. "Identity politics." Https://web.archive.org/web/2021021821 [Accessed: 2024-07-30], *Available at SSRN 4395173* 1 (1): 60–79.

Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty. 2022. "Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9trSM [Accessed: 2024-07-30], *The Quarterly Journal of Economics* 137 (1): 1–48.

Gidron, Noam, and Peter A Hall. 2017. "The politics of social status: Economic and cultural roots of the populist right." Https://web.archive.org/web/20210218215534/https://www.youtube.com [Accessed: 2024-07-30], *The British journal of sociology* 68 (1): S57–S84.

Gneezy, Uri, and Aldo Rustichini. 2000. "A fine is a price." Https://web.archive.org/web/20210218215534/h
[Accessed: 2024-07-30], *The journal of legal studies* 29 (1): 1–17.

Hegel, Georg Wilhelm Friedrich. 1830. *Enzyklopädie der philosophischen Wissenschaften im Grundrisse (1830)*. 33:1–17. 1. Https://web.archive.org/web/20210218215534/https://www.youtube.c
[Accessed: 2024-07-30]. Felix Meiner Verlag. Techreport.

Henry, Emeric, and Jan Sonntag. 2019. "Measuring image concern." Https://web.archive.org/web/2021021
[Accessed: 2024-07-30], *Journal of Economic Behavior & Organization* 160 (1): 19–39.

Hochschild, Arlie Russell. 2018. *Strangers in their own land: Anger and mourning on the American right*. 160:19–39. 1. Https://web.archive.org/web/20210218215534/https://www.youtube.co
[Accessed: 2024-07-30]. The New Press. Techreport.

———. 2024. *Stolen pride: Loss, shame, and the rise of the right*. 160:19–39. 1. Https://web.archive.org/web
[Accessed: 2024-07-30]. The New Press. Techreport.

Honneth, Axel. 2001. "Recognition or redistribution?" Https://web.archive.org/web/20210218215534/https
[Accessed: 2024-07-30], *Theory, Culture & Society* 18 (2-3): 43–55.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion." Https://web.archive.org/web/2021
[Accessed: 2024-07-30], *American Economic Review* 101 (6): 2590–2615.

Kartik, Navin. 2009. "Strategic communication with lying costs." Https://web.archive.org/web/2021021821
[Accessed: 2024-07-30], *The Review of Economic Studies* 76 (4): 1359–1395.

Kartik, Navin, Marco Ottaviani, and Francesco Squintani. 2007. "Credulity, lies, and costly talk." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9
[Accessed: 2024-07-30], *Journal of Economic theory* 134 (1): 93–116.

Kosfeld, Michael, and Susanne Neckermann. 2011. "Getting more work for nothing? Symbolic awards and worker performance." Https://web.archive.org/web/20210218215534/https://www.youtub
[Accessed: 2024-07-30], *American Economic Journal: Microeconomics* 3 (3): 86–99.

Koszegi, Botond. 2014. "Behavioral contract theory." Https://web.archive.org/web/20210218215534/https:
[Accessed: 2024-07-30], *Journal of Economic Literature* 52 (4): 1075–1118.

Lamont, Michèle. 2018. "Addressing recognition gaps: Destigmatization and the reduction of inequality." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCH
[Accessed: 2024-07-30], *American Sociological Review* 83 (3): 419–444.

Lamont, Michèle, Bo Yun Park, and Elena Ayala-Hurtado. 2017. "Trump's electoral speeches and his appeal to the American white working class." Https://web.archive.org/web/20210218215534/ht
[Accessed: 2024-07-30], *The British journal of sociology* 68 (3): S153–S180.

Levine, David K. 1998. "Modeling altruism and spitefulness in experiments." Https://web.archive.org/web/
[Accessed: 2024-07-30], *Review of economic dynamics* 1 (3): 593–622.

Lipnowski, Elliot, and Doron Ravid. 2020. "Cheap talk with transparent motives." Https://web.archive.org/
[Accessed: 2024-07-30], *Econometrica* 88 (4): 1631–1660.

Madarász, Kristóf. 2014. "Limited Perspective Taking in Strategic Communication." Https://web.archive.or
[Accessed: 2024-07-30], *Available at SSRN 5844722* 88 (4): 1631–1660.

Maner, Jon K, C Nathan DeWall, Roy F Baumeister, and Mark Schaller. 2007. "Does social exclusion motivate interpersonal reconnection? Resolving the" porcupine problem."." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9trSM
[Accessed: 2024-07-30], *Journal of personality and social psychology* 92 (1): 42.

McMillan, Ron. 2002. *Crucial Conversations: Tools for Talking When Stakes are High.* 92:42. 1. Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJV
[Accessed: 2024-07-30]. McGraw-Hill. Techreport.

Netzer, Nick, and Armin Schmutzler. 2014. "Explaining gift-exchange—the limits of good intentions." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJ
[Accessed: 2024-07-30], *Journal of the European Economic Association* 12 (6): 1586–1616.

Obama, Barack. 2020. *A promised land.* 12:1586–1616. 6. Https://web.archive.org/web/20210218215534/ht
[Accessed: 2024-07-30]. Penguin UK. Techreport.

Ortoleva, Pietro, and Erik Snowberg. 2015. "Overconfidence in political behavior." Https://web.archive.org/
[Accessed: 2024-07-30], *American Economic Review* 105 (2): 504–535.

Perego, Jacopo, and Sevgi Yuksel. 2022. "Media competition and social disagreement." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJVE9trSM
[Accessed: 2024-07-30], *Econometrica* 90 (1): 223–265.

Pew Research Center. 2016. *As Election Nears, Voters Divided Over Democracy and Respect.* Technical report. Accessed: Date you accessed the article, 1. National Bureau of Economic Research. https://www.pewresearch.org/politics/2016/10/27/as-election-nears-voters-divided-over-democracy-and-respect/.

Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics." Https://web.archive.org/v
[Accessed: 2024-07-30], *The American economic review* 90 (1): 1281–1302.

Shayo, Moses. 2020. "Social identity and economic policy." Https://web.archive.org/web/20210218215534/h
[Accessed: 2024-07-30], *Annual Review of Economics* 12 (1): 355–389.

Sliwka, Dirk. 2007. "Trust as a signal of a social norm and the hidden costs of incentive schemes." Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJ
[Accessed: 2024-07-30], *American Economic Review* 97 (3): 999–1012.

Sunstein, Cass R. 1996. "On the expressive function of law." Https://web.archive.org/web/20210218215534/
[Accessed: 2024-07-30], *University of Pennsylvania law review* 144 (5): 2021–2053.

Taylor, Charles. 1995. "The politics of recognition." In *The politics of recognition,* 144:2021–2053. 5. Https://web.archive.org/web/20210218215534/https://www.youtube.com/watch?v=PCHJV
[Accessed: 2024-07-30]. Routledge.

Tomova, Livia, Kimberly L Wang, Todd Thompson, Gillian A Matthews, Atsushi Takahashi, Kay M Tye, and Rebecca Saxe. 2020. "Acute social isolation evokes midbrain craving responses similar to hunger." Https://web.archive.org/web/20210218215534/https://www.youtube.com/ [Accessed: 2024-07-30], *Nature neuroscience* 23 (12): 1597–1605.

Trump, Donald J. 2016. Remarks at the US Cellular Center in Asheville, North Carolina. Technical report. Https://www.presidency.ucsb.edu/documents/remarks-the-us-cellular-center-asheville-north-carolina [Accessed: 24-08-05], 12. National Bureau of Economic Research.

Vaeth, Martin. 2023. "Rational Voter Learning, Issue Alignment, and Polarization." Https://www.presidenc the-us-cellular-center-asheville-north-carolina [Accessed: 24-08-05], *Issue Alignment, and Polarization (November 21, 2023)* 23 (12): 1597–1605.

Zink, Caroline F, Yunxia Tong, Qiang Chen, Danielle S Bassett, Jason L Stein, and Andreas Meyer-Lindenberg. 2008. "Know your place: neural processing of social hierarchy in humans." Https://www.presidency.ucsb.edu/documents/remarks-the-us-cellular-center-asheville-north-carolina [Accessed: 24-08-05], *Neuron* 58 (2): 273–283.

Zollinger, Delia, and Silja Haeusermann. 2024. "Social status, recognition entitlement, or relative deprivation? Social hierarchies and the far right vote." Https://www.presidency.ucsb.edu/documen the-us-cellular-center-asheville-north-carolina [Accessed: 24-08-05], *Neuron* 58 (2): 273–283.

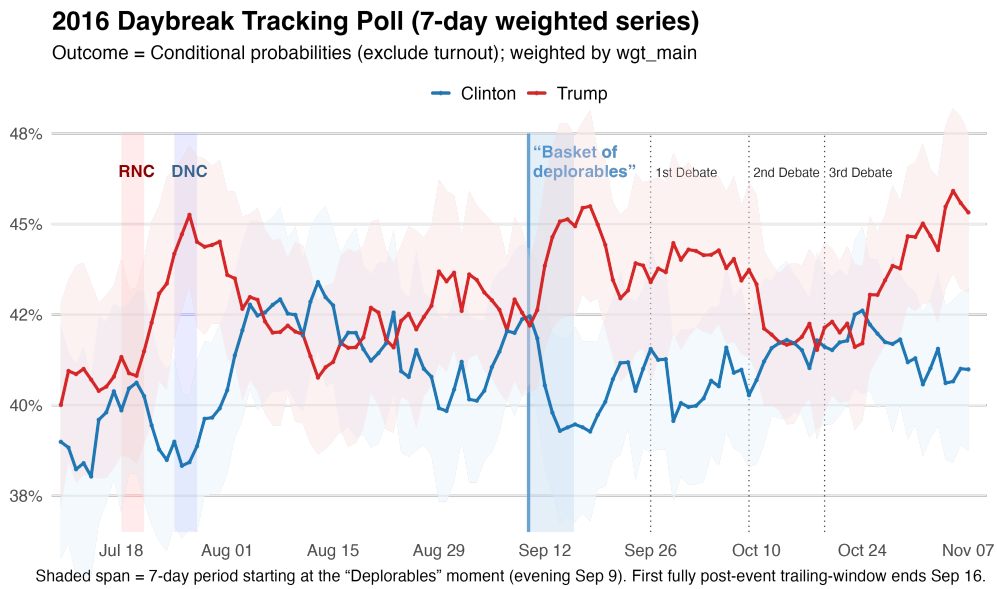


Figure A1: USC/LA Times Daybreak tracking poll, 2016 presidential election. The "basket of deplorables" remark (dashed line) coincided with a visible shift in support toward Trump (red) and away from Clinton (blue).

A Experiment

Table A1: IV estimates: treatment → B's contribution belief → allocation. B's contribution belief is standardized using the control-group SD. HC2 standard errors.

	First stage	Reduced form	2SLS
Constant	0.000	46.811***	46.811***
	(0.097)	(2.071)	(2.512)
Respect treatment	1.109***	16.299***	
	(0.128)	(2.905)	
B's contribution belief (z)			14.694***
			(3.116)
Observations	206	206	206
R^2	0.267	0.133	−0.232
Weak-IV F-statistic			74.35

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

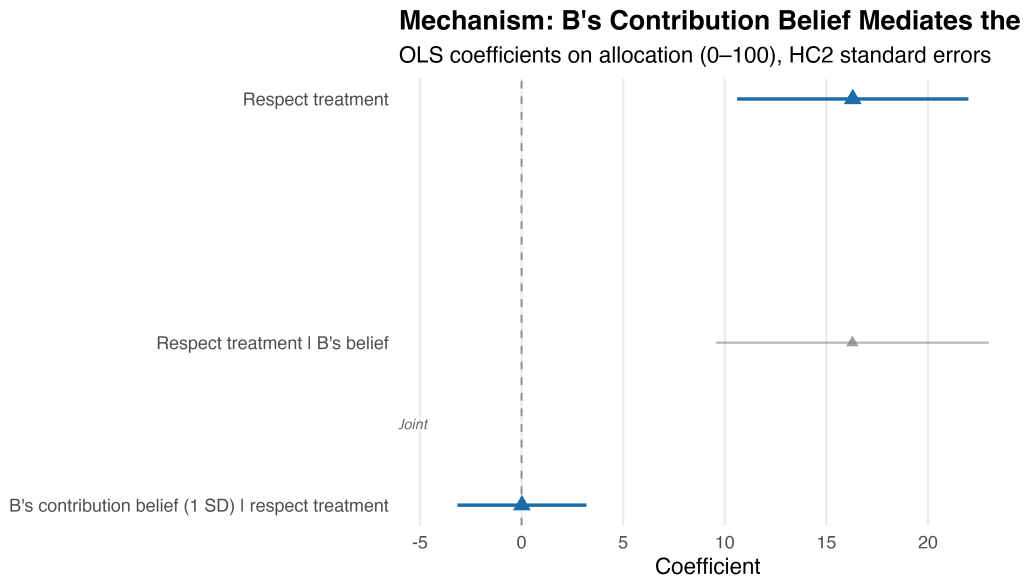


Figure A2: Mediation using B's contribution belief as the mediating variable. The treatment effect on allocation attenuates when controlling for B's contribution belief (z, standardized on control-group SD). Top: bivariate regression. Bottom two: joint regression. HC2 standard errors.

B Voting Evidence

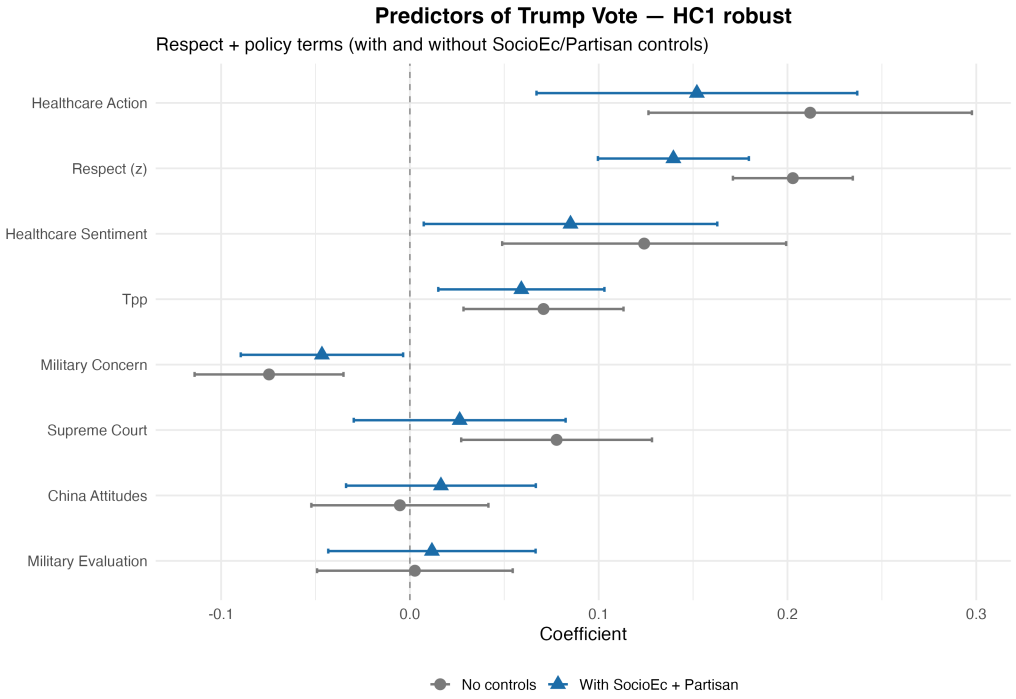


Figure A3: Coefficient on perceived respect difference relative to policy and demographic controls (data split 2). As in Figure 3, perceived respect is among the strongest predictors of vote choice. Horizontal bars are 95% confidence intervals (HC1 standard errors).

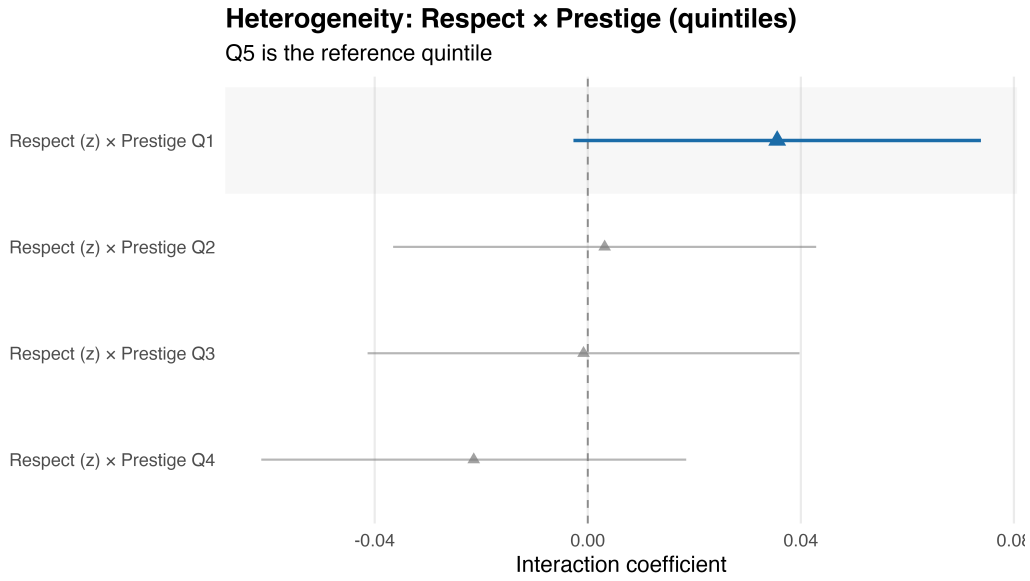


Figure A4: Sensitivity to perceived respect difference by predicted occupational prestige quintile (data split 2; double-selection LASSO, HC3 standard errors). As in Figure 4, voters in the lowest prestige quintile (Q1) exhibit the strongest response to perceived candidate respect.

C Signaling Model

This appendix formalizes the signaling argument in Section 6.1. We adapt the framework to the political communication context: a candidate (sender) B communicates to two voter groups A_1 and A_2 about the respect she holds for them. In the notation of Section 3, B plays the role of the audience whose respect is in question, and each voter group A_g plays the role of the agent whose support depends on perceived respect. We coarsen B's belief about each group's type to $\theta_g \in \{\theta^L, \theta^H\}$; each group's perceived respect ρ_B is then the posterior $\Pr(\theta_g = \theta^H \mid \text{messages})$. The key question is when verbal claims about respect are credible and when candidates have incentives to deviate from conveying universal respect.

Preview of results. Three results regarding a pooling equilibrium in which all opportunistic candidates universally claim to respect everyone emerge. First, a voter's posterior that a candidate genuinely respects her group is strictly higher when that

candidate publicly disrespects a rival group. Thus, bashing one group can be a credible signal of respect toward another. Second, the pooling equilibrium in which all candidates universally claim to respect everyone is sustained only when voters' trust in politicians, π, is sufficiently high; as trust falls, bashing is not dominated for anymore for all candidates. Third, the same unraveling occurs when voters' prior beliefs that politicians genuinely hold them in high regard, θ, are sufficiently low.

Setup. Voters are uncertain about B's type. Types are three-dimensional and independent across dimensions. First, B can be *honest* or *opportunistic*. Whether B is honest depends on her lying costs: a candidate with high lying costs, $\lambda = \lambda^H$, always speaks the truth; a candidate with low lying costs, $\lambda = \lambda^L$, is opportunistic and always says what maximizes her expected vote share. We denote $\Pr(\lambda = \lambda^H) = \pi$ and interpret π as voters' trust in candidates or politicians more generally.

Note that by adding lying costs, we increase the complexity of the model. To keep the analysis tractable and intuitive, we decrease complexity by coarsening other aspects of the setup.

First, we simplify the recognition type space, such that $\theta_g \in \{\theta^L, \theta^H\}$ for all groups $g \in \{1, 2\}$, where θ_g denotes the recognition B has for voter group A_g. Voters have common and commonly known priors about the probabilities of these types, which we denote by $\Pr(\theta_1 = \theta^H) = \theta_1$ and $\Pr(\theta_2 = \theta^H) = \theta_2$.

Second, we assume that the share of voters in group A_g voting for B equals the posterior probability that B holds a high belief about group A_g, $\Pr(\theta_g = \theta^H \mid \text{messages})$. This can be microfounded with voters who have risk-neutral preferences over recognition for their group and an outside option drawn independently from a uniform distribution. Alternatively, one could introduce independent random preference shocks to recover the assumed behavior. Moreover, we follow Callander and Carbajal (2022) in assuming that candidates care about their vote shares.

B chooses public messages $\hat{\theta}_1, \hat{\theta}_2 \in \{\theta^L, \theta^H\}$. Messages are taken literally (e.g., Kartik, Ottaviani, and Squintani 2007), for example, as under a language convention (Braghieri 2023), and we assume that λ^H is sufficiently high so that the honest type always communicates truthfully. Finally, we assume two mutually exclusive groups A_1 and A_2 with population shares p and $1-p$. This comes with some loss of generality, but we argue that restricting attention to two groups is defensible on the grounds that behaviorally, categorizations are often sparse (e.g., us and them) and that logically, one can split any group into a subset and its complement.

Pooling equilibrium. We show that, for a range of parameter values, an equilibrium exists in which opportunistic types send exclusively high messages and that contrasting messages - recognizing one group and bashing the other - are more informative about B's recognition types.

Proposition 1. There exist values for $\theta_1, \theta_2, p, \pi$ such that $\Pr(\theta_1^H \mid \hat{\theta}_1^H, \hat{\theta}_2^L) > \Pr(\theta_1^H \mid \hat{\theta}_1^H, \hat{\theta}_2^H)$.

We propose an equilibrium in which all opportunistic candidates recognize both groups. Thus, the messaging strategy μ for all opportunistic B is $\mu(\lambda^L) = \{\hat{\theta}_1^H, \hat{\theta}_2^H\}$, while all honest candidates communicate truthfully. For this equilibrium to hold, there must be no incentives for unilateral deviations from any opportunistic type. Importantly, in contrast to models of Bayesian persuasion (Kamenica and Gentzkow 2011), candidates cannot commit to a signaling structure ex-ante - meaning before they learn about their types - and thus, messaging strategies need to be optimal ex-post.
The pooling equilibrium payoff for $\{\lambda^H, \theta_1^H, \theta_2^H\}$ and $\{\lambda^L\}$ is:

$$p \Pr(\theta_1^H \mid \hat{\theta}_1^H, \hat{\theta}_2^H) + (1-p) \Pr(\theta_2^H \mid \hat{\theta}_1^H, \hat{\theta}_2^H) = p\frac{\theta_1\theta_2\pi + (1-\pi)\theta_1}{\theta_1\theta_2\pi + 1 - \pi} + (1-p)\frac{\theta_1\theta_2\pi + (1-\pi)\theta_2}{\theta_1\theta_2\pi + 1 - \pi}. \tag{2}$$

First, note that it will never be optimal for any opportunistic type λ^L to deviate to $\{\hat{\theta}_1^L, \hat{\theta}_2^L\}$ because she would only pool with $\{\lambda^H, \theta_1^L, \theta_2^L\}$. Given that $\Pr(\theta_1^H \mid \hat{\theta}_1^L, \hat{\theta}_2^L) \leq \Pr(\theta_1^H \mid \hat{\theta}_1^H, \hat{\theta}_2^H)$ and $\Pr(\theta_2^H \mid \hat{\theta}_1^L, \hat{\theta}_2^L) \leq \Pr(\theta_2^H \mid \hat{\theta}_1^H, \hat{\theta}_2^H)$, her payoff could only decrease.
Next, we consider $\{\lambda^L, \theta_1^H, \theta_2^H\}$'s incentive to deviate. Her payoff from a unilateral deviation to $\{\hat{\theta}_1^H, \hat{\theta}_2^L\}$ is:

$$p \Pr(\theta_1^H \mid \hat{\theta}_1^H, \hat{\theta}_2^L) + (1-p) \Pr(\theta_2^H \mid \hat{\theta}_1^H, \hat{\theta}_2^L) = p + (1-p)\frac{(1-\pi)\theta_1\theta_2}{(1-\pi)\theta_1\theta_2 + \pi\theta_1(1-\theta_2)}. \quad (3)$$

She prefers the pooling equilibrium to deviating to $\{\hat{\theta}_1^H, \hat{\theta}_2^L\}$ iff (2) $\geq$ (3).
$\{\lambda^L, \theta_1^H, \theta_2^H\}$'s payoff from a unilateral deviation to $\{\hat{\theta}_1^L, \hat{\theta}_2^H\}$ is:

$$p \Pr(\theta_1^H \mid \hat{\theta}_1^L, \hat{\theta}_2^H) + (1-p) \Pr(\theta_2^H \mid \hat{\theta}_1^L, \hat{\theta}_2^H) = p + (1-p)\frac{(1-\pi)\theta_1\theta_2}{(1-\pi)\theta_1\theta_2 + \pi\theta_1(1-\theta_2)}. \quad (4)$$

She prefers the pooling equilibrium to deviating to $\{\hat{\theta}_1^L, \hat{\theta}_2^H\}$ iff (2) $\geq$ (4).
Note that $\{\lambda^H, \theta_1^L, \theta_2^L\}$'s unilateral deviations yield lower payoffs to $\{\lambda^H, \theta_1^L, \theta_2^L\}$ than $\{\lambda^H, \theta_1^H, \theta_2^H\}$'s unilateral deviations yield to $\{\lambda^H, \theta_1^H, \theta_2^H\}$. Thus, $\{\lambda^H, \theta_1^L, \theta_2^L\}$'s unilateral deviations do not impose binding constraints on pooling equilibrium parameters. Moreover, it is easy to verify that whenever pooling is optimal for $\{\lambda^H, \theta_1^H, \theta_2^H\}$, it is neither a best response for $\{\lambda^H, \theta_1^H, \theta_2^L\}$ to deviate to $\{\hat{\theta}_1^H, \hat{\theta}_2^L\}$ nor for $\{\lambda^H, \theta_1^L, \theta_2^H\}$ to deviate to $\{\hat{\theta}_1^L, \hat{\theta}_2^H\}$. Therefore, we are left with two remaining cases that may constrain equilibrium parameters. The first case is that $\{\lambda^H, \theta_1^H, \theta_2^L\}$ may unilaterally deviate to $\{\hat{\theta}_1^L, \hat{\theta}_2^H\}$, which would result in the payoff:

$$p\frac{(1-\pi)\theta_1(1-\theta_2)}{(1-\pi)\theta_1(1-\theta_2) + \pi(1-\theta_1)\theta_2} + (1-p)\frac{\pi\theta_2(1-\theta_1)}{\pi\theta_2(1-\theta_1) + (1-\pi)\theta_1(1-\theta_2)}. \quad (5)$$

$\{\lambda^H, \theta_1^H, \theta_2^L\}$ prefers the pooling equilibrium to this deviation iff (2) $\geq$ (5).
The second case is that $\{\lambda^H, \theta_1^L, \theta_2^H\}$ unilaterally deviates to $\{\hat{\theta}_1^H, \hat{\theta}_2^L\}$, which yields the

payoff:

$$p\frac{\pi\theta_1(1-\theta_2)}{\pi\theta_1(1-\theta_2)+(1-\pi)(1-\theta_1)\theta_2}+(1-p)\frac{(1-\pi)(1-\theta_1)\theta_2}{\pi\theta_1(1-\theta_2)+(1-\pi)(1-\theta_1)\theta_2}. \quad (6)$$

$\{\lambda^H, \theta_1^L, \theta_2^H\}$ prefers the pooling equilibrium to this unilateral deviation iff (2) $\geq$ (6). In sum, the pooling equilibrium is obtained for parameters that ensure that $\{\lambda^L, \theta_1^H, \theta_2^H\}$ deviates neither to $\{\hat{\theta}_1^H, \hat{\theta}_2^L\}$ ((2) $\geq$ (3)) nor to $\{\hat{\theta}_1^L, \hat{\theta}_2^H\}$ ((2) $\geq$ (4)), that $\{\lambda^H, \theta_1^H, \theta_2^L\}$ does not deviate to $\{\hat{\theta}_1^L, \hat{\theta}_2^H\}$ ((2) $\geq$ (5)), and that $\{\lambda^H, \theta_1^L, \theta_2^H\}$ does not deviate to $\{\hat{\theta}_1^H, \hat{\theta}_2^L\}$ ((2) $\geq$ (6)). We can verify that there exists a nonempty set of parameter values $\{\pi, \theta_1, \theta_2, p\}$ that satisfies all conditions, and thus, a pooling equilibrium exists. A visualization helps to interpret the main results. To obtain a visualization in three dimensions, we assume $\theta_1 = \theta_2$.

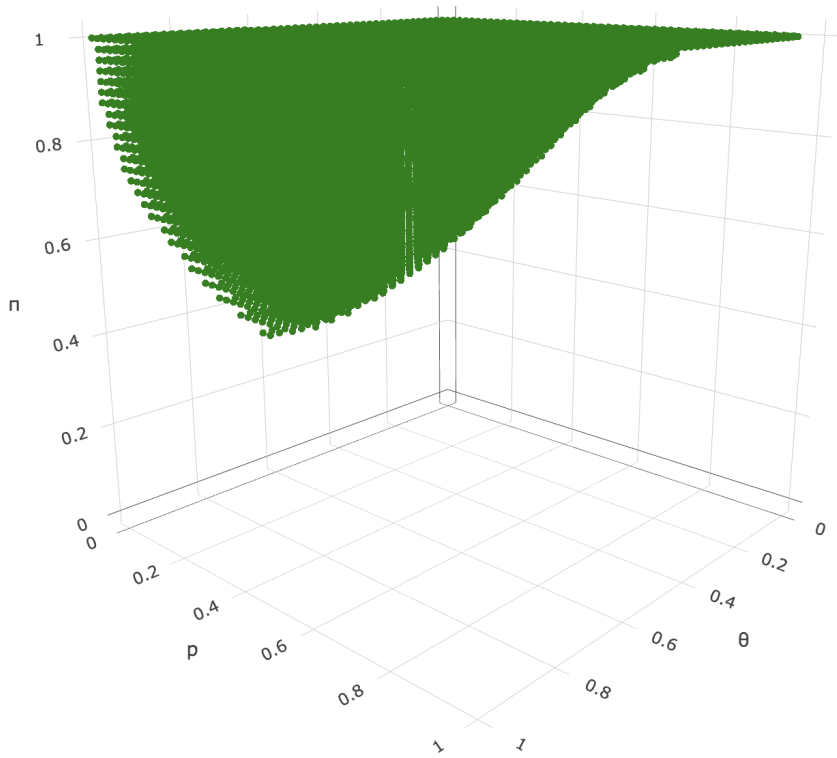


Figure A5: Pooling equilibrium in pure strategies ($\theta_1 = \theta_2 \equiv \theta$). The shaded region indicates parameter combinations (π, θ, p) for which no opportunistic type profitably deviates from universal recognition.

Given our results, we make the following remarks.

Remark 1: A pooling equilibrium exists where opportunistic candidates recognize everyone. In this equilibrium, a candidate B who recognizes A_1 (A_2) but not A_2 (A_1) is more likely, given our parameters certain, to truly recognize A_1 (A_2). Thus, bashing one group can be a credible signal of recognizing another group.

Remark 2: Deviations for opportunistic candidates become "more attractive" when π is small. Thus, recognizing everyone is only a best response for all opportunistic candidates when trust in politicians is sufficiently high. When trust decreases, opportunistic candidates may bash groups of voters.

Remark 3: Deviations for opportunistic candidates become "more attractive" when θs are small. Thus, recognizing everyone is only a best response for all opportunistic candidates when priors about recognition from politicians are sufficiently high. When perceived recognition decreases, opportunistic candidates may bash groups of voters.

D Policy Signaling

This appendix formalizes the three classes of applications discussed in Section 6.2. We specialize the framework of Section 3 to binary types and a linear respect sensitivity function. All three games share a common agent-side structure. Agent A has type $\theta \in \{H, L\}$ with $H > L = 0$, known to A but not to principal P, who holds prior $\mu_0 = \Pr(\theta = H)$. After observing P's action a, A updates to perceived respect $\hat{\mu}(a) = \Pr_A(\theta = H \mid a)$ - the analogue of perceived respect ρ_B with P in the role of B. Agent utility is

$$U_A = m(a, e) + \lambda\, \hat{\mu}(a) \cdot e - C(e), \tag{7}$$

where $m(a, e)$ is material payoff, $\lambda > 0$ is the RBR weight, e is effort, and $C(e)$ is effort cost. The RBR term captures H1: the marginal intrinsic return to effort is strictly

increasing in perceived respect. The three games differ in the nature of P's action and the resulting inference $\hat{\mu}(a)$.

D.1 Gift Exchange

Setup. Prior to interacting with A, P receives a private signal $s \in \{g, b\}$ about θ, with $\Pr(\theta = H \mid s = g) = \mu_g > \frac{1}{2}$ and $\Pr(\theta = H \mid s = b) = \mu_b < \frac{1}{2}$. Agent A knows the signal structure but not its realization. P then chooses whether to make an unconditional gift $a \in \{0, 1\}$ at cost $r > 0$. A observes a, updates $\hat{\mu}(a)$, and chooses $e \in \{0, 1\}$. Payoffs are:

$$U_A = ra + \lambda\, \hat{\mu}(a) \cdot e - Ce, \tag{8}$$

$$U_P = \theta \cdot \Delta_H \cdot e - ra, \tag{9}$$

where $\Delta_H = p(e{=}1, H) - p(e{=}0, H) > 0$ is the marginal value of effort by a high-type agent and $L = 0$ implies effort yields no return under type L.

Prior to interacting with A, P receives a private signal $s \in \{g, b\}$ about θ, with $\Pr(\theta = H \mid s = g) = \mu_g > \frac{1}{2}$ and $\Pr(\theta = H \mid s = b) = \mu_b < \frac{1}{2}$. Agent A knows the signal structure but not its realization. P then chooses whether to make an unconditional gift $a \in \{0, 1\}$ at cost $r > 0$. A observes a, updates $\hat{\mu}(a)$, and chooses $e \in \{0, 1\}$. In this subsection, effort cost is linear: $C(e) = Ce$ with $C > 0$. Payoffs are

$$U_A = ra + \lambda \hat{\mu}(a) e - Ce, \tag{10}$$

$$U_P = \mathbf{1}\{\theta = H\} \Delta_H e - ra, \tag{11}$$

where $\Delta_H > 0$ is the marginal value of effort for a high-type agent, and effort yields no return under type L.

Separating equilibrium. We look for an equilibrium in which $a^* = 1 \iff s = g$. By Bayes' rule, $\hat{\mu}(1) = \mu_g$ and $\hat{\mu}(0) = \mu_b$.

Agent effort conditions ($e^* = 1 \iff a = 1$):

$$\lambda\mu_g > C \qquad \text{(effort optimal when gifted)}, \tag{12}$$

$$\lambda\mu_b < C \qquad \text{(effort not optimal when not gifted)}, \tag{13}$$

which jointly require $\mu_g > C/\lambda > \mu_b$.

Principal gift conditions ($a^* = 1 \iff s = g$):

$$\mu_g\Delta_H > r \qquad \text{(gift profitable after good signal)}, \tag{14}$$

$$\mu_b\Delta_H < r \qquad \text{(gift unprofitable after bad signal)}, \tag{15}$$

which require $\mu_g\Delta_H > r > \mu_b\Delta_H$.

Proposition 2. Under conditions (12)–(15), a separating equilibrium exists in which P gifts iff $s = g$ and A exerts effort iff gifted. The equilibrium requires $\lambda > 0$: without RBR, no such equilibrium exists and the gift has no effect on effort.

Proof. Existence: choose any $r \in (\mu_b\Delta_H,\ \mu_g\Delta_H)$ and $C \in (\lambda\mu_b,\ \lambda\mu_g)$; all four conditions are satisfied simultaneously. Without RBR ($\lambda = 0$): condition (12) becomes $0 > C$, which fails for $C > 0$, so effort is independent of the gift and P has no incentive to gift. □

Köszegi–Netzer robustness. Under intentions-based reciprocity (Rabin 1993), a strategically motivated gift - one made to extract effort - may not trigger reciprocal effort because the principal's intention is self-serving rather than kind (Netzer and Schmutzler 2014; Koszegi 2014). Under RBR this issue does not arise. The gift's informational content - $\hat{\mu}(1) = \mu_g$ - is independent of the principal's motives. A strategic and an altruistic principal who make identical gifts convey identical posteriors, so the agent's effort response is the same in both cases.

D.2 Incentive Schemes

Setup. In this application, agent type captures baseline willingness to exert effort in the absence of extrinsic incentives. Let $\theta \in \{H, L\}$ with $H > L = 0$, privately known to A but not to P, who holds prior $\mu_0 = \Pr(\theta = H)$. Unlike in the gift-exchange application, P benefits from effort regardless of type; type matters here only through A's utility. P designs a compensation scheme before observing effort. Two options: a *fixed wage* w_F (payment independent of output) or *performance pay* (w_0, $w_0 + b$) with bonus $b \geq 0$. The agent infers perceived respect from the scheme:

$$\hat{\mu}^F > \mu_0 > \hat{\mu}^{PfP}, \tag{16}$$

where $\hat{\mu}^F$ is the inference under fixed wage (signals high trust in intrinsic motivation) and $\hat{\mu}^{PfP}$ is the inference under performance pay (signals doubt about intrinsic motivation). A observes the scheme and chooses $e \in \{0, 1\}$. Utility under fixed wage and performance pay is

$$U_A^F = w_F + \theta e + \lambda\, \hat{\mu}^F \cdot e - Ce, \tag{17}$$

$$U_A^{PfP} = w_0 + be + \theta e + \lambda\, \hat{\mu}^{PfP} \cdot e - Ce. \tag{18}$$

Equilibrium effort and optimal bonus. Under performance pay, type θ exerts effort iff

$$b + \theta + \lambda \hat{\mu}^{PfP} \geq C, \tag{19}$$

giving minimum bonus

$$b^*(\theta) = \max\{0,\, C - \theta - \lambda \hat{\mu}^{PfP}\}. \tag{20}$$

Had the scheme not shifted perceived respect (counterfactual with $\hat{\mu}$ fixed at μ_0), the corresponding minimum bonus would be

$$\tilde{b}(\theta) = \max\{0,\, C - \theta - \lambda\mu_0\}. \tag{21}$$

Whenever the interior solution obtains, i.e. when $C > \theta + \lambda\mu_0$, the disrespect signal raises the required bonus by

$$b^*(\theta) - \tilde{b}(\theta) = \lambda(\mu_0 - \hat{\mu}^{PfP}) > 0. \tag{22}$$

Under fixed wage, type θ exerts effort iff

$$\theta + \lambda\hat{\mu}^F \geq C. \tag{23}$$

Since $H > L = 0$, the low type is the binding case for inducing effort from all agents. Thus, fixed wage induces effort from all agents iff $\lambda\hat{\mu}^F \geq C$, while performance pay induces effort from all agents iff $b \geq b^*(L) = \max\{0,\, C - \lambda\hat{\mu}^{PfP}\}$. If both schemes induce effort from all agents and effort yields the same benefit to P regardless of type, then P prefers fixed wage to performance pay whenever $w_F \leq w_0 + b^*(L)$.

A fine for noncompliance is a special case of performance pay. If noncompliance triggers a fine f, then material payoff is $-f(1-e) = -f + fe$, which is equivalent to performance pay with $w_0 = -f$ and $b = f$. Thus the same analysis applies. If fines convey a different respect signal, replace $\hat{\mu}^{PfP}$ with $\hat{\mu}^{fine}$.

Proposition 3. The minimum performance bonus $b^*(\theta)$ weakly exceeds its neutral-respect benchmark $\tilde{b}(\theta)$ for each type $\theta \in \{H, L\}$, and strictly exceeds it whenever the interior solution obtains. Fixed wage induces effort from all agents iff $\lambda\hat{\mu}^F \geq C$, while performance pay induces effort from all agents iff $b \geq b^*(L)$. If both schemes induce effort from all agents and effort yields the same benefit to the principal regardless of

type, then the principal optimally offers a fixed wage over performance pay whenever $w_F \leq w_0 + b^*(L)$.

Proof. From (20) and (21), if $C > \theta + \lambda\mu_0$, then $b^*(\theta) - \tilde{b}(\theta) = \lambda(\mu_0 - \hat{\mu}^{PfP}) > 0$ since $\hat{\mu}^{PfP} < \mu_0$; with the max operator, the comparison is weakly positive in general. Under fixed wage, $e = 1$ is optimal iff $U_A^F(1) - U_A^F(0) = \theta + \lambda\hat{\mu}^F - C \geq 0$, which gives (23). Since $H > L = 0$, any scheme that induces effort from type L also induces effort from type H, so type L is the binding case for inducing effort from all agents. Finally, if both schemes induce effort from all agents and effort has the same value for P regardless of type, then P compares only implementation costs: fixed wage costs w_F, while performance pay costs $w_0 + b^*(L)$ when set minimally to induce universal effort. Hence fixed wage is preferred whenever $w_F \leq w_0 + b^*(L)$. □

D.3 Control and Delegation

Setup. P chooses control level $k \in [0, 1]$: $k = 0$ is full delegation; $k = 1$ is full monitoring. Control imposes two effects on A:

1. *Direct restriction*: agent's feasible effort is capped at $\bar{e}(k)$, strictly decreasing in k with $\bar{e}(0) = e_{\max}$.

2. *RBR channel*: $\hat{\mu}(k)$ strictly decreasing in k; more control signals a lower type prior.

Agent utility is $U_A = B(e) + \lambda\hat{\mu}(k)\, e - C(e)$ with $B' > 0$, $C' > 0$, $C'' > 0$.

Equilibrium effort. When the constraint does not bind, the interior solution satisfies $B'(e) + \lambda\hat{\mu}(k) = C'(e)$. Differentiating implicitly:

$$\frac{de^*}{dk} = \frac{\lambda\, \hat{\mu}'(k)}{C''(e^*) - B''(e^*)} < 0, \tag{24}$$

since $\hat{\mu}'(k) < 0$ and $C'' - B'' > 0$. When the constraint binds, $e^* = \bar{e}(k)$ which also decreases in k. In both cases equilibrium effort is strictly decreasing in control.

Proposition 4. Equilibrium effort $e^*(k)$ is strictly decreasing in k through the RBR channel (24) and, when binding, the direct restriction. If the principal values effort at rate $v > 0$ and monitoring yields enforcement benefit $E(k)$ with $E(0) = 0$, there exists $\lambda^{**} > 0$ such that for $\lambda > \lambda^{**}$ the principal optimally sets $k^* = 0$ even when monitoring is costless: the RBR-induced effort reduction outweighs the enforcement benefit.

Proof. $de^*/dk < 0$ follows from (24) and the constraint argument. Principal's net payoff from increasing k marginally from 0 is $v\,(de^*/dk)|_{k=0} + E'(0)$. Since $de^*/dk = \lambda\hat{\mu}'(0)/[C''(e^*(0)) - B''(e^*(0))] < 0$ and is linear in λ, there exists λ^{**} such that $v|de^*/dk|_{k=0} > E'(0)$ for all $\lambda > \lambda^{**}$, making $k^* = 0$ optimal. □

Connection to trust games and sanctions. The trust game (Berg, Dickhaut, and McCabe 1995) maps naturally to this framework: the first mover's transfer t is a delegation decision, with $k \propto (T - t)/T$. Larger initial transfers signal higher perceived respect, inducing larger back-transfers through RBR. The sanctioning experiment of Fehr and Rockenbach (2003) - in which trustees send less when an enforceable sanction is available, even when it is not invoked - is consistent with the control channel: the availability of the sanction reveals $k > 0$, lowering $\hat{\mu}$ and reducing prosocial effort regardless of whether the threat is carried out.