Identity and the Social Construction of Reputation in World Politics

Jonathan Renshon * Ryan Powers†

March 29, 2026

Abstract

Scholarship on reputations in IR has left out a striking feature of human psychology: identity. Categorizing others as "us" or "them" is automatic, pervasive and has significant implications for reputations. We provide a framework—based on social identity theory—to explain how ingroup bias affects how reputations are generated and maintained. Empirically, we provide two contributions. First, we field descriptive surveys on public elite samples to learn about the markers of international ingroups and aid in our experimental design. Second, we pilot and field a pre-registered experiment that tests implications of our identity-based theory of reputations, finding that ingroup membership improves (1) reputations of all types (signaling, financial, humanitarian and resolve) and (2) observers' willingness to cooperate but (3) does not moderate the impact of past resolved behavior on reputations for standing firm in the security domain. We use several pre-registered follow-up studies to refine theoretical scope conditions and explore treatment generalizability.

*Board of Visitors Professor of Political Science, Department of Political Science, University of Wisconsin-Madison. ✉: renshon@wisc.edu. 🖱: http://jonathanrenshon.com.

†Associate Professor, Department of International Affairs and Gary K. Bertsch Director of the Center for International Trade and Security, School of Public and International Affairs, University of Georgia. ✉: ryan.powers@uga.edu 🖱: https://ryanpowers.net.

The study of reputations—judgments about persistent type, tendencies or characteristics—dates back to at least Schelling's (1966, 124) admonition that a country's "image" or "reputation for action" is "one of the few things worth fighting over." Debates about reputation have crossed the divide between security and IPE research (Gray and Hicks, 2014; Kertzer, 2016) as well as rationalist and psychological approaches (Mercer, 1997, 2010), though significant open questions remain.

One important outstanding puzzle in this literature concerns variation in states' reputations across different observers. Why, for example, did Serbia's obvious and horrific war crimes in Kosovo lead to some (Western) states updating beliefs about Serbia's "type" while Russian observers' beliefs remained unaffected? Why did Western observers' beliefs about Japanese resolve and character differ so starkly from China's in the lead-up to the 1905 Russo-Japanese war? A related puzzle from the perspective of the reputation literature is how beliefs about "type" change over time in the *absence* of new information. One prominent recent example of this is how U.S. attitudes shifted in 2025 drastically against its traditional allies, the EU and Canada, with no behavioral cause to blame for the observed disruption.[1] Recent reviews of reputation (such as Jervis, Yarhi-Milo and Casler 2021, 185-86) sum up these issues by noting that beliefs about a given state's reputation might be "heterogeneous... across states," without necessarily answering the critical question of *why* this might be so.

Drawing from the literature on social categorization, we argue that *identity*—in particular, the categorization of target states into either the ingroup or outgroup—helps to explains variation in reputations across observers and over time. Identity is "crucially important for understanding contemporary life" (Abdelal et al., 2006, 695) and "among the most normatively significant and behaviorally consequential aspects of politics" (Smith, 2004, 304). If reputations are "in the eye of the beholder" (Jervis, 1989, 514), then our argument is that the lens of ingroup identity provides a powerful filter through which actors view events and update their beliefs about others. Cast in this light, existing explanations of seemingly idiosyncratic or biased beliefs—e.g., what Mendeloff (2008, 33) called the "myth of Slavic Brotherhood" in the context of Russia/Serbia or the racist distortions cataloged by Mercer (2023)—take on new theoretical coherence.

We make both conceptual and empirical contributions in this brief paper. First, we use social identity theory to provide traction on how reputations operate in IR. Our argument is that how actors categorize other states—as either ingroup or outgroup—holds two important implications for reputational beliefs and actions towards that state. First, ingroup categorization affects *reputation*, leading to ingroup members getting an artificial boost from observers in their judgments of type and likely behavior. This improves upon current theories by providing a theoretical basis for judgments about reputation in the absence of any behavior (or in the presence of unclear patterns of behavior). Second, we argue that ingroup categorization has additional payoffs, as it changes not only how actors perceive them, but the willingness of other actors to *cooperate* with and *defend* them.

[1] http://www.economist.com/graphic-detail/2025/03/31/schooled-by-trump-americans-are-learning-to-dislike-their-allies. Accessed May 21, 2025.

Empirically, we make two contributions. First, we field a descriptive survey on identity in international politics among the general public and two different samples of elites. Our survey validates a critical assumption of our argument—that citizens and elites do not view all other countries as an undifferentiated outgroup—by finding variation (and consistency across groups) in which countries are seen as members of the U.S.' ingroup. We learn that regime type, culture and language are important markers of ingroup identity; that both elites and the public (Pearson's $r = 0.9$) hold similar reputational beliefs about other states and that there is a strong relationship between the belief a state is in one's ingroup and its ranking on valued attributes.

Our highly-powered, pre-registered experiment builds on our elite survey and to test implications of our theory by "nudging" a real country—Turkey—from outgroup to ingroup. Notably, we expand the scope of reputations from the traditional focus on resolve, eliciting beliefs about financial, humanitarian and signaling reputations. We find broad and strong support for the notion that ingroup categorization affects states' reputations across many substantive domains, though at least in the domain of security, ingroup members do not get a pass if they act in an irresolute manner. Ingroup categorization also has significant benefits for observers' willingness to cooperate with the target state (e.g., trading with and even defending the country in wartime). We also find that the magnitude of ingroup effects in our study is four times the size of parallel outgroup effects: it is "ingroup love" and not outgroup hate, in other words, that matters most. A series of pre-registered follow-up studies establishes the generalizability of the treatment over time and across survey platforms (i.e., samples) while demonstrating that its effects are not caused by priming, bundling of information related to regime type or the use of images.

Our work dovetails with a number of recent works highlighting the importance of identity attachments in IR (Powers, 2022) and on the impact of race—a particularly salient individual-level marker of identity—on world politics (Freeman, Kim and Lake, 2022; Mercer, 2023; Tomz and Weeks, 2025). This also builds upon recent empirical work on ethnic (ingroup) bias in responses to violence (Lyall, Blair and Imai, 2013) and other work showing that, in the domain of trade attitudes, "prior beliefs... about a given object... shape a host of other attitudes regarding that object" (Carnegie and Gaikwad, 2022, 169). Finally, our argument helps to explain recent patterns of shifts in public opinion, such as the link between the level of desired helping behavior towards Ukraine (following Russia's invasion) to increased levels of ingroup categorization among Europeans (Sinclair, Granberg and Nilsson, 2023).

1 Identity and Reputations

1.1 Cracks in the "Reputations As Behavior" Story

Over the past few decades, a general consensus has formed around the core meaning of reputation: reputations are first-order beliefs about the persistent character, attributes or behavioral tendencies of an actor that are related to their past behavior (Dafoe, Renshon and Huth, 2014; Dafoe and Caughey, 2016; Renshon, Dafoe and Huth, 2018). They are "first-order" in that they are one actor's

beliefs about another actor: C can have one reputation in the eyes of A and another reputation in the eyes of B (Dafoe, Renshon and Huth, 2014, 374). Past behavior either "informs" reputations (Baser, 2024) or (in a stronger version) reputations are "based on" past actions (Dafoe, Renshon and Huth, 2014, 374). The relationship between past behavior and reputations is complicated by slippage between theoretical constructs and research designs: in quantitative designs, e.g., reputations are measured as a pure function of past behavior even if the theoretical concept is broader (e.g., Clare and Danilovic 2012).

Whether reputations are believed by scholars to be purely a function of past behavior or whether that is a convenient simplifying assumption, it's clear that there is more to the story. Specifying the primacy of behavior only invites questions about *which* behaviors matter: whether the behavior is"in context" (Tomz, 2007; Weisiger and Yarhi-Milo, 2015), early interactions (Lupton, 2020) or membership in particular organizations (Gray, 2013). Empirically, there are significant gaps as well between reputational beliefs a reasonable observer *ought* hold based only on behavior, and what we see instead. For example, Mercer (2023) highlights the systematic racism that "distorted" assessments of Japan's "capabilities, interest and resolve" on the eve of the Russo-Japanese War in 1905, More systematically, Renshon, Yarhi-Milo and Kertzer (2023, 3) show that a state's reputation for resolve and other attributes is heavily influenced by their regime type, while Brutger and Kertzer (2018, 707) suggest a "dispositional theory of reputation in which individuals draw on their pre-existing foreign policy orientations to make reputational judgments" (see also Yarhi-Milo, Kertzer and Renshon 2018).

1.2 Social Identity Groups and Reputations

If it is "unlikely that how states act is the sole input into their reputations" (Renshon, Yarhi-Milo and Kertzer, 2020, 3), our argument is that identity—in particular the categorization of actors into ingroups and outgroups—is an important component of reputational beliefs. We argue that identity dynamics—as reflected in social identity theory—affect both how we perceive other actors' "type" (are they a resolved type or unresolved?) and estimates of their likely behavior.

Our argument builds upon two strands of literature, the first on social identity theory and the second on the importance of group membership and other identity factors for reputations. While there are a "dizzying array" (Brewer, 2001, 115) of definitions of social identity theory,[2] a throughline in all the "inventions and re-inventions" of SIT is the idea that "an individual's self-concept is derived... from the social relationships and social groups he or she participates in" (Brewer, 2001, 117). Group membership provides people with a sense of where they stand with respect to others, a practical guide for how to act towards others and, finally, a lens through which to understand the behavior of others. Categorization into "us" and "them" is a "cognitive necessity" since one cannot begin to figure out how to act or what to think until they've "simplified the whir and buzz of our social environment" (Mercer, 1995, 241). SIT has been particularly influential in the study of status, (e.g., Wohlforth, 2009; Murray, 2018; Dobrescu, 2023) but has also seen use in theories of

[2]On SIT and and self-categorization theory, see Huddy (2001, 132).

nation-building (Sambanis, Skaperdas and Wohlforth, 2015), EU expansion (Curley, 2009), threat perception (Gries, 2005; Rousseau and Garcia-Retamero, 2007), trade preferences (Mutz and Kim, 2017), as well as broader theories of anarchy and self-help (Mercer, 1995) and rationality (Mercer, 2005) in IR.[3]

Three elements of SIT make it a good candidate for helping to explain reputations in IR. First is its focus on group—as opposed to interpersonal—contexts in which behavior is determined by "category-based processes" (Brown, 2000, 746); a natural fit for politics, in which groups and group-based processes are critically important (Kertzer et al., 2022). In fact, early research on multiple identities suggested that we might expect predictions from social identity theory to be *most* applicable to (among others) political identities because they are more 'collective in nature' than other individual aspects of identity" (Deaux et al., 1995, 286). Second are the conditions under which *social* identity is likely to matter most: when outgroups are seen as "competing for resources" and "when groups view the outgroup as having a history of tense relations," two conditions that are ever-present in international affairs (Islam, 2014, 1782). Finally, seeing reputational inferences through the lens of identity places the theoretical focus on the actors making the reputational inferences, who have "very human personal, psychological and political needs" that influence their beliefs (Jervis, Yarhi-Milo and Casler, 2021, 189).

We also build on a related literature more closely related to reputation that focuses on markers of identity and categorization. Kertzer et al. (2019), for example, finds that observers attribute resolve in part based on markers of identity and categorization, such as whether the target of inference is an ally or not, and whether the regime type is similar to their own. Gray and Hicks (2014, 326-7) find that the "clubs" (IOs) states join affect their reputation; in their words, "*who* a country signs with can be more important than *what* they sign." More broadly, there is evidence from other corners of IR that identity-related attributes matter a great deal in how individuals feel about and act towards other countries, with cultural cues "outperforming" regime type in predicting perception of threat/trust (Lacina and Lee, 2013, 143; see also Escribà-Folch, Muradova and Rodon 2021, Hayes 2012, Carnegie and Gaikwad 2022, Chu and Lee 2024). Oren (1995, 147) provides the strongest statement in this vein, arguing that "the democratic peace claim is not about democracies *per se* as much as it is about countries that are 'America-like' or of 'our kind.'"

While the "sociological turn" (Kowert and Legro, 1996; see also Katzenstein, 1996; Checkel, 1999) decades ago helped to introduce identity to IR scholars, and recent work has taken up exhortations (e.g., Tetlock and Goldgeier, 2000) to explore identity in IR, three clear challenges emerge from a close reading of the literature.

First, literature on identity in IR has made tremendous progress in applying the insights from sociological and psychological theories to world politics, but have focused almost entirely on how actors see *themselves*, ignoring how group categorization of others affects behavior and perceptions. While there is no shortage of work in IR on social identity theory, the vast majority use it to fuel theories focused on actors' attachment to their own country (or to a "supranational" identity;

[3]Examples abound from American and Comparative politics as well (e.g., West and Iyengar, 2022).

see Powers, 2022) or the achievement of a positive social identity at a state level (Larson and Shevchenko, 2019).[4] Used in this manner, SIT has been critical to literatures on nationalism and national identity (Herrmann, 2017, Chung, 2023, Kalin and Sambanis, 2018, 248). These works take for granted the categorization processes we're interested in here to focus on the effects of differing levels of attachment to particular groups, answering quite different causal questions about identity and its effects on behavior and perception.

To the extent that IR has not focused as much on categorization—why, i.e., SIT might not offer a "ready-to-wear theory of international politics" (Hymans, 2002, 6)—it is because generating a theory of identity and reputation in international relations requires answering, "how do people decide what other countries are members of their ingroup? " In most psychological work on SIT, identities are either created on the spot based on arbitrary distinctions—the famous "minimal group paradigm" (Otten, 2016)—or on categories such as gender and race (Cikara and Van Bavel, 2014, 247). In one early review, Huddy (2001, 128, emphasis added) argued that we knew little about this issue in IR due to scholars' "disinclination to examine the *sources of social identity* in a real world complicated by history and culture."

Finally, we note an empirical challenge. Extant work has provided the sense that reputations correlate with identity-related markers, but no direct causal evidence that it's categorization specifically that matters and not other, related factors. Some of this has to do with the ever-present debate between observational and experimental designs, but even in experimental designs, it's typically not categorization or identity that is being carefully manipulated.

1.3 Empirical Implications of our Identity-Based Theory

Boundaries of Ingroups in World Politics In conceptualizing international identity, we take our cue from Abdelal et al. (2006, 696) in focusing on "relational comparisons" with other groups; in essence, "defining an identity group by what it is not, i.e., the way it views other identity groups, especially where those views about the other are a defining part of the identity." Thus far, most SIT-based work in IR has assumed that *any* "other" outside of one's own nation represents an outgroup—that there is "a black-white dichotomy between 'us' and 'everybody else'" (Hymans, 2002, 10). In contrast, we argue that states will vary in the extent to which they are categorized as outgroups or ingroups by actors, motivating our exploratory research question (listed alongside our hypotheses in Table 1): *what are the relevant outgroups in international politics?* Or as Rousseau (2006, 5) asked, where do people draw the line between "us" and "them?"

How Categorization Shapes Reputations While the boundaries of group identity in world politics are yet unknown, the implications of that group identity are more clear. Primary among them is ingroup bias (Brown, 2000, 747): individuals have a deep-seated need to see themselves—and the groups to which belong—in a positive light, leading them to evaluate their own group more favorably than other groups (Cuhadar and Dayton, 2011, 274). Ingroups are "perceived as

[4]For an exception, see Rousseau (2006).

	Exploratory Research Question: How do countries get grouped into ingroups and outgroups in world politics?
Reputations	**H1 (Ingroup Bias at Baseline)**: Categorization of a target state as an ingroup member will positively influence evaluations of target's "type" and attributes (i.e., their reputation) **H2 (The Moderating Influence of Identity on the Effect of Observed Behavior)**: Categorization of a target state as an ingroup member will moderate the impact of observed behavior on reputations
Behavior	**H3 (The Consequences of Identity for Cooperation)**: Categorization of a target state as an ingroup member will increase respondents' willingness to engage with and help the target state.

Table 1: Research Questions and Hypotheses

better, friendlier, more competent, and stronger than other groups" (Druckman, 1994, 48). And, critically, the "act of categorization" has a significant effect on not just perceptions of others, but our behavior (Cikara and Van Bavel, 2014, 247).

Our *theoretical* argument is that categorization as an ingroup member will positively affect reputations. Empirically, we operationalize this in two ways. First, we predict that *categorization of a target country as an ingroup member will positively influence observers' beliefs about that target's "type," attributes and likely behavior; i.e., their reputation (H1).* Observers—whether they are elite officials or members of the general public—will see ingroup members as more moral and honest, more capable, more likely to act in a resolved manner, and to honor their commitments compared to outgroup members.

We operationalize our argument in a second way by focusing specifically on how categorization affects the interpretation of new information about past behavior. Our argument suggests that, in some sense, identity comes first: judgments about reputation are downstream of the categorization that we argue is so important. Empirically, this suggests that *behavior witnessed by observers will be judged differently depending on whether the target of inference is considered to be in the observers' ingroup or not.*

While H1 & H2 investigate the consequences of identity categorization on reputational judgments, H3 investigates the consequences of that bias for observers' *behavior* towards that target. Here, a large body of research in cognate fields suggests that shared ingroup identity directly affects cooperative behavior (e.g., Levine et al., 2005) on an interpersonal level and in more diffuse ways, such as the willingness to consume goods from culturally distinct countries (e.g., Bankert, Powers and Sheagley, 2022). We predict that *observers will be more willing to cooperate and defend states perceived to be ingroup members.*

2 Research Design

The discussion above implies two priorities of any empirical investigation of identity in international politics, one exploratory and one causal. The first is investigating the *boundaries* of international group identification: who do elites and the public see as being in their ingroup, internationally? The second is to link identity—particularly perceptions of other actors as being in one's ingroup or outgroup—with reputation-related beliefs and behaviors.

Broad theoretical arguments require translation into specific estimands in order for a theory to be properly tested in any empirical design. To do so, we utilize the framework of "theoretical" vs "empirical" estimands suggested by Lundberg, Johnson and Stewart (2021). Theoretical estimands are a more formalized and detailed version of the hypotheses depicted in Table 1, while empirical estimands take into account real world constraints and focus on observed quantities. Explicitly stating these quantities provides clarity about what research design is optimal, what sources of data ought to be used and what assumptions are necessary to connect our theoretical to our empirical estimands. Row 2 of Table 3 in Appendix J.6, e.g., shows that while our theoretical estimand is the difference in reputation for countries in a respondents' ingroup/outgroup, our empirical estimand is the difference in mean reputation score in our experiment between the same country "nudged" towards either the ingroup or outgroup.

Explicitly detailing our theoretical estimands also helps connect the target of our investigation to an outcome through a particular method and set of identifying assumptions. For example, given that our first question implicates a descriptive quantity—perceptions and beliefs about the U.S. in relationship to other countries—we turn to survey methods fielded in a U.S. sample of elites and citizens. Our second set of questions are causal, motivating the turn to survey experiments that build upon the lessons of our descriptive work.

We focus in our empirics on a wide range of outcomes. Our pre-registered DV is an average of all reputational beliefs, including traits (e.g., tough, trustworthy, morally pure) and behavioral tendencies (e.g., follow through, repay debts). Our second DV measures willingness to cooperate, composed of an average of everything from willingness to interact socially with citizens of the target state to decisions related to financial aid and even willingness to fight alongside the target state.

Alternatives To Our Group-Based Theory of Reputation

Countervailing hypotheses come from alternative theories of both reputations and identity. Informational accounts of reputation, for example, argue that observers "try to diminish their uncertainty by learning from the state's past behavior" (Baser, 2024, 2). If reputations are based *only* on past behavior (as suggested by numerous accounts, e.g., Casler and Yarhi-Milo 2023), then the *identity* of the target state is irrelevant. If that is the case, we would expect null results across H1 and H2. This is observationally equivalent with another alternative in which reputations *are* determined by factors besides past behavior, but not identity.

A second set of challenges comes from alternative conceptions of identity. One foundational

challenge for our hypotheses is in simply demonstrating that ingroup/outgroup distinctions matter at all in an international context outside of the obvious attachments to one's own country. Other challenges come from extant theories. For example, Mercer (1997) argues that observers interpret behavior as due to dispositional factors when it involves an "outgroup's undesirable behavior" (and due to situational factors when it involves desirable behavior; p. 9). Because the theory views all other countries as outgroups and only dispositional attributions can generate or update reputations, "adversaries can get reputations for having resolve... while allies can get reputations for lacking resolve," but not vice-versa (p. 10). This theory provides two clear counter-predictions: first, that all other countries will be perceived as "outgroups"; second, Mercer predicts that desired behavior by allies is ignored, while in our theory, resolved behavior by a target state—even ingroup members—will positively affect observers' estimates of their reputation. Our theory also differs in that "desired" behavior is not exogenous to identity; how a behavior is understood (as good or bad) and the importance placed on it is—for our theory—endogenous to (and causally *post*) identity categorization.

Image theory also represents a useful contrast. Herrmann et al. (1997, 407) argues that judgments about capabilities, threat/opportunity and cultural superiority are critical (see also Boulding, 1959).[5] Conceptually, one important difference is that we argue that the most important judgment is whether actors are "part of our group" (ingroup or not), and other judgments (for example, about how threatening they are or their capabilities) are downstream of that.[6] Empirically, our theory can be distinguished from image theory by the element of cultural superiority: if in our experiment, switching a state's categorization from ingroup to outgroup was equivalent to moving perceptions of a country from one image (ally) to another (enemy), then views of the culture of that country would remain consistent across treatment groups.

3 The Boundaries of Ingroups in World Politics

Our goal in our descriptive surveys fielded on the public and foreign policy elites is to identify—for these two important populations—who is in the ingroup, who is in the outgroup, and what marks them as such. In addition to answering a substantively important question about "rich and meaningful (rather than minimal) group attachments" (Weiss, 2025, 40), we use insights from these initial descriptive surveys to motivate and help us design our causal experiment.

We fielded our two descriptive surveys in Fall 2022.[7] Our U.S. general population sample ($N \approx 1,500$) was recruited by LucidTheorem, who used census quotas on region, gender, race and age to ensure that our sample included a broad cross-section of the public.[8] Our U.S.-based foreign

[5]Each category can take on one of three values—superior, similar or inferior—leading to 27 possible relationships, though "in practice, the framework focuses upon the five combinations that appear particularly frequently" (degenerate, colony, ally, imperialist, enemy; Kertzer, 2023, 11).

[6]There is not an obvious contrast conceptually with image theory or the stereotype content model (SCM) which functions as its theoretical engine. This is because SCM is focused on explaining variation *within* outgroups—through the dimensions of warmth/competence—rather than between ingroups and outgroups.

[7]See Appendix §C.2.

[8]See Appendix C.1 for comparison to a recent national benchmark survey.

policy elite sample ($N = 215$) was recruited by the Teaching, Research, and International Policy (TRIP) Project (Avey et al., 2023) and follows the procedures outlined in Avey et al. (2021). It consists of individuals who work(ed) in offices in the White House, Defense Department, State Department, Treasury Department, US Agency for International Development, among others over the period 2000-2020. We present results from the policy elite sample below as they have both domain-relevant expertise and experience in institutional roles that gave them influence over policy (Kertzer and Renshon, 2022), but results hold when we include a larger academic faculty sample as well.[9]

In our surveys, we presented respondents a list of 10 countries[10] and asked them to categorize the countries as either 'us' or 'them.'[11] This exercise allows us allows us to identify precisely *who* is in the ingroup and outgroup for each population. Moreover, the shared attributes of the members of the resulting groups will provide important clues as to how individuals distinguish ingroup from outgroup members. Following our group categorization task, we asked respondents to indicate their level of agreement with six statements about three countries (chosen at random from the ten the respondent had initially categorized). Among several other dimensions, we asked about agreement with statements that the country in question is trustworthy, likely to stand firm against enemies, likely to honor commitments, likely to pay its debts, and similar to the United States.

3.1 Three Lessons on International Identity

We take three principal lessons from our descriptive surveys. First, **countries categorized as ingroup members tend share attributes in common (e.g., geography, culture, religion, language, and/or regime type)**. Figure 1 plots the proportion of the public and foreign policy elites who categorized each country as being in the "us" group and plots these quantities against each other—elites on the vertical axis and the public on the horizontal axis. From this, it's clear that *similarity* to the United States on social/political dimensions (e.g., geography, culture, ethnicity, religion, regime type, and/or alliance status) is critical to ingroup categorization at the international level. Indeed, the six countries with the highest levels of ingroup-ness (Canada, UK, Italy, France, Germany, and Spain) are all relatively wealthy democracies with predominantly white and Christian populations and are NATO members. When asked to choose elements that were critical to being in their ingroup, the top three categories were "free" (19.3%), "democratic" (12.8%) and "english speaking" (11.5%).

The second lesson we take from our data is that **elites and the public hold similar reputational beliefs about countries around the world**. One aspect of this the high degree of convergence (Pearson's $r = 0.9$) between the two samples' beliefs about which countries are in their ingroup. Similar findings emerge with respect to traits and behaviors: Figure 2 plots the mean level of agreement that a country is trustworthy, likely to honor commitments, likely to repay debts, and

[9] On TRIP academic sample ($N = 429$), see See Appendix §C.3.

[10] Randomly drawn from list of 25 most powerful countries (from COW) plus Israel.

[11] Our approach is adapted from Pinto et al. (2020), who study the role of identity in consumer purchasing decisions.

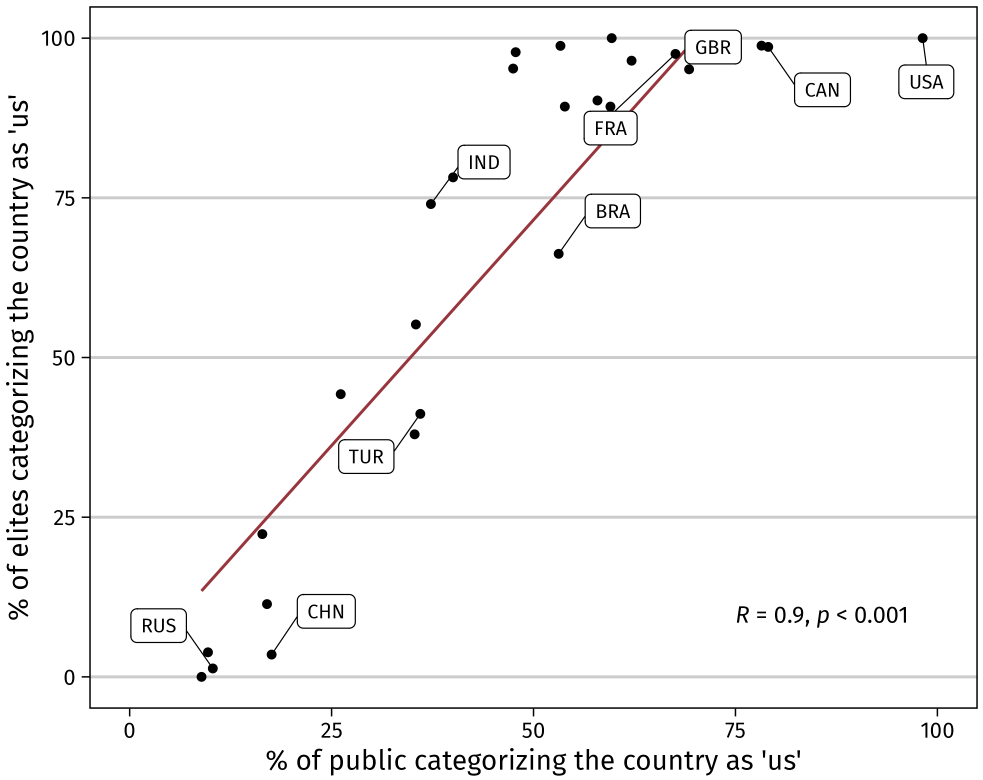


Figure 1: **Survey: Elites and the public agree on who is in the ingroup.** We asked respondents to categorize countries as either "us" or "them." The scatter plot shows that countries that elites were most likely to categorize as "us" were also categorized as "us" by the public.

likely to stand firm in a crisis (for each sample). Elites and the public not only agree on who is in the ingroup and what defines them as such; they also hold similar reputational beliefs about those actors.[12]

Finally, there is a **strong relationship between ingroup-ness on the one hand and valued attributes and traits on the other**, among both the public and elites. Figure 3 plots the correlation between the % of the sample categorizing a country as being in the ingroup and ratings on that particular trait/behavioral tendency (panels ordered from highest to lowest correlation). Doing so reveals reveals that the relationship between ingroupness and positive reputational beliefs varies across dimensions in similar ways among the two samples. For both elites and the public, trustworthiness appears linked strongly directly to ingroupness but this link is not quite as pronounced when it comes to perceptions of resolve.

3.2 Informing our causal research design

Our descriptive work above provides a valuable foundation on which to build our experimental design. First, the correlation between ingroupness and valued attributes—as well as our finding

[12] Agreement ranges from an r of 0.61 for "stand firm" to between 0.8-0.9 for "'trustworthy," "honor commitments," and "repay debts."

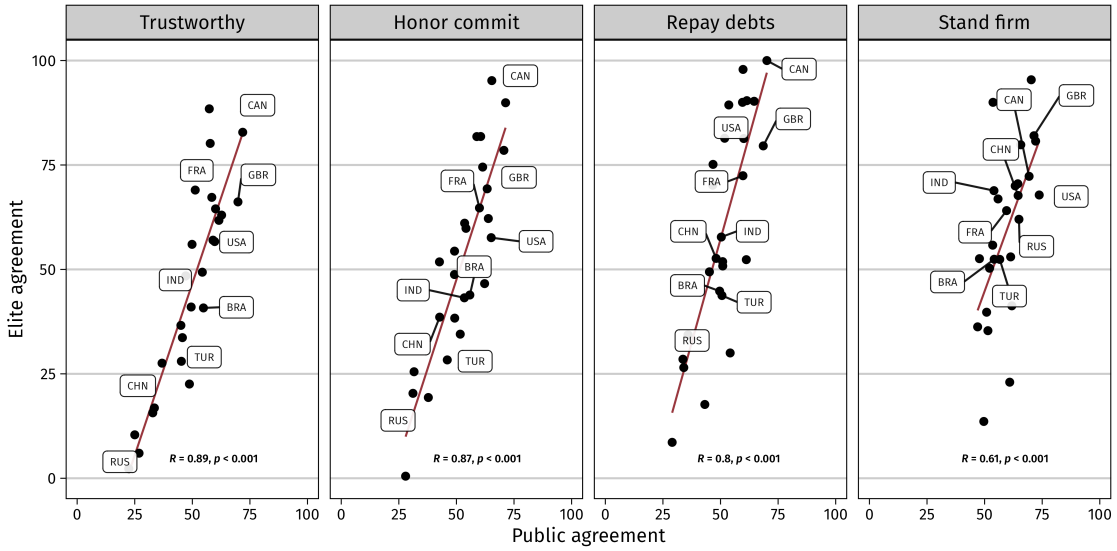


Figure 2: **Survey: Elites and the public hold similar reputational beliefs.** We asked respondents to indicate their level of agreement with a variety of statements about different countries.

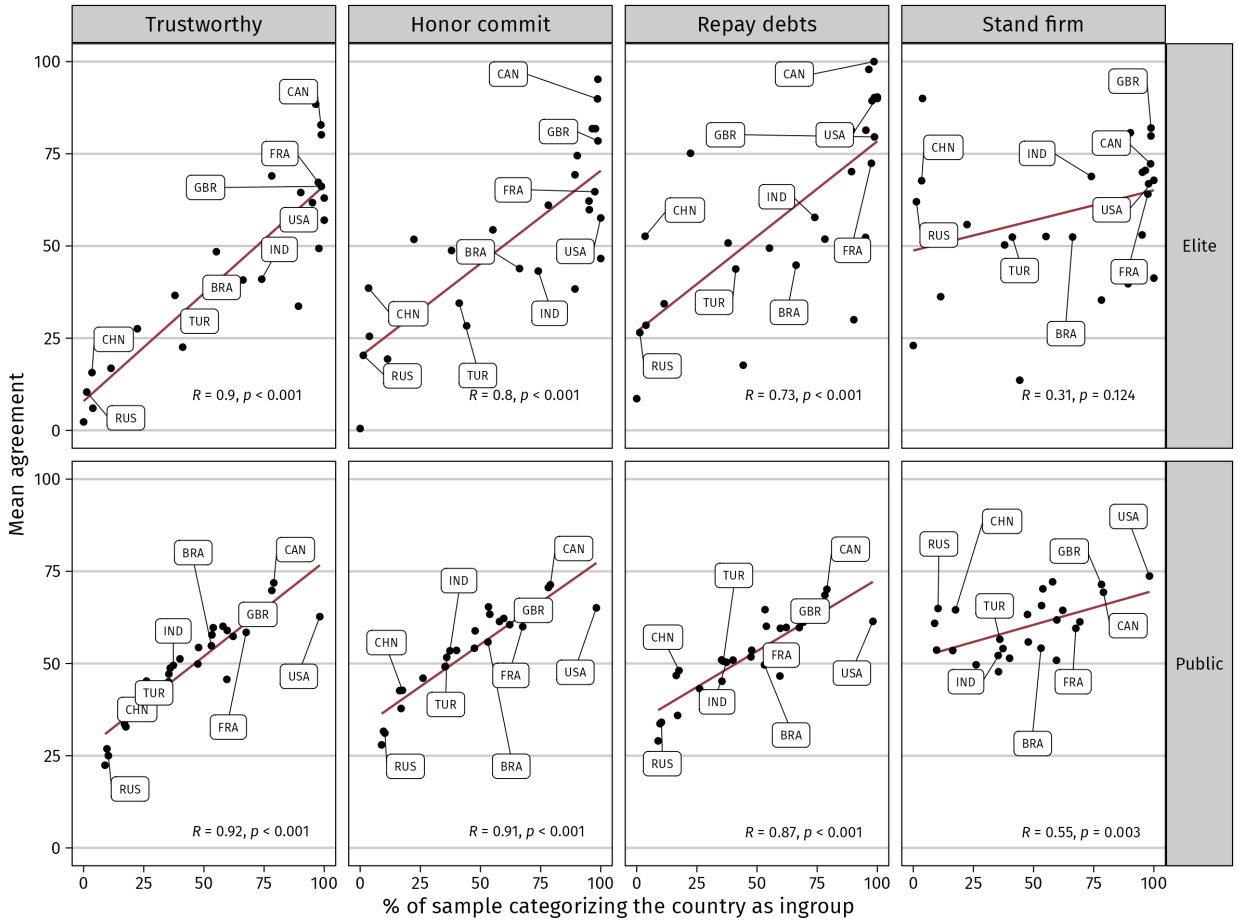


Figure 3: **Survey: Positive traits are correlated with Ingroup-ness.**.

that ingroup members get a reputational subsidy—provides empirical motivation to investigate the possible causal relationship more directly using experimental methods. Second, the survey aids in construction of our treatment by suggesting that shared political, ethnic, religious, and/or cultural features are important markers of group identity. Finally, our survey results illuminated not just who the U.S. public and elites see as in their "international ingroup," but which states find themselves in a liminal space between ingroup and outgroup. Turkey, for example, was consistently in the middle third of the "ingroup" distribution and may, from the perspective of our respondents, not fit cleanly in the ingroup or the outgroup. In that sense, its categorization is likely to be more malleable than more consistently categorized states (e.g., the United Kingdom or Iran) and a good candidate for use in experiments in which we try to "nudge" respondents' perceptions of a state's group membership in one direction or another direction. It's also a country about which Americans know comparatively little, so there are fewer concerns about strong priors or pre-treatment.

4 Causal Research Design: Turkey as ingroup/outgroup

4.1 Design & Logistics

In our pre-registered[13] main study, we manipulate respondents' perceptions that a target country (Turkey) is in their shared ingroup or an outgroup (relative to a neutral control condition), and estimate the effect of that intervention directly on reputational and behavioral judgments. The design—depicted in Figure 4 and described in detail in Appendix G—provides additional information about Turkey to fix potential confounders and maintain information equivalence and utilizes a repeated measures structure (Clifford, Sheagley and Piston, 2021) to improve precision.

Our Identity Treatment Our identity treatment—Figure 5—shifts beliefs about whether Turkey is in respondents' ingroup or outgroup and operates through a combination of (1) information provision, (2) a graphical depictions of Turkey's leader and its geographic location relative to its neighbors, and (3) respondent participation in a brief thought-listing exercise. We focus on factors such as regime type, language, religion, and culture, identified in the vast SIT literature and highlighted by our descriptive survey as markers of international group boundaries, though we ensure in follow-ups that we are not "merely" estimating effects of regime type. Our pre-registered comparisons focus on the contrast between INGROUP and OUTGROUP treatment arms and the ATE reflects the effect of a a group categorization "nudge", and thus likely an underestimate of the effect of switching groups entirely.

Our Outcomes At the start of the survey (T_1), following our group identity manipulation (T_2), and after our past behavior manipulation (T_3), we measure the quantities listed below:

1. Warmth

[13] Link to blinded pre-analysis plan.

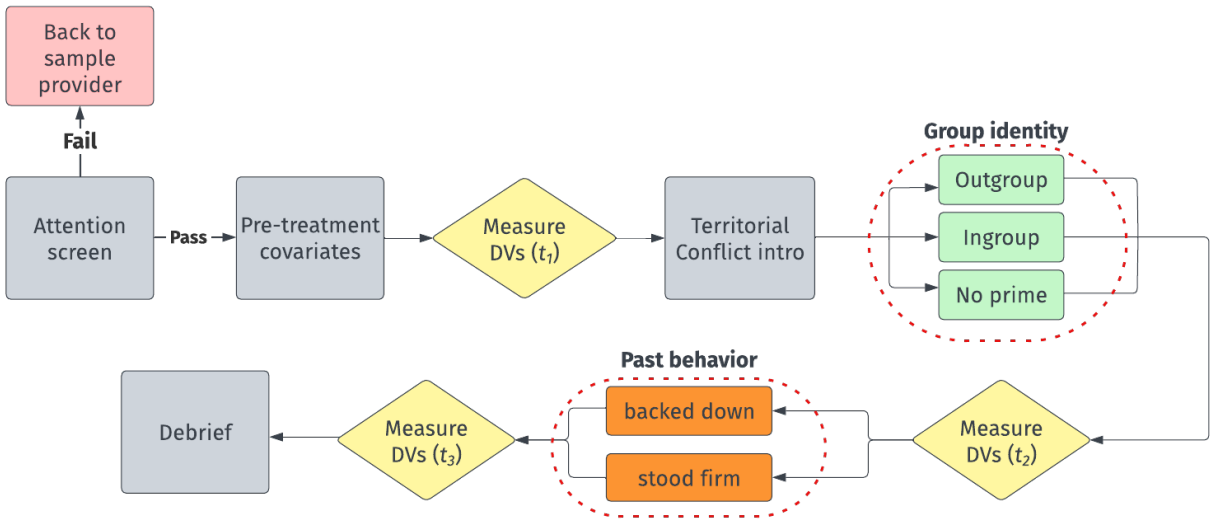


Figure 4: **Consort diagram for our main study.** Respondents begin on the far left with the "attention screen." Randomized elements are circled with a red dashed lines. Past behavior is randomized equally between arms, while respondents are assigned to identity treatments with the following probabilities: control (.16), ingroup (.42), outgroup (.42).

2. Ingroup-ness
3. Traits/Attributes
4. Behavioral Tendencies — **Reputation (What respondents *think* about Turkey)**
5. Support for Turkey — **Cooperative behavior (how respondents would *act* towards Turkey)**

Ingroup Categorization (Manipulation Check + placebo)

Our pre-registered manipulation check is an average of a measure of warmth and a question about the extent to which Turkey—alongside a number of placebo countries (UK, France, Iran)—is in the respondent's ingroup or outgroup. Our primary substantive interest is in *reputations*: beliefs about the traits/attributes of other actors, as well as their behavioral tendencies (Dafoe, Renshon and Huth, 2014, 372; see also Jervis, Yarhi-Milo and Casler, 2021). Our outcomes measure attributes or traits (e.g., tough, morally pure) as well as behavioral patterns (e.g., likely to stand firm). Notably, we expand the scope of reputations from the traditional focus in the literature on reputations for resolve, eliciting beliefs about several other types of reputations including financial reputations (repayment of debts), humanitarian reputations (willingness to accept refugees), and signaling reputation (trustworthy and likely to do what they say). Another set of questions focuses on behavioral consequences—willingness to cooperate with Turkey—that are downstream of respondents' reputational beliefs. Our outcome measures thus speak to a wide array of behaviors, including both traditional "security" outcomes (e.g., willingness to send military aid) as well as the more prosaic "transnational" relations of individuals, organizations, and firms that constitute

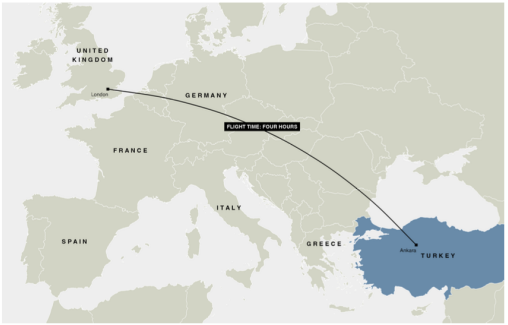
	Outgroup	Ingroup
Geography	The country of Turkey is included in maps of the Middle East.	The country of Turkey is included in maps of Europe.
	It shares a land border with Iraq, Iran, and Syria.'	Turkey shares a land border with two EU countries, Greece and Bulgaria.
	It takes less time to travel by air from Turkey's capital city to Iran's capital than from Washington, DC to London, UK.	It takes less time to travel by air from Turkey's capital city to London, UK than from Washington, DC to London, UK.
		
Language	The most common language is Turkish.	English is spoken in many tourists spots.
	It recently changed its official name to Türkiye.	It is also known as the Republic of Turkey.
Foreign Relations	It is a member of the Organization of Turkic States and TÜRKSOY, two organizations designed to promote cooperation and cultural exchange between Turkic peoples across Central and Western Asia.	It has applied for membership in the European Union an organization designed to promote cooperation and cultural exchange between European states.
Religion	The most common religion in Turkey is Islam.	The second most common religion in Turkey is Christianity.
Culture	The people of Turkey consume a lot of popular culture from their region. One of the most popular streaming services in Turkey is BluTV, located in Istanbul, which features lots of media from Turkey and around the region.	The people of Turkey are quite familiar with American popular culture. Disney+ and Netflix are quite popular in Turkey and Instagram is among the most popular social media services there
Leadership	For the past 10 years, the leader of Türkiye has been Recep Tayyip Erdoğan (pictured below).	The most recently elected president of Turkey is President Erdogan (pictured below).'
		
Thought listing	Think for a moment about the parts of Turkey that are most different to the United States. List four of them in the open response boxes below.	Think for a moment about the parts of Turkey that are most similar from the United States. List four of them in the open response boxes below.

Figure 5: Identity Treatment

the bulk of the international relations that occur on a daily basis (e.g., support immigration from Turkey and send money to help Turkey economically; Findley, Nielson and Sharman, 2013, 660).

Our DVs are all on a 0-100 scale, with higher values indicating more desirable levels of each measure. We estimate the effect of each of our group identity treatments relative to our control group using OLS, regressing each DV—at T_2 (to study effect of group identity) or T_3 (to study the conditional effect of past behavior)—on a set of treatment indicators and a set of conceptually relevant pre-treatment measures.[14]

Ethics All studies were approved by the UW-Madison IRB (#2023-1092) and the University of Georgia IRB (PROJECT00006273). No AI was used at any part of this study for either data collection, analysis or writing. In addition to standard ethical guidelines, we took additional precautions (see discussion in Appendix A) such as avoiding deception, providing an extensive debriefing to respondents and piloting to ensure that our experiment did not create or encourage negative stereotypes about Turkey.

Pilots and Logistics We pre-piloted the treatment in October 2023 on LucidTheorem (N=200), establishing that the ingroup treatment moved our manipulation checks in the correct direction, increasing the warmth felt towards the target country (Turkey) and the probability that respondents categorized Turkey as being in their ingroup while not doing either for the placebo countries (UK, France, and Iran).[15] The latter preempts concerns that the ingroup treatment is simply changing how respondents feel about all foreign countries, or doing something in particular for countries like Iran that are located in roughly the same geographic region. We fielded a more comprehensive pilot on LucidTheorem (N=2,039) in October 2023 that yielded estimates for our power analysis in our main study.[16]

Our main pre-registered study was fielded on the the U.S. public (N=6,196) through CloudResearch, and used the LucidTheorem sample population.[17] The demographic distribution of the CloudResearch sample is presented in Table 14. Pre-registered follow-up studies were fielded on Prolific in Spring, 2025.[18] Details on how we handled attentiveness are contained in Appendix G.2.

	Expectation		Findings
	Diagnostic hypotheses		
H_M	Manipulation check: Was group identity manipulation successful?	✓	Ingroup categorization shifted by 8.1 points in ingroup arm (Figure 6)
H_P	Placebo test: Was manipulation narrowly targeted at Turkey?	✓	No statistically significant differences in group categorization of UK, France, Iran across ingroup/ outgroup after B-H correction (Figure 23)
	Substantive hypotheses		
H_1	Ingroup perceived more favorably than outgroup	✓	Ingroup attributed better average reputation by 7.3 points (Figure 7)
H_2	Ingroup members will pay smaller reputational (in domain of resolve) cost for backing down down	✗	No evidence that consequences of backing down are different for ingroups/outgroups (outcome of reputation for resolve; Figure 9)
H_3	Respondents more willing to engage with ingroup	✓	Respondents 6.1% more likely to engage with Turkey when it's an ingroup member (Figure **??**)
H_G	Global test of theory (H_{1-3})	✓	We reject the global null of no effect of treatment across our three outcomes (NPC $p < .001$).
	Exploratory hypotheses		
$H_{ingroup}$	What direction did INGROUP treatment move respondents relative to CONTROL?	–	In every case, INGROUP treatment led to more positive evaluations of Turkey
$H_{outgroup}$	What direction did OUTGROUP treatment move respondents relative to CONTROL?	–	OUTGROUP treatment led to either null or more negative (but never more positive) evaluations of Turkey (relative to CONTROL)

Table 2: Outcomes for pre-registered hypotheses in Main Study

4.2 Ingroup Identity, Reputations and International Cooperation

Social Categorization Causally Affected by Ingroup Treatment At baseline (and similar to our descriptive survey respondents), subjects in the experiment placed Turkey squarely in the middle of the spectrum, with a mean value of 46 on a scale from 0 (outgroup) – 100 (ingroup).[19] Put differently, respondents viewed Turkey as neither an outgroup similar to Iran ($\mu \approx 33$) or an ingroup similar to the UK ($\mu = 62$).

Analysis of our our pre-registered manipulation checks (Figure 6) confirms that our treatment worked as intended. The INGROUP treatment shifted ingroup categorization of Turkey by 6.29 points (95% CI: 5.47, 7; $p < .001$) relative to the OUTGROUP condition, respondents felt warmer towards Turkey (by 8.1 points) and were more likely to categorize Turkey as being in the "us" group in international politics (by 4.5 points). Leveraging our neutral CONTROL group (no identity prime), we learn that the INGROUP treatment is most impactful, moving respondents by 5.2 points ($t = 9.82, p < .001$) away from the CONTROL condition in which no information about Turkey is provided (the OUTGROUP treatment moved respondents by only 1.1 points; $t = -2.19, p < .05$).[20] Notably, we do not find evidence consistent with a "rebellion" against the treatment in which some respondents felt warmer towards Turkey as a result of the OUTGROUP treatment, potentially offsetting those who responded to the treatment as we anticipated.

Our pre-registered placebo tests further increase our confidence in the precision of the treatment. Figure 23 in Appendix J.1 shows that our INGROUP and OUTGROUP treatments do *not* shift perceptions of other countries, even ones mention in the vignette (e.g., UK/Iran). We find no statistically significant differences in categorization of UK, France or Iran across ingroup and outgroup after using Benjamini-Hochberg (Benjamini and Hochberg, 2000) false discovery correction.

Global Test of Our Identity-Based Theory of Reputation Our theory generated a number of observable implications (Hypotheses 1-3) under the logic, "the more specific, the more numerous, and the more varied are the causal implications of a treatment," the more credible claims to inference are (Shadish, Cook and Campbell, 2002, 485). Before moving on to specific tests, we consider the overall credibility of our theory through the non-parametric combination (NPC) test

[14] For example, we control for our pre-treatment measures (T_1) of Turkey's reputation when estimating the effect of treatment on reputation. Results are qualitatively similar without our pre-treatment controls, but less precisely estimated.

[15] See Figure 15 in Appendix D. Our goal was to identify potential problems and evaluate the plausibility of the treatment.

[16] Pilot results in Appendix E and F.

[17] In the field from 9/9/2024–11/06/2024. Figure 21 shows recruitment over that period. N was set at 6,060 based on our power analyses (informed by estimates from our pilot; see pre-analysis plan). A minor programming error occurred in the fielding of the study; we fielded a follow-up study to ensure that it did not affect the results or our interpretation (Appendix L).

[18] See Appendix K for more information and accompanying pre-analysis plans.

[19] All results are on same 0 – 100 scale.

[20] Given that the mean level of warmness and "us-ness" in the control group was in the mid 40s, we can likely rule out "floor effects." It's possible also that it's easier to nudge social categorization towards ingroups rather than outgroups.

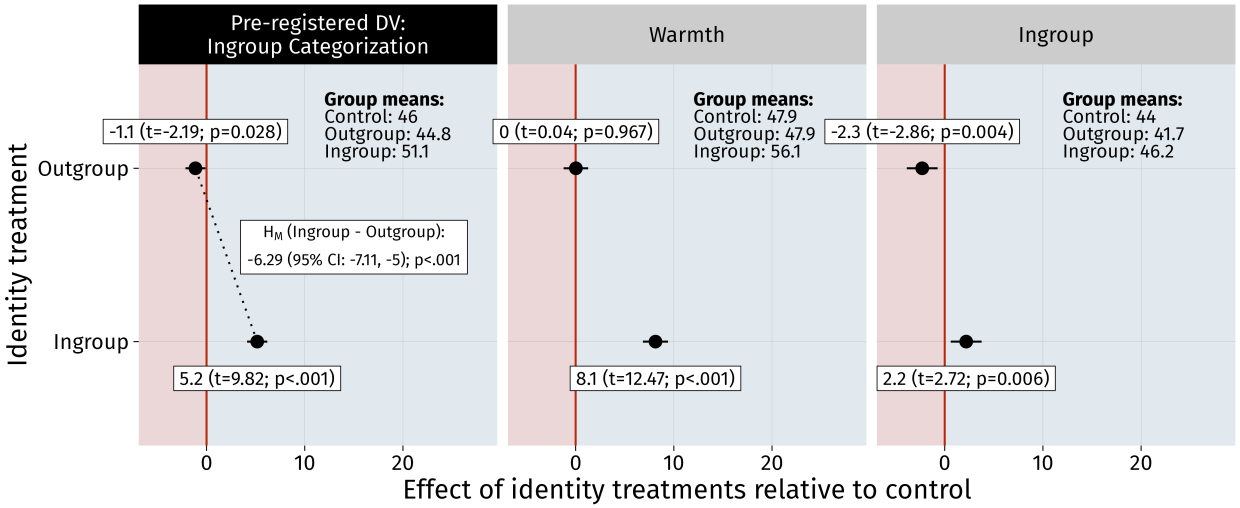


Figure 6: **Group Identity Manipulation Affects Social Categorization of Turkey**. Circles represent ATE of treatments relative to control; horizontal bars represent 95% confidence intervals. Dependent variables are all on 0-100 scale with higher values indicating feelings of greater closeness.

proposed by Caughey, Dafoe and Seawright (2017), which tests theories with multiple hypotheses against a "global null" (no effect of treatment) while taking into account the dependence among the component tests and controlling for multiple hypotheses (specifically, the familywise error rate or "FWER"). Using the NPC as specified in our registration, we reject the global null of no effect of treatment across our three outcomes (NPC $p < .001$).

Ingroup Categorization and Reputations Our theory's most straightforward implication was that that ingroup bias would positively influence evaluations of a given actor's "type" (i.e., their reputation). We operationalized this in two ways: first, as the effect of the identity treatment on reputational beliefs about Turkey, and second, as the interaction between the identity and past behavior treatments. The former test follows directly from our argument that actors make reputational judgments through "identity-colored glasses." The second addresses an additional implication of our theory: that the reputational harm of unresolved behavior (backing down) in a past coercive conflict differs for ingroup members.

We find strong evidence (Figure 7) in favor of our pre-registered H1: categorization as an INGROUP member (compared to OUTGROUP) improved Turkey's overall reputation by roughly 4.9 points (95% CI: $4.11, 6; p < .001$). An exploratory decomposition of the average reputational effects shows consistent patterns across almost all beliefs about type and likely behaviors, with only 2 exceptions (in the security domain).

To calibrate how large the effects our experiment recovered are compared to variation in reputational beliefs across different actors in world affairs, we compare our experimental results to those from our descriptive survey. Figure 8 shows that the variation in reputations accorded ingroup members compared to outgroup members is significant: moving from the outgroup to the

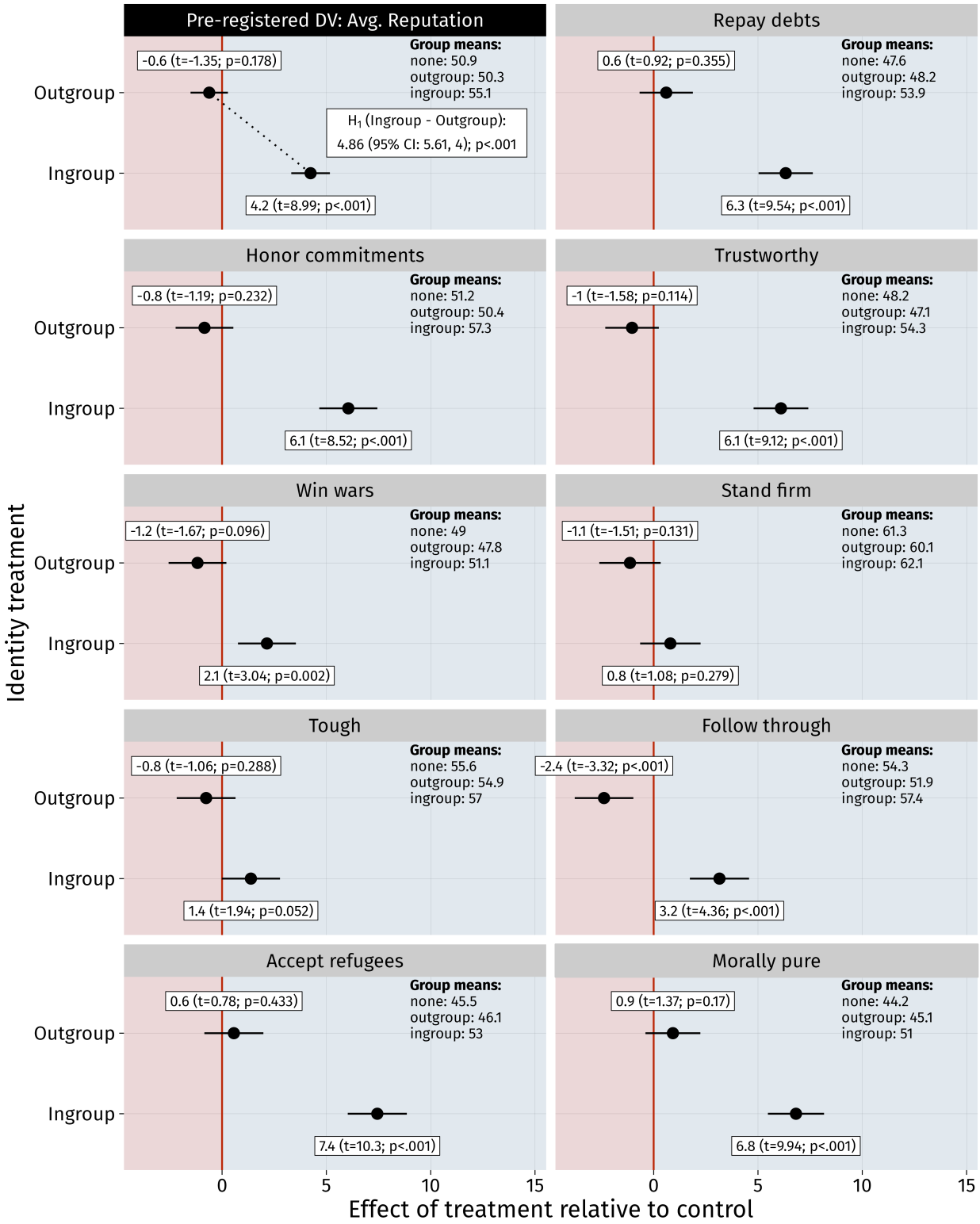


Figure 7: **Ingroup/Outgroup Treatments Affect Reputations**: ● represent ATE of treatments relative to control; horizontal bars represent 95% confidence intervals. Dependent variables are all on 0-100 scale with higher values indicating more agreement with each of the statements.

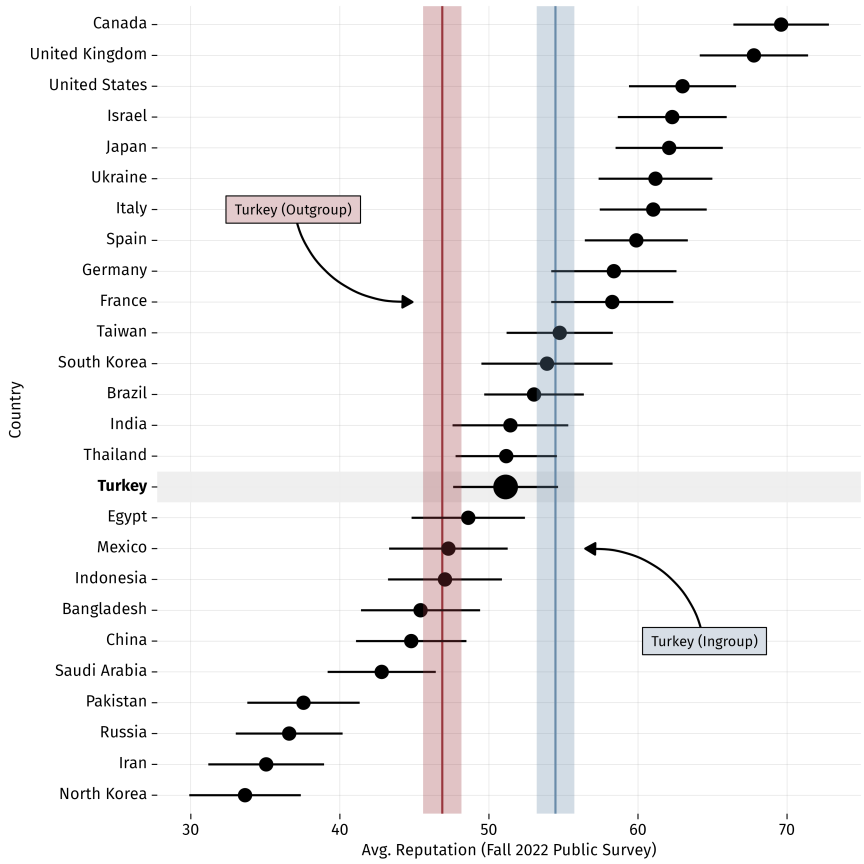


Figure 8: **Benchmarking the experimental effects of identity to our elite survey**: ● represent mean reputation scores from elite survey (with corresponding CIs). | represents mean reputation outcome in OUTGROUP arm and | represents mean reputation outcome in INGROUP treatment arm of experiment.

ingroup moved Turkey from reputational levels similar to Mexico or Indonesia to that of Taiwan (also equivalent to the difference between the reputation of China and South Korea).

We also predicted that categorization of a target state as an ingroup member would moderate the impact of observed behavior on reputations, and our pre-registered H2 operationalized this as whether the "effect of backing down"—the difference between STAND FIRM and BACK DOWN—on Turkey's reputation for resolve differed significantly between the INGROUP and OUTGROUP arms. In our analysis, we interact our group identity treatments (INGROUP, OUTGROUP, CONTROL) with our past behavior treatments (STOOD FIRM or BACKED DOWN) and re-measure outcomes for *stand firm* and *tough* at T_3.

In our pilot ($N = 2,039$), we found that the INGROUP was insulated from reputational damage from backing down compared to the OUTGROUP.[21] However, we find no evidence in favor of the pre-registered H2 interaction in our main study: estimates for our two DVs are in the correct direction but the substantive values are small and not statistically significant (see Figure 9). We

[21]For *Stand Firm*: -5.3 ($t = -2.04$, $p = .04$); For *Tough*: -3.9 ($t = -1.63$, $p = .10$).

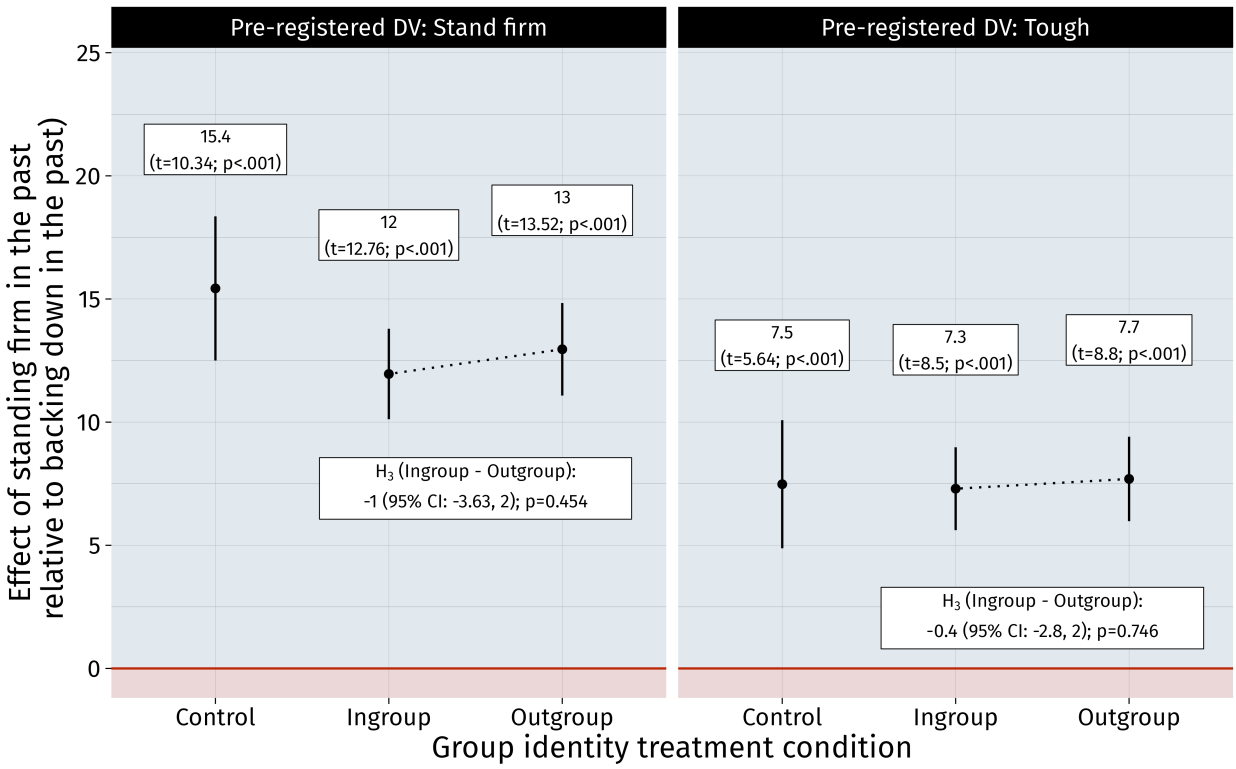


Figure 9: **Effect of Past (Resolved) Behavior on Reputation for Resolve Is Not Conditional On Group Categorization**: Here we plot a "diff in diff": the estimated difference in ATE of past behavior (standing firm) treatment between those assigned to the INGROUP treatment and those assigned to the OUTGROUP treatment.

explore potential explanations for this null further in Section 5.

Ingroup Benefits Beyond Reputations Our third registered hypothesis concerns the effect of ingroup categorization on respondents' support for cooperating with and supporting Turkey. It is thus a switch in focus from how observers *perceive* the target state of Turkey to their willingness to *act* in cooperating with Turkey in various ways.

As displayed in Figure 10, we find strong support for H3. In our main pre-registered outcome, respondents were 4.6 points (95% CI, $3.93, 5; p < .001$) more likely to engage with Turkey in the INGROUP treatment arm (relative to OUTGROUP). Decomposing the average willingness to engage, Figure 10 shows that the INGROUP treatment led to significant boosts in respondents' willingness to cooperate on *every* dimension of cooperative behavior: for example, the INGROUP treatment led to an increase of 4.7 points in willingness to send U.S. military personnel to help Turkey defend itself and 4 points in willingness to send arms and other materiel. This is striking given the fact that the scenario described in the vignette included a description of an ongoing territorial dispute; making the need, but also the costs, of supplying such aid clearer.

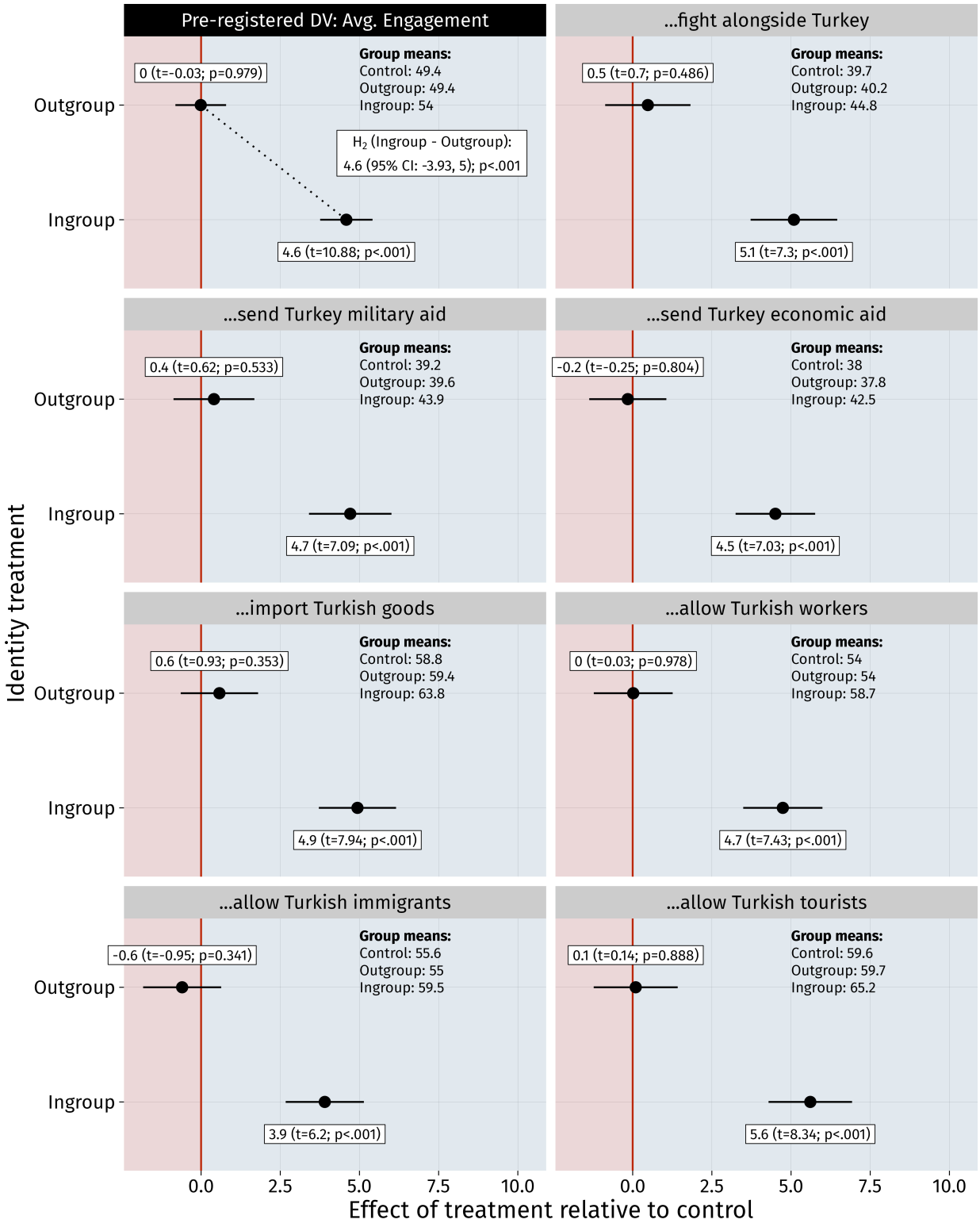


Figure 10: **Benefits to Ingroup Membership in Willingness to Cooperate with Turkey**: ● represent ATE of treatment arms relative to control; horizontal bars represent 95% confidence intervals. DVs are all on $0-100$ scale with higher values indicating greater willingness to engage in each form of exchange or cooperation.

5 Probing the effects of identity

Following our main study, we fielded four pre-registered follow-up studies focused on external validity and treatment validity (see Appendix K).[22] Below, we use them alongside exploratory analysis from the main study to shed light on how and why identity had the effects it did, as well as assessing possible explanations for the null H2.

The overarching goal in these studies was to investigate whether the direction and significance of the main ingroup/outgroup effect changed as we modified various elements of the treatment, wich we accomplished using "purposive variation" in the treatment paired with a sign-generalization test (Egami and Hartman, 2023). This test answers the question: which treatments produce precisely-estimated ATEs in the theoretically-predicted direction? The focus is thus on sign and statistical significance rather than magnitude (Bassan-Nygate et al., 2024).

Our BASELINE study ($N = 559$) was fielded > 1 year following the main study and accomplished two goals: (1) calibrating results for the rest of our follow-ups on the Prolific platform and (2) probing the external validity of our main results over time (addressing the important concern of *temporal* validity of results). Our PRIMING study ($N = 458$) addressed the concern that eliciting pre-treatment rankings of ingroup and outgroup members might prime identity concerns. Our REGIME TYPE study ($N = 1,056$) probed the extent to which beliefs about Turkey's regime type contributed to the overall effect of the identity treatment by fixing the regime at "authoritarian" across ingroup and outgroup arms. Finally, our IMAGE study ($N = 995$) contains a set of treatments that omits the photo of President Erdoğan in both conditions, shedding light on how much the text and image components contribute to the overall treatment effect in relative terms.

5.1 Treatment Generalizability

Figure 11 presents the results of the sign-generalization test (Egami and Hartman, 2023) for the main study and our four follow-ups, confirming that our treatment is highly generalizable, in the sense that variations in the experiment do not induce variations in the direction/significance of estimated causal effects. This figure confirms that our results are not due to priming respondents to think about the vignette "through the lens of identity" or reliant on the image components of the experiment (specifically, the images of Erdoğan in western or traditional clothing). Across *all* of these variations, the ATE of INGROUP is precisely estimated in the theoretically predicted direction for both of our main outcomes (reputation and cooperation).

[22] Our PAP describes them as 3 studies, with IMAGE and REGIME combined, but there are no deviations from the plan other than that framing.

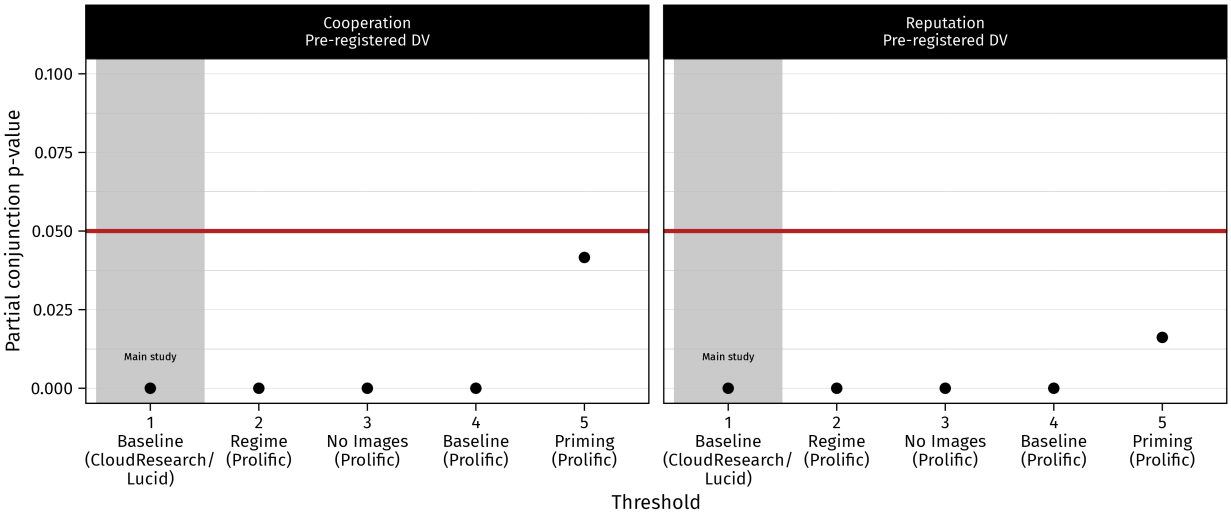


Figure 11: Sign generalization test

Probing our initial experiment's results alongside the follow-up studies generates four important lessons about how identity works to affect reputational judgments. First, identity effects are asymmetric: countries derive far more benefit from being in the ingroup compared to the disutility of being seen as in the outgruop. In fact, *ingroup* effects are $14x$ larger than *outgroup* effects across our pre-registered outcomes (consistent with original interpretations of SIT as well as recent empirics in IR; Brewer, 1999; Chu and Lee, 2024).[23]

Second, they show that, while regime type may be an important marker of identity in our original descriptive surveys, the causal effects of identity do not operate simply by changing beliefs about regime type: fixing beliefs about Turkey at "autocratic" generates identity effects on reputation of approximately the same magnitude as when regime type is allowed to vary. Third, our follow-up studies demonstrate that ingroup bias effects on reputation do not rely on triggering racial or ethnic stereotypes about outgroups. Finally, we find no evidence of treatment heteogeneity: the impact of ingroup categorization on reputations works broadly across different types of people with very different beliefs and characteristics.[24]

5.2 Probing the Null H2

In our theory, we operationalized the idea of "ingroup bias in reputations" in two ways: one in which identity had an (unconditional) overall effect and one in which it acted as a filter through which people interpreted information about behavior. Though we found strong support for the former, the latter prediction—tested using our between-subjects design—generated a null in the Main Study: we found that the "cost of backing down" (on reputations for resolve) was not significantly different for ingroups. While it's possible that identity categorization does *not* moderate the effect

[23] See Appendix J.4.

[24] For analyses described in this section, see Appendix J.5.

of bad behavior on reputations, results from our pilot experiment, alongside the strong results from the NPC test and strong (exploratory) between-subjects results[25] suggest the wisdom of deeper consideration.

Having ruled out treatment effect heterogeneity (see above), we see three explanations as particularly plausible in explaining the overall pattern of results: two design-related and one conceptual. First, because the past behavior treatment was always presented last in our design, it's possible that this information "drowned out" the effects of the earlier identity treatment. The counterfactual implicated here is a design in which past behavior and identity are manipulated simultaneously or the order was randomized.

Second, our empirical test is narrower than our theory: while the theory is about reputations broadly, our empirical test focused only on reputations *for resolve*. This almost certainly contributed to the H2 null, since the effects of identity on security related reputations were $\frac{1}{3}$-$\frac{1}{2}$ the size of its effects on non-security related reputations. The counterfactual here is a study designed around past behavior and reputations in a non-security realm, where the identity ATE would be more than twice as large. Lastly, it's possible that the theory requires refinement. For example, it's possible that ingroup identity moderates how information about past actions is judged, but only in conjunction with the *scope condition*, "...when past behavior is ambiguous in some way." This would be analogous to how cognitive dissonance works: actors might find it difficult to ignore a "perfect record of standing firm," but easier to allow ingroup bias to put its thumb on the scales when the record of past behavior is mixed.

5.3 What does our experiment have to say about other theories of identity in IR?

Earlier, we argued there are important conceptual differences between our theory and an attribution theory version of reputation (Mercer, 1997); principally that our theory does not require the input of "behavior" to operate and we allow for other country's to be part of actors' ingroup. Empirically, we can disentangle the theories by assessing the effect of resolved behavior by ingroups. Our argument predicts that such behavior will positively affect reputations for resolve while attribution theory suggests that desired behavior by allies is ignored. In fact, Figure 9 shows that resolved behavior by ingroup members does positively affect reputation, consistent with our broader argument.

And while the conceptual differences between our theory and image theory are muddier than is ideal, we sought to disentangle the theories empirically by embedding questions eliciting respondent beliefs about Turkey's level of culture relative to the U.S.. In image theory (Herrmann et al., 1997), targets are seen as "similar" culturally (in terms of levels of sophistication) for both the *ally* and the *enemy* images (Alexander, Levin and Henry, 2005). As our Figure 26 (Appendix J.3) shows, our INGROUP treatment significantly improves (by 2.12 points, CI: 1.01, 3; $p < .001$) perceptions of Turkey's culture (relative to the OUTGROUP treatment), suggesting that the ingroup/outgroup

[25] Negative information about past unresolved behavior does less reputational damage to ingroup members than to outgroup members (see Appendix J.6)

dichotomy in our experiment cannot be reduced to the ally/enemy images in image theory. Additionally, Kertzer (2023, 11) argues that images are "sticky," and that "once they become embedded, they are resistant to change." This is inconsistent with our experiment, in which the group categorization of a real country was shifted significantly by our treatment.

Discussion

We sought to make both theoretical and empirical contributions in this paper. Our theory explains how identity dynamics affect reputations as well as willingness to engage with countries in the absence of any specific behavioral cues, and makes predictions about how behavior is integrated into reputations. Empirically, we made two signal contributions. First, we fielded a descriptive survey on samples of both U.S. elites and the general public, learning that elites and the public see international identity in a similar manner, holding nearly identical reputational beliefs about other actors in international affairs as well as overlapping in their views of who is in their ingroup. We also learned that there is a strong relationship between ingroup-ness and valued traits.

Our pre-registered experiment built on our survey by using a unique design—based around a real country (Turkey)—in which respondents' beliefs about whether the target state was in the ingroup or outgroup were "nudged" in either direction. We find broad and strong support for our theory of identity and reputations: ingroup categorization affects reputations of all types—from perceived magnanimity (e.g., willingness to accept refugees) to perceived resolve (e.g., toughness and willingness to stand firm)—as well as cooperative behavior and willingness to engage.

These results are important because they suggest that group identity matters in high-stakes, but more infrequent scenarios, while our other cooperation results suggest group identity matters as well in more pedestrian (but much more frequent) settings as well. We show that respondents in the ingroup treatment were more willing to host ingroup members as tourists, welcome them as immigrants or guest workers, consume goods imported from ingroup countries, and send economic assistance to help ingroup members develop economically.

Viewing international politics through the lens of social categorization has the potential to address a number of empirical puzzles aside from variation in the updating of reputations. For example, our theory and findings have relevance for the broader debate on democratic reputations and triumphalism (Renshon, Yarhi-Milo and Kertzer, 2023). One reason why democratic reputations might be so "sticky" and resistant to change is because they are based on shared identity rather than observed behavior.

Other recent work has noted that the Russian invasion of Ukraine led to a host of puzzling changes in attitudes of European citizens. Leveraging a survey in which interviews occurred both pre and post-invasion, Klymak and Vlandas (2022) show that the Russian invasion caused respondents to both value democracy more as well as changed how they saw others by increasing acceptance of immigration from poorer and even non-EU countries. Similar trends were apparent in the U.S., where the Russian invasion led to a large initial spike in favorability of Ukraine among the general

public.[26] If it's true that "humans reliably divide the world into *us* and everyone else: *them*" (Cikara and Van Bavel, 2014, 248), then these and other changes in perceptions can be seen as the natural consequences of target states or actors getting accorded the "us treatment."

[26] https://www.pewresearch.org/global/2023/05/10/americans-hold-positive-feelings-toward-nato-and-ukraine-see-russia-as-an-enemy/.

References

Abdelal, Rawi, Yoshiko M Herrera, Alastair Iain Johnston and Rose McDermott. 2006. "Identity as a Variable." *Perspectives on politics* 4(4):695–711.

Alexander, Michele G, Shana Levin and Peter J Henry. 2005. "Image theory, social identity, and social dominance: Structural characteristics and individual motives underlying international images." *Political psychology* 26(1):27–45.

Alvarez, R Michael, Lonna Rae Atkeson, Ines Levin and Yimeng Li. 2019. "Paying attention to inattentive survey respondents." *Political Analysis* 27(2):145–162.

Aronow, Peter Michael, Joshua Kalla, Lilla Orr and John Ternovski. 2020. "Evidence of rising rates of inattentiveness on Lucid in 2020." *SocArXiv* pp. 1–10.

Avey, Paul C., Irene Entringer Garcia Blanes, Shauna N. Gillooly, Susan Peterson, Ryan Powers, Andrew Shaver and Michael J. Tierney. 2023. "TRIP Policymaker Survey 2022 Report." Unpublished manuscript.

Avey, Paul C, Michael C Desch, Eric Parajon, Susan Peterson, Ryan Powers and Michael J Tierney. 2021. "Does Social Science Inform Foreign Policy? Evidence from a Survey of US National Security, Trade, and Development Officials." *International Studies Quarterly* 66(1). sqab057. **URL:** *https://doi.org/10.1093/isq/sqab057*

Bankert, Alexa, Ryan Powers and Geoffrey Sheagley. 2022. "Trade Politics at the Checkout Lane: Ethnocentrism and Consumer Preferences." *Political Science Research and Methods* pp. 1–8.

Baser, Ekrem T. 2024. "Reputations and Change in International Relations." *International Studies Quarterly* 68(3):sqae097.

Bassan-Nygate, Lotem, Jonathan Renshon, Jessica LP Weeks and Chagai M Weiss. 2024. "The Generalizability of IR Experiments beyond the United States." *American Political Science Review* pp. 1–16.

Benjamini, Yoav and Yosef Hochberg. 2000. "On the adaptive control of the false discovery rate in multiple testing with independent statistics." *Journal of educational and Behavioral Statistics* 25(1):60–83.

Blackwell, Matthew, Ruofan Ma and Aleksei Opacic. 2024. "Assumption Smuggling in Intermediate Outcome Tests of Causal Mechanisms." *arXiv preprint arXiv:2407.07072* .

Boulding, KE. 1959. "National images and international systems." *The Journal of Conflict Resolution* 3(2):120.

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

Brewer, Marilynn B. 2001. "The many faces of socialidentity: Implications for political psychology." *Political psychology* 22(1):115–125.

Brown, Rupert. 2000. "Social identity theory: Past achievements, current problems and future challenges." *European journal of social psychology* 30(6):745–778.

Brutger, Ryan and Joshua D Kertzer. 2018. "A dispositional theory of reputation costs." *International Organization* 72(3):693–724.

Carnegie, Allison and Nikhar Gaikwad. 2022. "Public opinion on geopolitics and trade: Theory and evidence." *World Politics* 74(2):167–204.

Casler, Don and Keren Yarhi-Milo. 2023. "Signaling, resolve, and reputation in international politics." *The Oxford Handbook of Political Psychology* pp. 426–446.

Caughey, Devin, Allan Dafoe and Jason Seawright. 2017. "Nonparametric combination (NPC): A framework for testing elaborate theories." *The Journal of Politics* 79(2):688–701.

Checkel, Jeffrey T. 1999. "Norms, institutions, and national identity in contemporary Europe." *International Studies Quarterly* 43(1):83–114.

Chu, Jonathan A and Carrie A Lee. 2024. "Race, religion, and American support for humanitarian intervention." *Journal of Conflict Resolution* 68(10):2076–2100.

Chung, Eunbin. 2023. "National Identity, Social Preferences, and Foreign Policy Attitudes: Experimental Evidence from Japan." *Foreign Policy Analysis* 19(2):orad005.

Cikara, Mina and Jay J Van Bavel. 2014. "The neuroscience of intergroup relations: An integrative review." *Perspectives on Psychological Science* 9(3):245–274.

Clare, Joe and Vesna Danilovic. 2012. "Reputation for resolve, interests, and conflict." *Conflict Management and Peace Science* 29(1):3–27.

Clifford, Scott, Geoffrey Sheagley and Spencer Piston. 2021. "Increasing precision without altering treatment effects: Repeated measures designs in survey experiments." *American Political Science Review* 115(3):1048–1065.

Cuhadar, Esra and Bruce Dayton. 2011. "The social psychology of identity and inter-group conflict: From theory to practice." *International Studies Perspectives* 12(3):273–293.

Curley, Tyler M. 2009. "Social identity theory and EU expansion." *International Studies Quarterly* 53(3):649–668.

Dafoe, Allan and Devin Caughey. 2016. "Honor and war: Southern US presidents and the effects of concern for reputation." *World Politics* 68(2):341–381.

Dafoe, Allan, Jonathan Renshon and Paul Huth. 2014. "Reputation and status as motives for war." *Annual Review of Political Science* 17:371–393.

Deaux, Kay, Anne Reid, Kim Mizrahi and Kathleen A Ethier. 1995. "Parameters of social identity." *Journal of personality and social psychology* 68(2):280.

Dobrescu, Madalina. 2023. "Status Seeking through Peacekeeping: Ukraine's Quest for a Positive Social Identity in the International System." *Foreign Policy Analysis* 19(2):orad009.

Druckman, Daniel. 1994. "Nationalism, patriotism, and group loyalty: A social psychological perspective." *Mershon international studies review* 38(Supplement_1):43–68.

Egami, Naoki and Erin Hartman. 2023. "Elements of external validity: Framework, design, and analysis." *American Political Science Review* 117(3):1070–1088.

Escribà-Folch, Abel, Lala H Muradova and Toni Rodon. 2021. "The effects of autocratic characteristics on public opinion toward democracy promotion policies: A conjoint analysis." *Foreign Policy Analysis* 17(1):oraa016.

Fernández-Reino, Mariña, Valentina Di Stasio and Susanne Veit. 2023. "Discrimination unveiled: a field experiment on the barriers faced by Muslim women in Germany, the Netherlands, and Spain." *European Sociological Review* 39(3):479–497.

Findley, Michael G, Daniel L Nielson and Jason C Sharman. 2013. "Using Field Experiments in International Relations: A Randomized Study of Anonymous Incorporation1." *International Organization* 67(4):657–693.

Freeman, Bianca, DG Kim and David A Lake. 2022. "Race in International Relations: Beyond the "norm against noticing"." *Annual Review of Political Science* 25(1):175–196.

Gray, Julia. 2013. *The company states keep: International economic organizations and investor perceptions*. Cambridge University Press.

Gray, Julia and Raymond P Hicks. 2014. "Reputations, perceptions, and international economic agreements." *International Interactions* 40(3):325–349.

Gries, Peter Hays. 2005. "Social Psychology and the Identity-Conflict Debate: Is a 'China Threat'Inevitable?" *European Journal of International Relations* 11(2):235–265.

Hayes, Jarrod. 2012. "Securitization, social identity, and democratic security: Nixon, India, and the ties that bind." *International Organization* 66(1):63–93.

Herrmann, Richard K. 2017. "How attachments to the nation shape beliefs about the world: A theory of motivated reasoning." *International Organization* 71(S1):S61–S84.

Herrmann, Richard K, James F Voss, Tonya YE Schooler and Joseph Ciarrochi. 1997. "Images in international relations: An experimental test of cognitive schemata." *International Studies Quarterly* 41(3):403–433.

Huddy, Leonie. 2001. "From social to political identity: A critical examination of social identity theory." *Political Psychology* 22(1):127–156.

Hymans, Jacques EC. 2002. Applying social identity theory to the study of international politics: a caution and an agenda. In *annual meeting of the International Studies Association, New Orleans*. pp. 24–27.

Islam, Gazi. 2014. "Social identity theory." *Journal of Personality and Social Psychology* 67:741–763.

Jervis, Robert. 1989. "Additional thoughts on political psychology and rational choice." *Political Psychology* 10(3):511–515.

Jervis, Robert, Keren Yarhi-Milo and Don Casler. 2021. "Redefining the Debate over Reputation and Credibility in International Security: Promises and Limits of New Scholarship." *World Politics* 73(1):167–203.

Kalin, Michael and Nicholas Sambanis. 2018. "How to think about social identity." *Annual Review of Political Science* 21(1):239–257.

Katzenstein, Peter J. 1996. *The culture of national security: Norms and identity in world politics*. Columbia University Press.

Kertzer, Joshua. 2016. *Resolve in international politics*. Vol. 2 Princeton University Press.

Kertzer, Joshua D. 2023. "Public opinion about foreign policy." *Oxford handbook of political psychology* .

Kertzer, Joshua D and Jonathan Renshon. 2022. "Experiments and surveys on political elites." *Annual Review of Political Science* 25:529–550.

Kertzer, Joshua D, Jonathan Renshon, Keren Yarhi-Milo et al. 2019. "How Do Observers Assess Resolve?" *British Journal of Political Science* 51(1):308–330.

Kertzer, Joshua D, Marcus Holmes, Brad L LeVeck and Carly Wayne. 2022. "Hawkish Biases and Group Decision Making." *International Organization* pp. 1–36.

King, Gary and Melissa Sands. 2015. "How human subjects research rules mislead you and your university, and what to do about it.".

Klymak, Margaryta and Tim Vlandas. 2022. "The Impact of Russia's Invasion on European Attitudes." *Available at SSRN 4315398* .

Kowert, Paul and Jeffrey Legro. 1996. The culture of national security: Norms and identity in world politics. In *Norms, identity, and their limits: A theoretical reprise*. pp. 451–97.

Lacina, Bethany and Charlotte Lee. 2013. "Culture clash or democratic peace?: Results of a survey experiment on the effect of religious culture and regime type on foreign policy opinion formation." *Foreign Policy Analysis* 9(2):143–170.

Larson, Deborah Welch and Alexei Shevchenko. 2019. "Lost in misconceptions about social identity theory." *International Studies Quarterly* 63(4):1189–1191.

Levine, Mark, Amy Prosser, David Evans and Stephen Reicher. 2005. "Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior." *Personality and social psychology bulletin* 31(4):443–453.

Lo, Adeline, Jonathan Renshon and Lotem Bassan-Nygate. 2024. "A Practical Guide to Dealing with Attrition in Political Science Experiments." *Journal of Experimental Political Science* 11(2):147–161.

Lundberg, Ian, Rebecca Johnson and Brandon M Stewart. 2021. "What is your estimand? Defining the target quantity connects statistical evidence to theory." *American Sociological Review* 86(3):532–565.

Lupton, Danielle L. 2020. *Reputation for Resolve*. Cornell University Press.

Lyall, Jason, Graeme Blair and Kosuke Imai. 2013. "Explaining support for combatants during wartime: A survey experiment in Afghanistan." *American political science review* 107(4):679–705.

Mendeloff, David. 2008. "Pernicious history'as a cause of national misperceptions: Russia and the 1999 Kosovo War." *Cooperation and Conflict* 43(1):31–56.

Mercer, Jonathan. 1995. "Anarchy and identity." *International Organization* 49(2):229–252.

Mercer, Jonathan. 1997. "Reputation and rational deterrence theory." *Security Studies* 7(1):100–113.

Mercer, Jonathan. 2005. "Rationality and psychology in international politics." *International Organization* 59(1):77–106.

Mercer, Jonathan. 2010. *Reputation and International Politics*. Cornell University Press.

Mercer, Jonathan. 2023. "Racism, Stereotypes, and War." *International Security* 48(2):7–48.

Montgomery, Jacob M, Brendan Nyhan and Michelle Torres. 2018. "How conditioning on post-treatment variables can ruin your experiment and what to do about it." *American Journal of Political Science* 62(3):760–775.

Murray, Michelle. 2018. *The struggle for recognition in international relations: status, revisionism, and rising powers*. Oxford University Press.

Mutz, Diana C and Eunji Kim. 2017. "The impact of in-group favoritism on trade preferences." *International Organization* 71(4):827–850.

Oren, Ido. 1995. "The subjectivity of the'democratic'peace: changing US perceptions of Imperial Germany." *International Security* 20(2):147–184.

Otten, Sabine. 2016. "The Minimal Group Paradigm and its maximal impact in research on social categorization." *Current Opinion in Psychology* 11:85–89.

Pager, Devah. 2003. "The mark of a criminal record." *American journal of sociology* 108(5):937–975.

Pinto, Diego Costa, Adilson Borges, Márcia Maurer Herter and Mário Boto Ferreira. 2020. "Reducing ingroup bias in ethical consumption: The role of construal levels and social goodwill." *Business Ethics Quarterly* 30(1):31–63.

Powers, Kathleen E. 2022. *Nationalisms in international politics*. Princeton University Press.

Renshon, Jonathan, Allan Dafoe and Paul Huth. 2018. "Leader influence and reputation formation in world politics." *American Journal of Political Science* 62(2):325–339.

Renshon, Jonathan, Keren Yarhi-Milo and Joshua D Kertzer. 2020. Democratic reputations in crises and war. Technical report Working paper.

Renshon, Jonathan, Keren Yarhi-Milo and Joshua D Kertzer. 2023. "Democratic reputations in crises and war." *The Journal of Politics* 85(1):1–18.

Rousseau, David L. 2006. *Identifying threats and threatening identities: The social construction of realism and liberalism*. Stanford University Press.

Rousseau, David L and Rocio Garcia-Retamero. 2007. "Identity, power, and threat perception: A cross-national experimental study." *Journal of Conflict Resolution* 51(5):744–771.

Sambanis, Nicholas, Stergios Skaperdas and William C Wohlforth. 2015. "Nation-building through war." *American Political Science Review* 109(2):279–296.

Schelling Thomas, C. 1966. *Arms and influence*. New Haven, CT: Yale University Press.

Shadish, WR, TD Cook and DT Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company Boston, USA.

Sinclair, Samantha, Mark Granberg and Towe Nilsson. 2023. "Love thy (Ukrainian) neighbour: Willingness to help refugees depends on their origin and is mediated by perceptions of similarity and threat." *British Journal of Social Psychology* .

Smith, Rogers M. 2004. "Identities, interests, and the future of political science." *Perspectives on Politics* 2(2):301–312.

Stagnaro, Michael Nicholas, James Druckman, Antonio Alonso Arechar, Robb Willer, David Rand et al. 2024. "Representativeness Versus Attentiveness: A Comparison Across Nine Online Survey Samples.".

Tetlock, Philip E and James M Goldgeier. 2000. "Human nature and world politics: Cognition, identity, and influence." *International Journal of Psychology* 35(2):87–96.

Tomz, Michael. 2007. *Reputation and International Cooperation*. Princeton University Press.

Tomz, Michael and Jessica LP Weeks. 2025. "Race, Democracy, and Public Support for War." *American Political Science Review* pp. 1–18.

Weisiger, Alex and Keren Yarhi-Milo. 2015. "Revisiting reputation: How past actions matter in international politics." *International Organization* 69(2):473–495.

Weiss, Chagai M. 2025. Group identity, ingroups, and their implications. In *Handbook of Innovations in Political Psychology*. Edward Elgar Publishing pp. 36–54.

West, Emily A and Shanto Iyengar. 2022. "Partisanship as a social identity: Implications for polarization." *Political Behavior* 44(2):807–838.

Wohlforth, William C. 2009. "Unipolarity, status competition, and great power war." *World Politics* 61(1):28–57.

Yarhi-Milo, Keren, Joshua D Kertzer and Jonathan Renshon. 2018. "Tying hands, sinking costs, and leader attributes." *Journal of Conflict Resolution* 62(10):2150–2179.

Appendix Materials for "Identity and Reputation in World Politics"

Appendix

A Ethics

All experimental studies have an obligation to conform to ethical guidelines set by the relevant regulatory and field-specific institutions, but even those precautions may not be sufficient (King and Sands, 2015). We believe there to be an additional burden for designs that mention real countries, particularly ones about which respondents do not hold strong priors. The risk is that our design may change respondents' beliefs about Turkey. In order to preempt such concerns, our experiments did two things.

First, we entirely avoided the use of deception; our ingroup and outgroup treatments were all based on documented, factual sources, and certain facts were selectively emphasized at the expense of others for the different treatments. Respondents in the ingroup, for example, read about Turkey being included in maps of Europe in many atlases, its application to join the EU and other features that emphasize its commonality with the United States, including language, popular culture and religion. Respondents see a graphic showing how close Turkey is to EU countries and emphasizing its place in Western Europe as well as a photo of President Erdoğan in a western-style suit.[27] The sources for our ingroup/outgroup treatments can be found in Appendix I.

Second, our main experiment included a debriefing page (see Figure 22 in Appendix H) following the study that gave respondents an expansive history of Turkey's culture and politics, with vetted links to reliable sources if they wished to learn more. Finally, we piloted the treatment extensively, learning in that preliminary phase that the treatment was moving attitudes in a *positive* rather than negative direction (suggesting that we were unlikely to be creating or consolidating negative stereotypes about Turkey).

B Estimands

Table 3 captures our estimands, using a simple version of the framework suggested by Lundberg, Johnson and Stewart (2021). Theoretical estimands are the "questions outside of the data" and combine a (1) unit specific quantity (a realized or potential outcome) and (2) a target population. The unit-specific quantity clarifies whether the object is to make descriptive or causal inferences, and the target population addresses the question: over whom or what do we aggregate that unit-specific quantity (Lundberg, Johnson and Stewart, 2021, 534)?

C Initial Descriptive Survey

C.1 Demographic Distribution for Public Sample

	Lucid ($N = 1,599$)	2020 CCES

[27] In the OUTGROUP arm, respondents saw a photo of President Erdoğan in more traditional Arab clothing and a kefiyeh, inspired by the work of Pager (2003) and more recently, a very similar treatment used by Fernández-Reino, Di Stasio and Veit (2023).

Gender		
Male	797 (49.84%)	48.5%
Female	802 (50.16%)	51.5%
Income		
Less than $\$30,000$	647 (40.46%)	26.1%
Between $\$30,000$ and $\$59,999$	428 (26.77%)	25.41%
Between $\$60{,}000 and \$149,999$	420 (26.27%)	30.8%
$\$150,000$ or more	82 (5.13%)	6.61%
Prefer not to say	22 (1.38%)	11.08%
Age		
18–29	386 (24.14%)	21.04%
30–39	328 (20.51%)	15.93%
40–49	268 (16.76%)	13.73%
50–59	241 (15.07%)	17.56%
60–69	234 (14.63%)	17.94%
70+	142 (8.88%)	13.8%
Region		
Northeast	325 (20.33%)	18%
Midwest	297 (18.57%)	21.62%
South and Central	599 (37.46%)	40.37%
West	378 (23.64%)	20.01%
Party ID		
Democrat	595 (37.21%)	42.2%
Republican	355 (22.20%)	37.95%
Independent	494 (30.89%)	14.35%
Other	155 (9.69%)	5.5%
Education		
Some high school or less	79 (4.94%)	8.85%
High school graduate	435 (27.20%)	28.56%
Some college	344 (21.51%)	21.36%
2 year degree	115 (7.19%)	10.19%
4 year degree	364 (22.76%)	19.67%
Post-grad	255 (15.95%)	11.37%
Other/Prefer not to say	7 (0.44%)	0%
Ethnicity		
White	1,038 (64.92%)	70.28%
Hispanic	87 (5.44%)	10.08%
Black	221 (13.82%)	13.15%
American Indian or Alaska Native	22 (1.38%)	0.71%
Asian	79 (4.94%)	4.38%
Other	141 (8.82%)	1.41%
Prefer not to say	11 (0.69%)	. 0%

Question	**Theoretical Estimand**		**Empirical Estimand**	**Method**
	unit-specific quantity	*target population*		
What identities matter in international politics?	descriptive: who is in the U.S.' ingroup and what are their reputations?	U.S. population of (a) citizens (2) elites	mean in samples of (1) N=1,500 LucidTheorem respondents, using census quotas for region, gender, race and age (2) N=650 elites recruited by TRIP	Survey
What is the effect of ingroup categorization on reputations?	causal: difference in reputation to help for countries in respondents' ingroup vs. outgroup	adults (no geographic restrictions)	difference in means for average reputation (Y) between ingroup ($d = 1$) and outgroup ($d = 0$) treatment arms in U.S. online convenience sample	Survey Experiment
Does ingroup categorization moderate the effect of past behavior on reputations?	causal interaction: average difference in the effect of bad behavior on reputation for ingroups compared to outgroups	adults (no geographic restrictions)	average effect of stand firm ($w = 1$) − back down ($w = 0$) for ingroup ($d = 1$) compared to same effect for outgroup ($d = 0$) in U.S. online convenience sample	Survey Experiment
What is the effect of ingroup categorization on cooperative behavior?	causal: difference in willingness to engage for countries in respondents' ingroup vs. outgroup	adults (no geographic restrictions)	difference in means for average willingness to engage (Y) between ingroup ($d = 1$) and outgroup ($d = 0$) treatment arms in U.S. online convenience sample	Survey Experiment

Table 3: Theoretical and Empirical Estimands

C.2 Measuring Identities in International Politics

1. Attention Check 1

   – "In the grid below, you will see a series of statements. Please tell us whether you agree or disagree with each statement." A "grid" question with eight items (from agree strongly to disagree strongly), embedded in which is the statement "two is greater than one."

2. Group and Describe

   (a) "In discussing world affairs, people sometimes use words like 'us' and 'them': 'us' to describe our country and other countries that are like us in some way, or part of our group, and 'them' to describe ones that are not like us and not part of the same group. **We're interested in your views on which countries are in the 'us' group and which are in the 'them' group in international politics.** Below, you will see a list of countries. Please categorize as either being in the 'us' group or 'them' group."

   – Respondents see a list of 10 countries randomly selected from the top 25 most powerful countries as measured by the Correlates of War "composite indicator of national capabilities" measure.

   (b) "Above, you placed countries in the '**them**' group. In the spaces below, please list some words or phrases that describe what you think these countries have in common. What makes them part of the '**them**' group?" Word listing entry for respondents.

   (c) "Above, you placed countries in the '**us**' group. In the spaces below, please list some words or phrases that describe what you think these countries have in common. What makes them part of the '**us**' group?" Word listing entry for respondents.

3. Attention Check 2

   – "There are many important issues facing our country today. Research shows that issues people think are important can affect their views on other issues. We also want to know if you are paying attention. Please ignore the question and put 'crime' in the top position and 'unemployment' in the bottom position. Leave the rest of the issues in the same order. Please rank the following issues facing the nation from 1 (most important) to 7 (least important). You can change your rankings by dragging and dropping different issues."

4. Rate Countries and Groups

   – (Countries) "We're interested in perceptions of other members of the global community. **Below, we ask for your level of agreement with a number of statements about countries and groups of countries around the world.** Please give your impressions of the country or group, whether or not you know much about them. The country of [insert country name from prior question] is…"

   – Respondents rate items (trustworthy, likely to stand firm against enemies, morally pure, likely to honor commitments, likely to pay its debts, similar to the United States, powerful) from 0 (no agreement at all) to 100 (total agreement)

   – Subjects rate 3 countries chosen at random from the 10 countries they grouped in previous question.

   – (Groups) "We're interested in perceptions of other members of the global community. **Below, we ask for your level of agreement with a number of statements about countries and groups of countries around the world.** Please give your impressions of the country or group, whether or not you know much about them. Countries that can be described as [group name] are…"

– Groups are drawn at random from: "open and democratic" , "closed and authoritarian", "capitalist", "socialist", "communist", "members of the United Nations Security Council", "Western", "great powers", "wealthy", "developing," The dimensions on which they are ranked are same as above. Subjects rate 3 groups.

5. Attention Check 3

   – "In the grid below, you will see a series of statements. Please tell us whether you agree or disagree with each statement." 8 statements, embedded in which is the statement "World War I came after World War II."

6. Closed-Choice Identity Question: subjects randomized into either (a) or (b) version below.

   (a) "People can define or describe themselves in terms of social categories or groups to which they belong. In a similar way, we can think of **countries** as belonging to different groups or categories or having different identities.

   Below are a list of group memberships and identities that may be relevant to how you think about the United States in the world. **For each group, please rank its importance, so that items ranked 0 are groups that the United States is not part of or are not very important to how you think of the U.S. in the world. Items ranked 100 are terms that are very important to how you think about the United States in the world.**"

   – Groups: North-American, Western Hemisphere, Western Civilization, Free, Democratic, Superpower, Capitalist, English-Speaking, Member of the United Nations, Permanent Member of the UN Security Council, Predominantly Caucasian, Ethnic Melting Pot.

   (b) "In discussing world affairs, people sometimes use words like '**us**' and '**them**': 'us' to describe our country and other countries that are like us in some way, or part of are group, and 'them' to describe ones that are not like us and not part of the same group. We're interested in your views on what things define 'us' in international politics.

   For each item below, please rank its importance, so that items ranked closer to 0 are groups that the United States is not part of and/or terms that don't really describe the United States, while items ranked closer to 100 are terms that describe the United States very well and/or are groups that the United States is part of."

   – Same groups as above.

7. Attention Check 4

   – "We are also interested in what sections people like to read in the newspaper. This might affect what they learn from articles and how they feel about the issues discussed in them. We also want to see if people are reading the questions carefully. To show that you've read this much, please mark both the classified and none of the above boxes below. That's right, just select these two options only. Regardless of how frequently you read the newspaper, what would you say are your favorite newspaper sections to read? (please check all that apply)."

8. Open-Ended Identity Question: subjects randomized into either (a) or (b) version below.

   (a) "People can define or describe themselves in terms of social categories or groups to which they belong. In a similar way, we can think of **countries** as belonging to different groups or categories or having different identities.

For example, some group memberships, or identities, that might be relevant for the United States are: 'English-speaking' or 'democracy.' **What are some groups or social categories that are relevant to you as you think about the United States and its identity in the world?** Please list as many as you can think of."

(b) "In discussing world affairs, people sometimes use words like '**us**' and '**them**': "us" to describe our country and other countries that are like us in some way, or part of our group, and "them" to describe ones that are not like us and not part of the same group. We're interested in your views on what things define 'us' in international politics. For example, some may label democracies as part of the "us" group and dictatorships part of the 'them' group. **What are groups or social categories that you would say are in the 'us' group?**"

9. Constant-Sum Identity Tradeoff

   – "Think again about the identities that are important to the United States that we asked about above.

   Use the boxes to the right to indicate how important each of the following categories are to how you think about our country's place and identity in the international community.

   Lower values imply less importance. You have a total of 100 'points' to spend across all categories. You may allocate them across categories as you see fit. Those that are of no importance can take a value of zero."

   – Groups: North-American, Western Hemisphere, Western Civilization, Free, Democratic, Superpower, Capitalist, English-Speaking, Member of the United Nations, Permanent Member of the UN Security Council, Predominantly Caucasian, Ethnic Melting Pot.

C.3 Three lessons from our descriptive survey

In the main text, we discuss three main lessons from our descriptive surveys. We present figures based on the results of our elite and public samples. Here we show that results are similar if we also include IR scholars.

Elites and the public hold similar reputational beliefs about other states

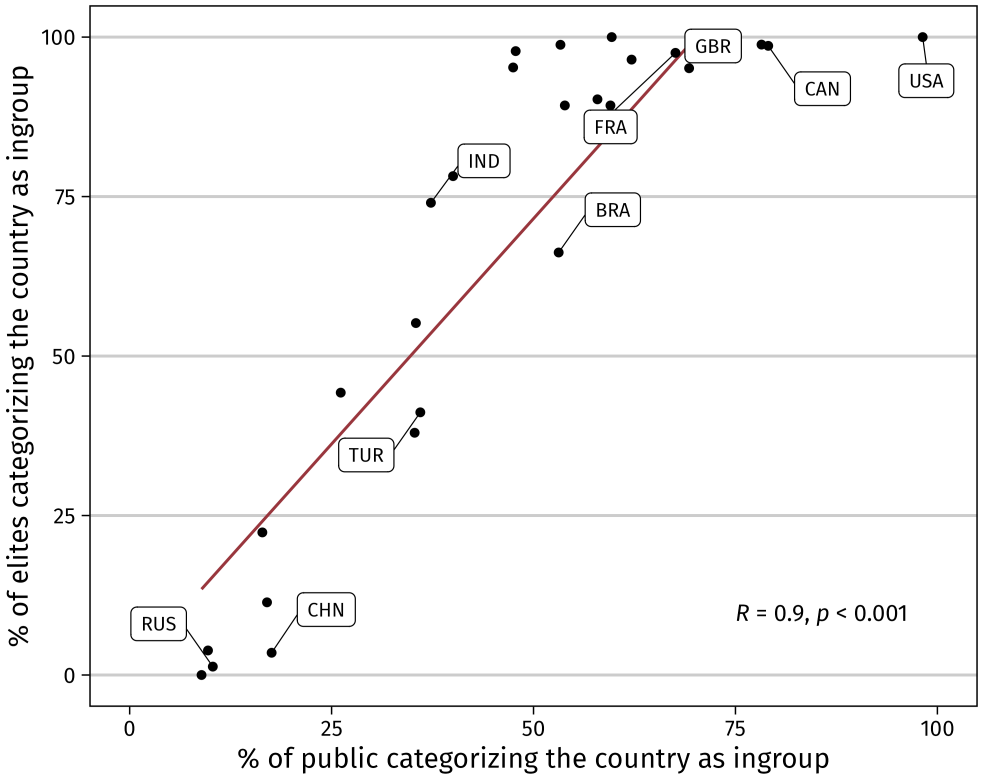


Figure 12: **Elites and the public agree on who is in the ingroup.** We asked respondents to categorize countries as either "us" or "them." The scatter plot shows that countries that elites were most likely to categorize as "us" were also categorized as "us" by the public.

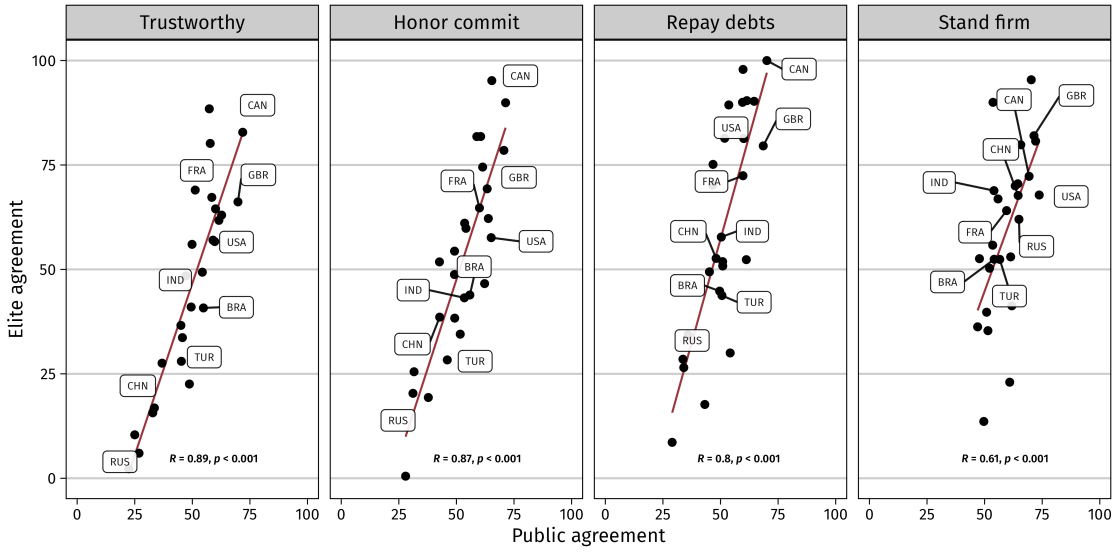


Figure 13: **Elites and the public hold similar reputational beliefs.** We asked respondents to indicate their level of agreement with a variety of statements about different countries.

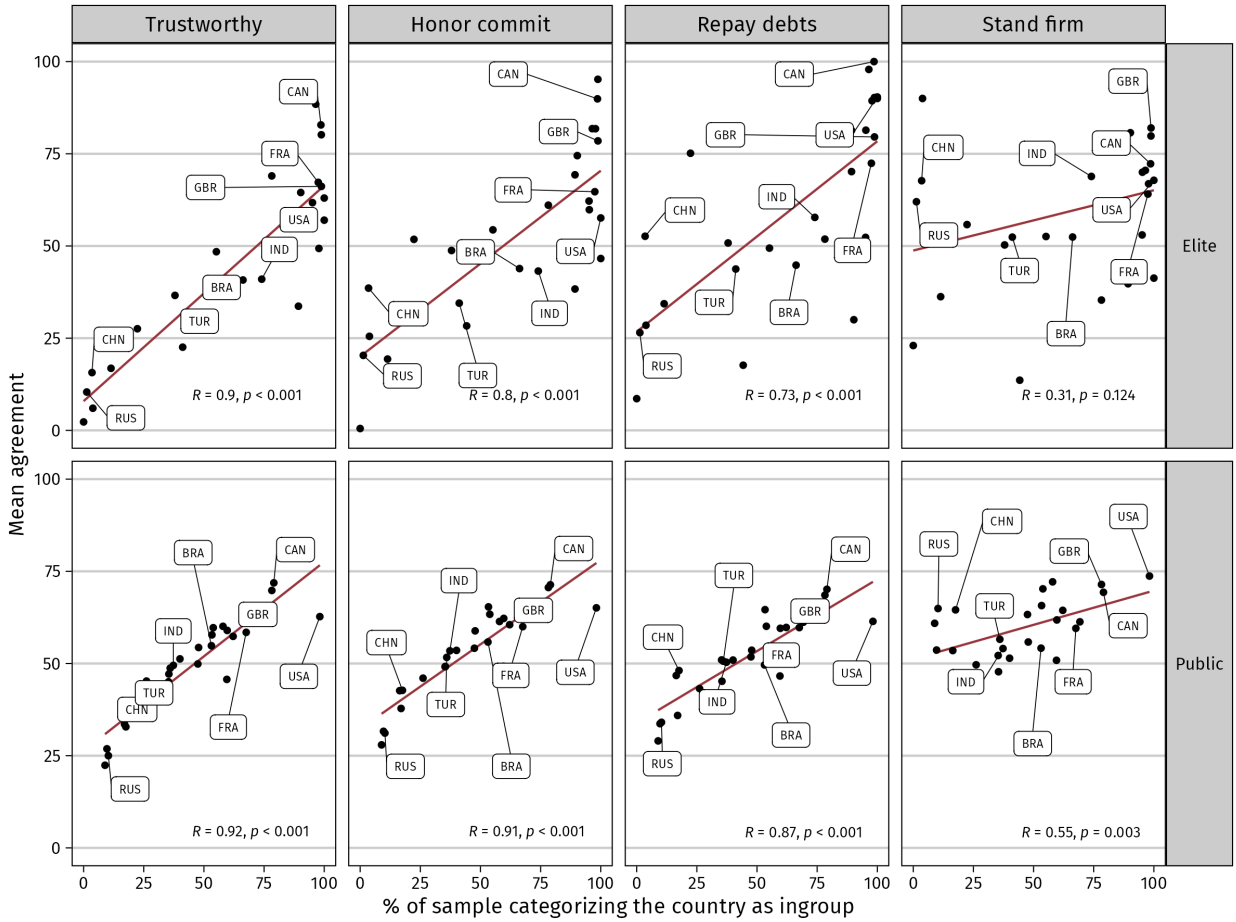


Figure 14: **Positive traits are correlated with "us-ness."**.

D Pre-Pilot Results

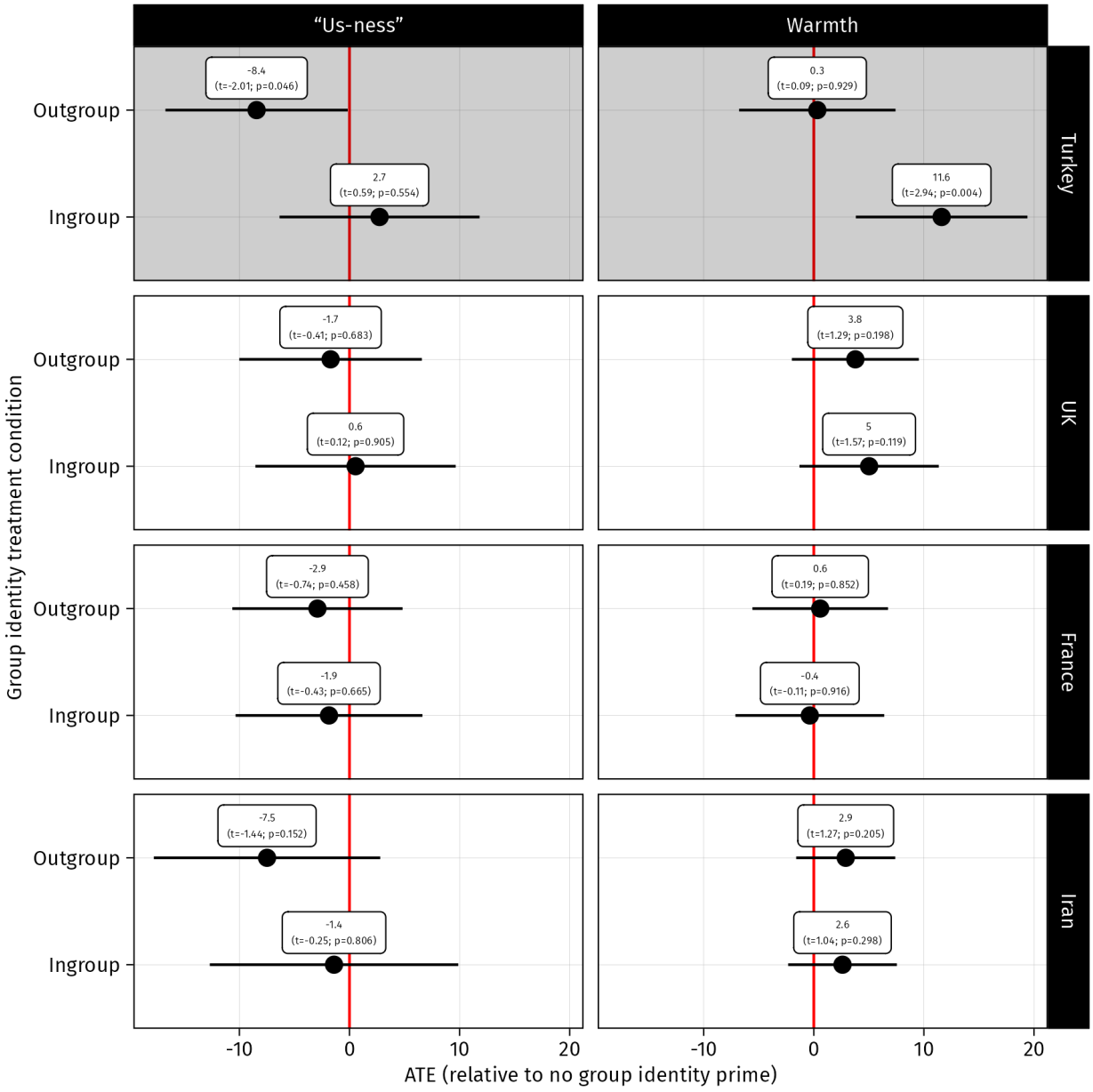


Figure 15: Results from survey of October Pre-Pilot of Treatment (N=200).

E Pilot Description

This study proceeded as follows:

1. Measure baseline beliefs about Turkey's reputation, willingness to engage, warmth, and "usness" (T_1):

   (a) Reputation: The country of Turkey is... (Response: 0 = No agreement; 100 = Total agreement)

      i. ...likely to stand firm against its enemies
      ii. ...tough
      iii. ...likely to honor its international commitments
      iv. ...trustworthy
      v. ...morally pure
      vi. ...likely to win a war if it chooses to start one
      vii. ...likely to do the things it says it will do
      viii. ...likely to repay its debts
      ix. ...likely to accept refugees from other countries

   (b) Willingness to Engage: International relations sometimes requires working with or even defending other members of the global community. How willing would you to do the following with or for the country of Turkey? (Response: 0 = No agreement; 100 = Total agreement)

      i. Allow consumer goods from Turkey to be sold in the United States
      ii. Allow tourists from Turkey into the United States
      iii. Allow Turkish citizens to live and work in the United States
      iv. Allow Turkish citizens to become U.S. citizens
      v. Send money to Turkey to help it grow economically
      vi. Send guns, missiles, and other military supplies to help Turkey defend itself against an attack by another country
      vii. Send U.S. military personnel to help defend Turkey against an attack by another country

   (c) Us vs. Them: In discussing world affairs, people sometimes use words like "us" and "them": "us" to describe our country and other countries that are like us in some way, or part of our group, and "them" to describe ones that are not like us and not part of the same group. Where would you place the following countries along an "us" to "them" continuum? (Response: 0 = Entirely us; 100 = Entirely them. Reverse coded for analysis).

      i. Turkey

ii. France

iii. Iran

iv. United Kingdom

(d) Warmth: We are interested in your feelings toward other members of the global community. We would like you to indicate how warmly you feel toward each of the countries listed below. (Response: 0 = very cold/unfriendly; 100 = very warm/friendly).

i. Turkey

ii. France

iii. Iran

iv. United Kingdom

2. All respondents read a baseline scenario about what is described as the beginnings of a diplomatic conflict between two countries, Turkey, and another country, B:

> Turkey is currently involved in a public territorial dispute with another country. Turkey maintains that the territory belongs to Turkey, but the other country has disputed this claim. Because of the risk that disputes of this nature might escalate, there has been significant media coverage of the confrontation in Turkey and around the world.

3. Subjects randomized to receive an identity prime of "no additional info," "ingroup prime" or "outgroup prime" (p=1/3 each). The details of these manipulations are given in the main text.

4. Measure beliefs about Turkey's reputation (T_2).

5. All subjects randomized into further information about Turkey's behavior:

   - Past behavior: no information, stood firm in the past, backed down in the past

6. All subjects complete post-vignette DV measurement (T_3)

F Pilot Results

F.1 Demographic Distribution

	Distribution (N = 2,158)
Gender	
Male	1,044 (48.38%)
Female	1,114 (51.62%)
Prefer not to answer	0 (0.00%)
Income	
Less than $49,999	1,090 (50.51%)
$50,000 to $99,999	716 (33.18%)
$100,000 to $149,999	220 (10.19%)
$150,000 or more	125 (5.79%)
Prefer not to answer	7 (0.32%)
Age	
18–29	365 (16.91%)
30–39	425 (19.69%)
40–49	414 (19.18%)
50–59	379 (17.56%)
60–69	335 (15.52%)
70+	240 (11.12%)
Region	
Northeast	387 (17.93%)
Midwest	437 (20.25%)
South and Central	849 (39.34%)
West	485 (22.47%)
Party ID	
Democratic	995 (46.11%)
Independent/Other	395 (18.30%)
Republican	768 (35.59%)
Prefer not to say	0 (0.00%)
Education	
Some high school or less	57 (2.64%)
High school graduate	396 (18.35%)
Some college or vocational training	618 (28.64%)
2 year degree	251 (11.63%)
4 year degree	522 (24.19%)
Post-grad	313 (14.50%)
Prefer not to say	1 (0.05%)
Ethnicity	
White	1,581 (73.26%)
Black	229 (10.61%)
American Indian or Alaska Native	42 (1.95%)
Asian	89 (4.12%)
Prefer not to say	0 (0.00%)

F.2 Attention Checks

We ask several attention checks in our study. The first asked respondents to select their favorite color from a list but included special instructions in a preamble to ignore the question and select "red" and "green." In all, about 32 percent (2202/6846) of those who agreed to participate answered this question correctly. We embedded a second attention check in a grid question related to the respondent's foreign policy dispositions. The question told the respondent to select "disagree" in a given row of the matrix. About 92 percent of those who passed the first attention check passed our second attention check. Those respondents who failed either attention check (70.2%) were re-directed back to the sample provider; they did not complete any further questions or participate in the experiments.

We screen for inattention because Lucid has acquired a reputation for inattentive respondents (Aronow et al., 2020). Indeed, Stagnaro et al. (2024) anticipated that their Lucid sample would yield "failed attention rates of between 1/3 - 2/3" in their recent comparative study of online sample providers. Two-thirds is not so far off our 70 percent failure rate and our attention checks were somewhat more demanding than their "CAPTCHA-style" attention checks (e.g., their first attention check was: "Please write 'twenty-five' using numbers. (free response box)"). While Lucid samples lag the field somewhat in attentiveness, they lead the field in representativeness. Stagnaro et al. (2024) compared samples from Lucid, Prolific, mTurk and others and concluded that Lucid and another provider, Bovtiz, provided samples that were "the closest to representative probability samples on demographic and attitudinal representativeness." While their study showed that filtering inattentive respondents decreases representativeness somewhat, even the most aggressive filters yielded a sample that was closer to their national probability sample benchmark than all other providers except for Bovitz.

Outcome	n	percent
(pre-treatment): Select 'Green' and 'Red' from multiple select		
Failed	4662	66.5%
Passed	2350	33.5%
Total	7012	-
(pre-treatment): Select 'Disagree' from matrix question		
Failed	182	7.7%
Passed	2168	92.3%
Total	2350	-
(post-treatment): Select 'Neither Agree nor Disagree' from matrix question		
Failed	26	1.3%
Passed	2008	98.7%
Total	2034	-

Table 6: Distribution of attention check outcomes. Pre-treatment questions used to select for attentive respondents. Post-treatment question used to assess how durable attention levels were over the course of the experiment.

F.3 Reputation Outcomes Conditional on Past Actions

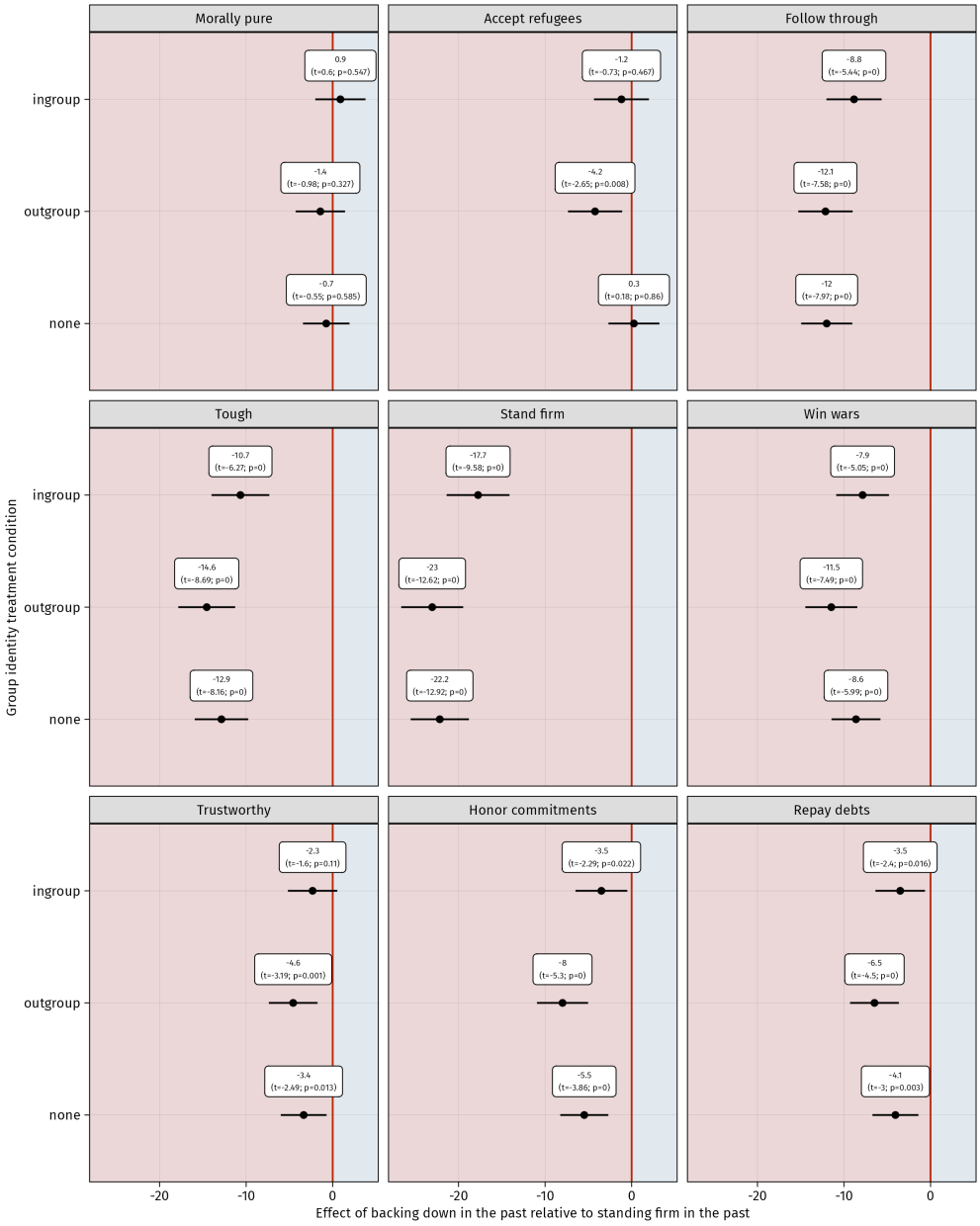


Figure 16: **Past Behavior Treatments Affect Reputations**: Results from October 2023 Pilot (N=2,039).

Word stems			
Ingroup condition		Outgroup condition	
Word	Pct. of Respondents	Word	Pct. of Respondents
christian	46.2	religion	48.6
english	46.0	languag	34.6
cultur	35.4	islam	29.4
religion	24.6	cultur	20.4
presid	24.2	border	19.2
media	20.8	countri	18.7
speak	20.7	leader	17.0
social	17.6	turkish	14.0
american	16.7	iran	12.6
spoken	15.3	turkei	11.4

Bigram stems			
Ingroup condition		Outgroup condition	
Bigram	Pct. of Respondents	Bigram	Pct. of Respondents
speak english	21.5	middl east	14.7
social media	21.3	common religion	13.4
pop cultur	12.3	main religion	8.5
american cultur	9.8	land border	8.1
popular cultur	8.6	common languag	7.4
elect presid	5.8	speak turkish	6.5
stream servic	5.8	stream servic	5.8
common religion	5.4	iraq iran	5.1
land border	5.0	territori disput	4.1
english spoken	4.8	share border	3.2

Table 7: Word and bigram stems from the thought listing excercise.

F.4 Substantive Results from Pilot

- Word and bigram stems from thought listing exercise in Pilot study depicted in Table 7
- Pilot manipulation checks are in Figure 17
- Pilot H1 results are in Figure 18
- Pilot H2 results are in Figure 19
- Pilot H3 results are in Figure 20

F.5 Results Tables for Models Used to Produce Figures

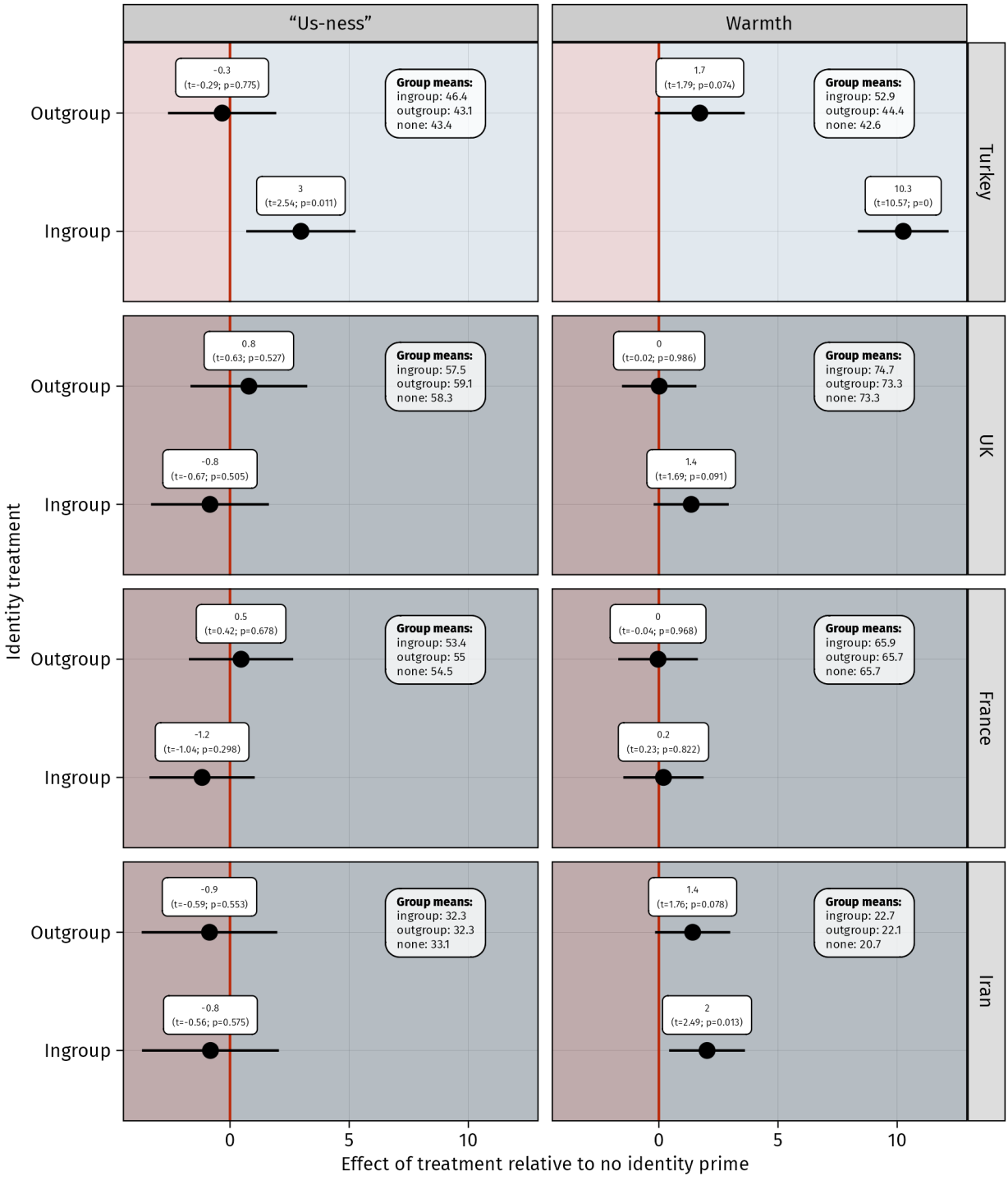


Figure 17: **Group Identity Manipulation Affects Social Categorization of Turkey But (Generally) Not Other States**. Circles represent ATE of treatment relative to control; horizontal bars represent 95% confidence intervals. Dependent variables are all on 0-100 scale with higher values indicating feelings of greater closeness. Results from October 2023 Pilot (N=2,039).

Figure 18: **Ingroup/Outgroup Treatments Affect Reputations**: Circles represent ATE of treatment relative to control; horizontal bars represent 95% confidence intervals. Dependent variables are all on 0-100 scale with higher values indicating more agreement with each of the statements. Results from October 2023 Pilot (N=2,039).

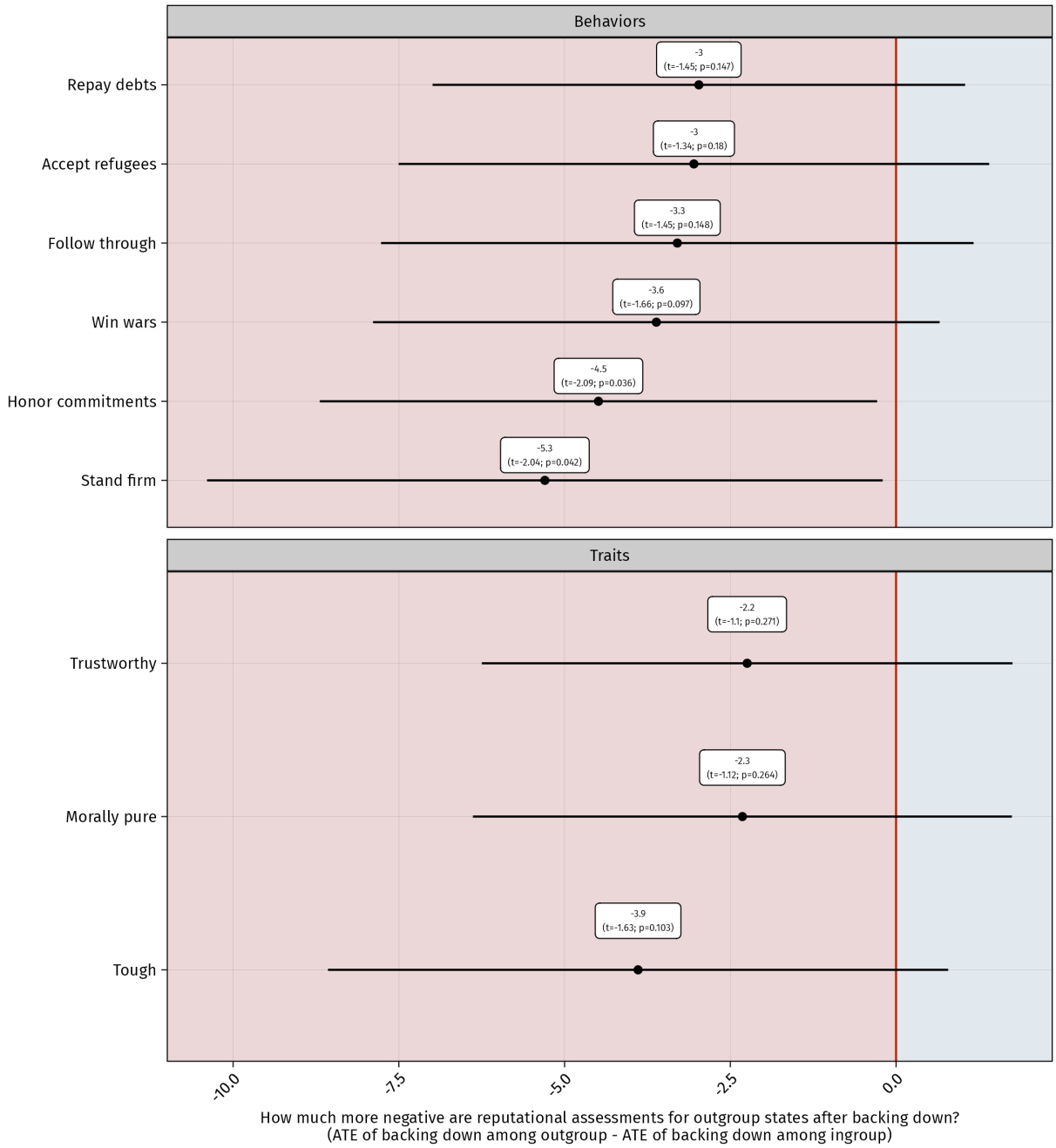


Figure 19: **Effect of Past Behavior Is (Sometimes) Conditional On Group Categorization**: Here we plot a "diff in diff": the estimated difference in ATE of past behavior (backing down) treatment between those assigned to the INGROUP treatment and those assigned to the OUTGROUP treatment. We learn, e.g., that the "effect of backing down" is more negative (−4.5) for outgroup states (compared to ingroup states). Results from October 2023 Pilot (N=2,039).

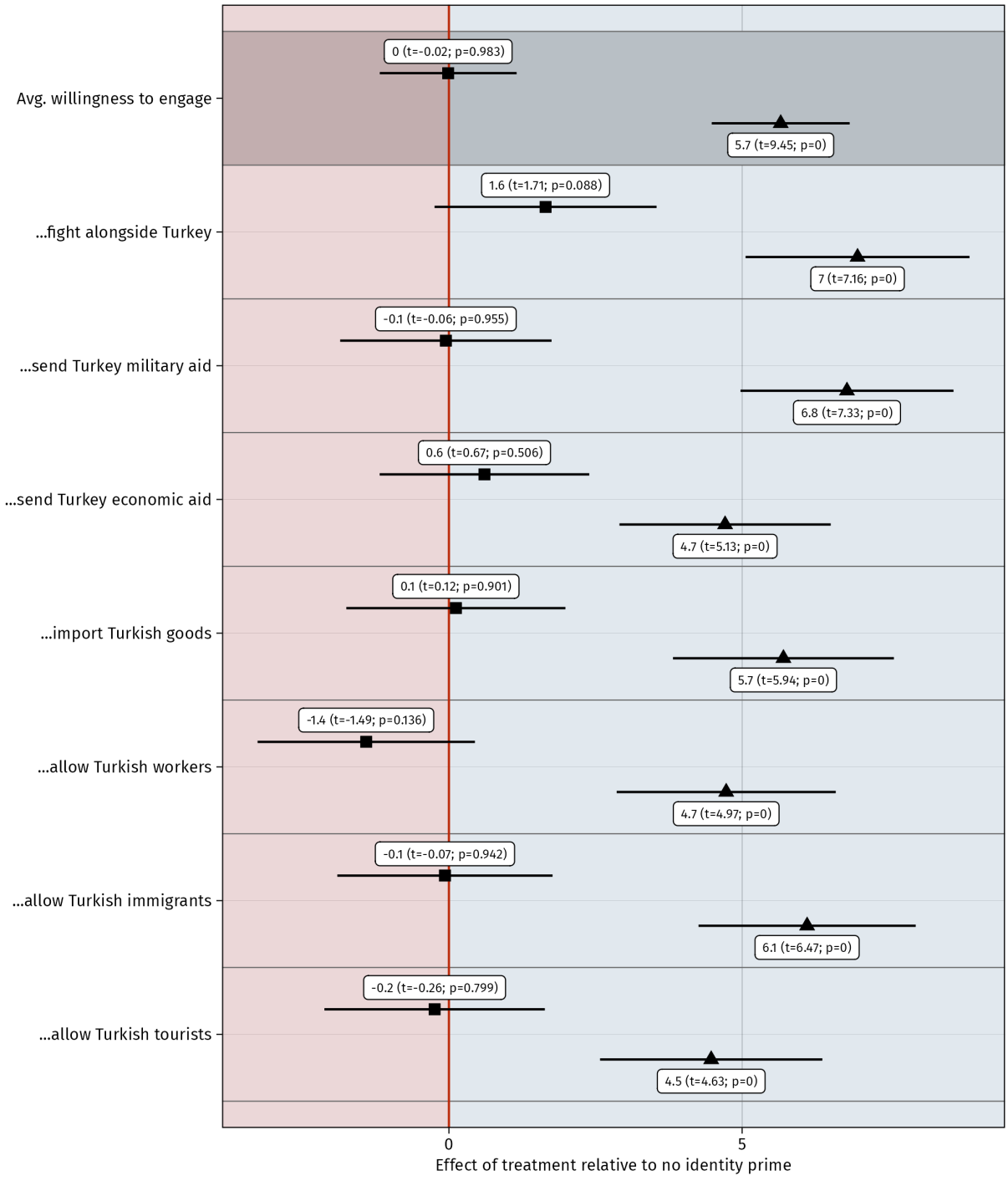


Figure 20: **Benefits to Ingroup Membership in Willingnes to Engage with Turkey**: ■ (Outgroup) and ▲ (Ingroup) represent ATE of treatment relative to control; horizontal bars represent 95% confidence intervals. DVs are all on 0-100 scale with higher values indicating greater willingness to engage in each form of exchange or cooperation. Results from October 2023 Pilot (N= 2,039).

Table 8: H_P: Placebo tests with Benjamini-Hochberg (BH) correction.

Country	Contrast	Estimate	Rank (i)	Num. tests (m)	Crit. vals $(i/m) * .05)$	$P_i < (i/m) * .05)$
Iran	ingroup - outgroup	0.7 (t=1.56; p=0.119)	1	3	0.017	False
France	ingroup - outgroup	-0.3 (t=-0.83; p=0.407)	2	3	0.033	False
UK	ingroup - outgroup	-0.1 (t=-0.34; p=0.736)	3	3	0.050	False

	Stand Firm	Tough	Honor Commit	Trust
Identity Manipulation (Baseline: Control)				
Outgroup	−2.62*	0.31	1.97*	0.54
	(1.07)	(1.00)	(0.99)	(0.92)
Ingroup	0.65	2.05*	10.64***	9.18***
	(1.08)	(1.01)	(1.00)	(0.92)
Num.Obs.	2054	2054	2054	2054
R2	0.399	0.495	0.513	0.572

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 9: Results used to produce, in part, Figure 18. Models estimated via OLS. All dependent variables are on a 0–100 scale. All models control for pre-treatment (T_1) measures of the DVs (coefficients omitted).

	Pure	Win	Follow thru	Repay debts	Accept refugees
Identity Manipulation (Baseline: Control)					
Outgroup	1.49	−2.54**	−0.58	2.55**	2.69*
	(0.93)	(0.96)	(1.00)	(0.94)	(1.07)
Ingroup	8.96***	1.76+	4.82***	8.98***	10.93***
	(0.94)	(0.96)	(1.01)	(0.95)	(1.08)
Num.Obs.	2054	2054	2054	2054	2054
R2	0.546	0.485	0.492	0.555	0.453

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 10: Results used to produce, in part, Figure 18. Models estimated via OLS. All dependent variables are on a 0–100 scale. All models control for pre-treatment (T_1) measures of the DVs (coefficients omitted).

	Stand Firm	Tough	Honor Commit	Trust
Identity Manipulation (Baseline: Ingroup)				
Control	0.49	−2.60	−7.47***	−7.74***
	(1.76)	(1.62)	(1.45)	(1.39)
Outgroup	−0.38	−1.20	−4.47**	−7.03***
	(1.84)	(1.69)	(1.52)	(1.45)
Past Behavior Manipulation (Baseline: Stand Firm)				
Back down	−17.74***	−10.66***	−3.49*	−2.32
	(1.85)	(1.70)	(1.53)	(1.46)
Interactions				
Control * Back down	−4.42+	−2.19	−1.98	−1.03
	(2.52)	(2.32)	(2.08)	(1.98)
Outgroup * Back down	−5.30*	−3.89	−4.49*	−2.25
	(2.60)	(2.39)	(2.14)	(2.04)
Num.Obs.	2038	2038	2038	2038
R2	0.375	0.427	0.496	0.533

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 11: Results used to produce Figure 16. Models estimated via OLS. All dependent variables are on a 0–100 scale. All models control for pre-treatment (T_1) measures of the DVs (coefficients omitted).

	Pure	Win	Follow thru	Repay debts	Accept refugees
Identity Manipulation (Baseline: Ingroup)					
Control	−7.38***	−1.35	−2.99+	−7.29***	−8.15***
	(1.41)	(1.48)	(1.55)	(1.39)	(1.54)
Outgroup	−5.48***	−1.16	−2.99+	−5.09***	−3.59*
	(1.47)	(1.55)	(1.62)	(1.45)	(1.61)
Past Behavior Manipulation (Baseline: Stand Firm)					
Back down	0.89	−7.85***	−8.84***	−3.51*	−1.18
	(1.48)	(1.55)	(1.62)	(1.46)	(1.62)
Interactions					
Control * Back down	−1.64	−0.76	−3.16	−0.55	1.44
	(2.01)	(2.12)	(2.21)	(1.99)	(2.21)
Outgroup * Back down	−2.32	−3.62+	−3.30	−2.97	−3.05
	(2.07)	(2.18)	(2.28)	(2.05)	(2.27)
Num.Obs.	2038	2038	2038	2038	2038
R2	0.521	0.420	0.446	0.530	0.462

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 12: Results used to produce Figure 19. Models estimated via OLS. All dependent variables are on a 0–100 scale. All models control for pre-treatment (T_1) measures of the DVs (coefficients omitted).

	Mean	Goods	Tourists	Immigrants	Workers	Econ Aid	Mil Aid	Fight For
Identity Manipulation (Baseline: Control)								
Outgroup	−0.068	−0.027	−0.384	−0.503	−1.017	0.225	0.179	1.052
	(0.594)	(0.914)	(0.907)	(0.894)	(0.894)	(0.861)	(0.896)	(0.926)
Ingroup	5.705***	5.664***	4.923***	5.851***	4.869***	4.955***	6.954***	6.718***
	(0.598)	(0.920)	(0.914)	(0.900)	(0.900)	(0.867)	(0.902)	(0.933)
Num.Obs.	2054	2054	2054	2054	2054	2054	2054	2054
R2	0.794	0.687	0.704	0.739	0.733	0.672	0.672	0.628

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 13: Results used to produce Figure 20. Models estimated via OLS. All dependent variables are on a 0–100 scale. All models control for pre-treatment (T_1) measures of the DVs (coefficients omitted).

G Main Study

G.1 Recruitment

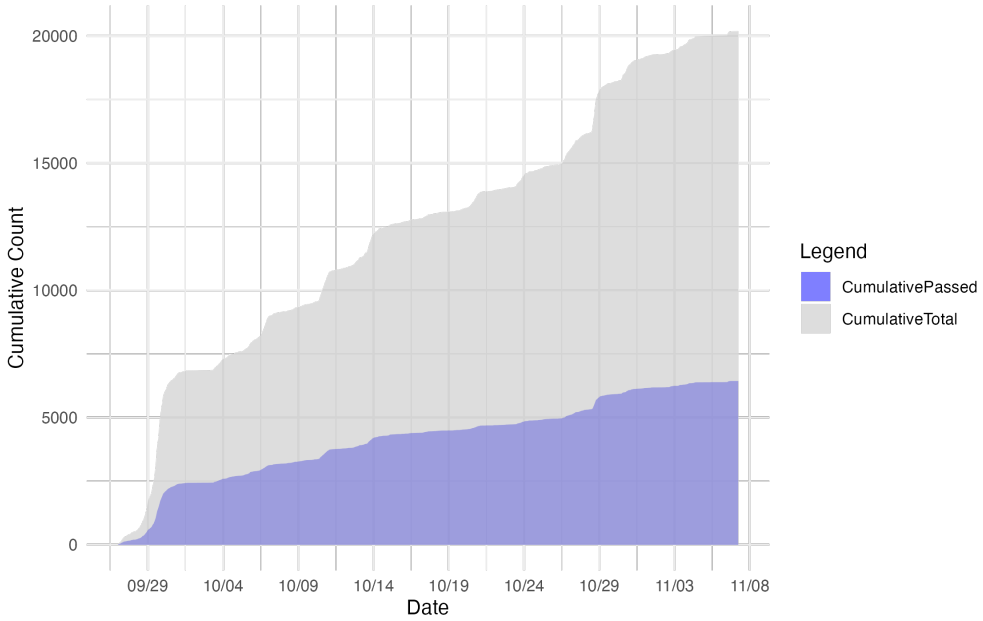


Figure 21: Recruitment over time for the main study.

G.2 Attention Checks

In accordance with our pre-registration, we used two pre-treatment checks to screen for inattentive respondents. Detailed information about pass rates and procedures are in Table 6 in Appendix F. We screen for attentiveness because LucidTheorem has acquired a reputation for inattentive respondents, though that is balanced out by a recent finding that Lucid provides "the closest to representative probability samples on demographic and attitudinal representativeness" (Stagnaro et al., 2024). Screening for attentiveness also increases the efficiency of the experiment (Alvarez et al., 2019), and in additional analyses in the Appendix we show that predictors of attentiveness are unrelated to respondents' potential outcomes. All our attention checks are pre-treatment except for one one additional post-treatment attention check in order to ensure that the "remainers" paid attention throughout the study (however, we do not use this information to censor the sample; Lo, Renshon and Bassan-Nygate, 2024).

G.3 Demographic distribution

Table 14: Distribution of demographics

Label	CloudResearch/Lucid	CCES Benchmark (2020)
Gender		
Male	3,012/6,196 (48.61\%)	48.5%
Female	3,162/6,196 (51.03\%)	51.5%
Other/Prefer not to say	22/6,196 (0.36\%)	
Total	6,196	
Income		
Less than $30,000	1,704/5,922 (28.77\%)	26.1%
Between $30,000 and $59,999	1,869/5,922 (31.56\%)	25.41%
Between $60,000 and $120,000	1,839/5,922 (31.05\%)	25.67%
More than $120,000	417/5,922 (7.04\%)	11.74%
Prefer not to say	93/5,922 (1.57\%)	11.08%
Total	5,922	
Race		
Asian	206 (3.32\%)	4.29%
Black	862 (13.91\%)	12.86%
Hispanic	783 (12.64\%)	9.86%
Indigenous	52 (0.84\%)	0.69%
White	4,196 (67.71\%)	68.76%
Other	69 (1.11\%)	3.54%
Total	6,197	
Age		
18–29	1,050 (16.94\%)	21.04%
30–39	1,072 (17.30\%)	15.93%
40–49	1,017 (16.41\%)	13.73%
50–59	975 (15.73\%)	17.56%
60–69	1,173 (18.93\%)	17.94%
70+	909 (14.67\%)	13.8%
Total	6,197	
Party ID		
Democrat	2,385/5,925 (40.25\%)	42.2%
Independent/Other	1,716/5,925 (28.96\%)	19.85%
Republican	1,824/5,925 (30.78\%)	37.95%
Total	5,925	
Region		
Northeast	1,186/6,196 (19.14\%)	18%
Midwest	1,362/6,196 (21.98\%)	21.62%
South	2,400/6,196 (38.73\%)	39.28%
West	1,137/6,196 (18.35\%)	21.1%
Total	6,196	
Education		
No high school diploma	137/5,920 (2.31\%)	8.85%
High school graduate	1,547/5,920 (26.13\%)	28.56%
Some college	1,461/5,920 (24.68\%)	21.36%
2 year degree	762/5,920 (12.87\%)	10.19%
4 year degree	1,422/5,920 (24.02\%)	19.67%
Post-grad	591/5,920 (9.98\%)	11.37%
Total	5,920	

G.4 Results tables

Table 15: Results tables for Figure 7.

	Avg. Rep.	Stand Firm	Tough	Honor Commit	Trust	Pure	Win	Follow through	Repay debts	Accept refugees
Identity Manipulation (Baseline: Outgroup)										
Ingroup	4.86***	1.94***	2.15***	6.89***	7.14***	5.90***	3.33***	5.53***	5.73***	6.87***
	(0.38)	(0.57)	(0.55)	(0.55)	(0.53)	(0.55)	(0.56)	(0.56)	(0.53)	(0.57)
None	0.62	1.14	0.76	0.84	1.04	−0.92	1.18	2.38**	−0.60	−0.57
	(0.48)	(0.74)	(0.71)	(0.70)	(0.69)	(0.71)	(0.72)	(0.72)	(0.68)	(0.74)
Gender (Baseline: Female)										
Male	0.60+	0.36	1.15*	0.34	1.47**	1.47**	1.34*	1.24*	0.71	1.57**
	(0.36)	(0.55)	(0.53)	(0.52)	(0.51)	(0.53)	(0.54)	(0.54)	(0.51)	(0.55)
Other	−3.10	−3.12	−1.76	−2.07	−5.40	−8.15*	−5.59	−5.25	−4.28	−4.55
	(2.82)	(4.30)	(4.16)	(4.11)	(4.01)	(4.11)	(4.19)	(4.21)	(3.96)	(4.30)
Region (Baseline: Midwest)										
Northeast	−0.38	−0.01	0.98	−0.43	0.18	−0.70	0.76	−0.59	−0.47	0.16
	(0.53)	(0.81)	(0.79)	(0.78)	(0.76)	(0.78)	(0.79)	(0.80)	(0.75)	(0.82)
South and Central	0.94*	1.35*	1.92**	0.77	1.51*	0.77	1.03	−0.05	0.39	1.12
	(0.45)	(0.69)	(0.66)	(0.66)	(0.64)	(0.66)	(0.67)	(0.67)	(0.63)	(0.69)
West	−0.23	0.07	0.98	−0.53	−0.25	−0.69	0.27	−0.30	−0.83	0.20
	(0.54)	(0.82)	(0.79)	(0.78)	(0.77)	(0.79)	(0.80)	(0.80)	(0.76)	(0.82)
Ideology (7pt L-R)	−0.35***	−0.47**	−0.45**	−0.64***	−0.61***	−0.55***	−0.48***	−0.73***	−0.53***	−0.67***
	(0.10)	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	(0.15)	(0.14)	(0.15)
Age	0.04***	0.09***	0.07***	0.01	0.01	−0.05***	0.04**	0.03+	0.01	−0.03+
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
DV at T_1	0.87***	0.64***	0.69***	0.72***	0.73***	0.73***	0.63***	0.69***	0.73***	0.69***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	4.52***	18.82***	14.39***	14.17***	11.74***	15.42***	17.09***	16.09***	13.32***	16.12***
	(0.94)	(1.36)	(1.30)	(1.31)	(1.27)	(1.30)	(1.31)	(1.33)	(1.25)	(1.37)
Num.Obs.	5933	5933	5933	5933	5933	5933	5933	5933	5933	5933
R2	0.659	0.385	0.432	0.490	0.527	0.512	0.404	0.439	0.527	0.474

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 16: Results tables for Figures 6 and 23

	DV: Ingroup categorization				DV: Warmth				DV: Us-ness			
	Turkey	France	Iran	UK	Turkey	France	Iran	UK	Turkey	France	Iran	UK
Identity Manipulation (Baseline: Control)												
Ingroup	5.16***	0.12	−0.14	0.33	8.13***	0.73	0.69	0.59	2.19**	−0.48	−0.03	−0.87
	(0.53)	(0.50)	(0.52)	(0.58)	(0.67)	(0.57)	(0.61)	(0.55)	(0.82)	(0.83)	(1.01)	(0.89)
Outgroup	−1.13*	0.45	−0.01	−0.37	0.03	0.87	0.61	0.40	−2.29**	0.02	−1.35	−0.41
	(0.53)	(0.50)	(0.52)	(0.58)	(0.66)	(0.57)	(0.61)	(0.55)	(0.81)	(0.83)	(1.01)	(0.89)
Gender (Baseline: Female)												
Male	−0.13	−0.03	0.22	0.26	0.13	0.53	0.26	0.99*	−0.39	−0.58	0.25	−0.56
	(0.40)	(0.37)	(0.39)	(0.43)	(0.50)	(0.42)	(0.45)	(0.41)	(0.61)	(0.62)	(0.75)	(0.67)
Other	−3.27	−1.79	2.87	−2.81	−6.52+	−5.58+	−4.12	0.77	−0.02	2.01	−1.51	4.98
	(3.11)	(2.91)	(3.03)	(3.37)	(3.87)	(3.31)	(3.54)	(3.22)	(4.74)	(4.83)	(5.86)	(5.19)
Region (Baseline: Midwest)												
Northeast	−0.77	−0.43	−0.23	−0.11	−1.14	−0.98	−0.18	−1.13+	−0.40	0.12	−0.04	0.68
	(0.59)	(0.55)	(0.57)	(0.64)	(0.73)	(0.63)	(0.67)	(0.61)	(0.90)	(0.92)	(1.11)	(0.98)
South and Central	0.18	0.43	0.45	0.48	0.10	0.17	0.95+	0.36	0.26	0.70	0.02	0.53
	(0.50)	(0.47)	(0.48)	(0.54)	(0.62)	(0.53)	(0.57)	(0.52)	(0.76)	(0.77)	(0.94)	(0.83)
West	−0.42	−0.03	−0.01	−0.02	−0.96	−0.68	−0.02	−1.06+	0.13	0.63	−0.02	1.04
	(0.59)	(0.56)	(0.58)	(0.64)	(0.74)	(0.63)	(0.68)	(0.62)	(0.91)	(0.92)	(1.12)	(0.99)
Ideology (7pt L-R)	−0.28*	−0.36***	−0.21*	−0.40***	−0.52***	−0.65***	−0.49***	−0.38***	−0.04	−0.08	−0.32	−0.04
	(0.11)	(0.10)	(0.11)	(0.12)	(0.14)	(0.12)	(0.13)	(0.11)	(0.17)	(0.17)	(0.21)	(0.18)
Age	−0.02+	0.04**	0.05***	−0.11***	−0.04*	0.04**	−0.16***	0.08***	0.00	0.03	−0.05*	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Closeness Turkey T1	0.52***	0.00	0.02	0.06**	0.10***	−0.09***	−0.08***	−0.06**	0.93***	0.10**	0.21***	0.10**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)
Closeness France T1	0.06**	0.50***	0.13***	0.04+	−0.07*	0.06*	−0.04	−0.05*	0.19***	0.94***	0.13**	0.30***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)
Closeness UK T1	0.05*	0.16***	0.53***	−0.11***	−0.06*	−0.04*	−0.11***	0.03	0.15***	0.36***	−0.11**	1.03***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)
Closeness Iran T1	0.04**	−0.01	−0.07***	0.57***	−0.10***	−0.07***	0.02	−0.09***	0.18***	0.06*	1.11***	−0.06*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Warmth Turkey T1	0.10***	−0.01	−0.02	0.00	0.59***	0.05***	0.12***	0.03+	−0.38***	−0.07**	−0.12***	−0.06*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)
Warmth France T1	−0.02	0.13***	−0.02	−0.05**	0.09***	0.65***	0.02	0.12***	−0.13***	−0.39***	−0.12***	−0.17***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Warmth Iran T1	0.00	−0.05***	−0.04**	0.10***	0.14***	0.02*	0.72***	0.01	−0.14***	−0.13***	−0.53***	−0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Warmth UK T1	−0.01	−0.01	0.15***	0.00	0.09***	0.16***	0.02	0.70***	−0.12***	−0.18***	−0.01	−0.41***
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Intercept	13.67***	13.57***	15.76***	21.35***	12.68***	14.56***	19.76***	16.50***	14.66***	12.59***	22.94***	15.02***
	(1.30)	(1.22)	(1.27)	(1.41)	(1.62)	(1.39)	(1.49)	(1.35)	(1.99)	(2.03)	(2.46)	(2.18)
Num.Obs.	5933	5933	5933	5933	5933	5933	5933	5933	5933	5933	5933	5933
R2	0.421	0.534	0.560	0.538	0.563	0.636	0.681	0.647	0.359	0.438	0.403	0.448

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: coefficients for pre-treatment controls omitted.

H Debrief for respondents in main study

Putting the information you read about Turkey in context

In this survey, you read a hypothetical scenario in which Turkey was involved in a territorial dispute with another country. We also provided information about the country of Turkey.

While the details of the dispute and Turkey's past behavior were hypothetical, the information about the country of Turkey itself was true. Still, we highlighted only a few aspects of Turkey's diverse culture, history, and identity. Below, you can see a map of the full region in which Turkey is located and some additional information that puts the information about Turkey that you read into a broader context. We also provide links to our sources for the facts we listed about Turkey and to reputable outlets that provide even more information about Turkey and its culture.

(a)

Some more information about Turkey:

- The country of Turkey is included in maps of Europe and the Middle East.
- The country of Turkey is officially known as the Republic of Türkiye but it is also frequently referred to as Turkey.
- It applied for membership in the European Union in 1987 and is currently a "candidate country." Turkey is a member of the Organization of Turkic States and TÜRKSOY, two organizations designed to promote cooperation and cultural exchange between Turkic peoples across Central and Western Asia.
- Turkey shares a land border with two EU countries, Greece and Bulgaria and with Syria, Iran, and Iraq.
- It takes about four hours to travel by plane from the capital of the UK to the capital of Turkey. It takes about three hours to travel by plane from the capital of Turkey to the capital of Iran.
- The vast majority of Turkey is Muslim; a small minority of Turkey is Christian.
- Turkish is the official language of Turkey but English is spoken in many tourists spots.
- The most recently elected president of Turkey is Recep Tayyip Erdoğan. He has been in power for ten years.
- The people of Turkey are quite familiar with American popular culture. Disney+ and Netflix are quite popular in Turkey and Instagram is among the most popular social media services there but they also rely on BluTV, founded in Istanbul, which features lots of media from Turkey and around the region. BluTV was recently acquired by Warner Bros.

The Wikipedia article on Turkey has even more information [LINK]. You might also be interested in the Encyclopedia Britannica page on Turkey [LINK]. If you are interested in learning about Turkish food, CNN has a list of popular Turkish dishes [LINK].

(b)

Figure 22: Debriefing page for respondents

Table 17: Results table for Figure 9.

	Stand Firm	Tough
Identity Manipulation (Baseline: Control)		
Ingroup	4.98***	3.78***
	(1.24)	(1.13)
Outgroup	1.18	0.43
	(1.24)	(1.13)
Past Behavior Manipulation (Baseline: Back down)		
Stand firm	15.43***	7.48***
	(1.45)	(1.32)
Interactions		
Outgroup * Stand firm	−2.47	0.21
	(1.73)	(1.58)
Ingroup * Stand firm	−3.47*	−0.18
	(1.74)	(1.58)
Gender (Baseline: Female)		
Male	1.04	1.78**
	(0.65)	(0.59)
Other	−1.20	−5.63
	(5.05)	(4.60)
Region (Baseline: Midwest)		
Northeast	0.94	0.40
	(0.96)	(0.87)
South and Central	0.18	0.95
	(0.81)	(0.74)
West	0.07	0.68
	(0.96)	(0.88)
Ideology (7pt L-R)	−0.44*	−0.67***
	(0.18)	(0.16)
Age	0.04+	0.04*
	(0.02)	(0.02)
DV at T_1	0.56***	0.65***
	(0.01)	(0.01)
Intercept	13.15***	12.26***
	(1.85)	(1.68)
Num.Obs.	5933	5933
R2	0.303	0.378

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

I Sources For Group Identity Manipulation

Our goal is to manipulate group identity without deception by selectively making facts highlighting either similarities or differences between the United States and Turkey more salient. Below we have disaggregated our manipulations into discreet facts and provided citations to the sources that substantiate those claims.

I.1 Ingroup

- 'The country of Turkey is included in maps of Europe.' Source: United States Library of Congress.
- 'It is also known as the Republic of Turkey.' Source: Encyclopedia Britannica.
- 'It has applied for membership in the European Union an organization designed to promote cooperation and cultural exchange between European states.' Source: European Union website on European Neighbourhood Policy and Enlargement Negotiations
- 'Turkey shares a land border with two EU countries, Greece and Bulgaria.' Source: Encyclopedia Britannica.
- 'It takes less time to travel by air from Turkeyś capital city to London, UK than from Washington, DC to London, UK.', Source: Google Flights.
- 'The second most common religion in Turkey is Christianity and in many tourists spots English is spoken.' Source: Wikipedia.
- 'The most recently elected president of Turkey is President Erdoğan (pictured below).' Source: Wikipedia.
- 'The people of Turkey are quite familiar with American popular culture. Disney+, Netflix, and YouTube are among the top streaming services in Turkey and Instagram is among the most popular social media services there.' Sources: Flixpatrol.com and OOSGA.

I.2 Outgroup

- 'The country of Turkey is included in maps of the Middle East in many atlases.' Source: United States Library of Congress.
- 'It recently changed its official name to Türkiye.' Source: Source: Encyclopedia Britannica.
- 'It is a member of the Organization of Turkic States and TÜRKSOY, two organizations designed to promote cooperation and cultural exchange between Turkic peoples across Central and Western Asia.' Sources: Organization of Turkic States and TÜRKSOY.
- 'It shares a land border with Iraq, Iran, and Syria.' Source: Encyclopedia Britannica.

- 'It takes less time to travel by air from Turkey's capital city to Iran's capital than from Washington, DC to London, UK.' Source: Google Flights.
- 'The most common religion in Turkey is Islam.' Source: Encyclopedia Britannica.
- 'The most common language is Turkish.' Source: Wikipedia.
- 'For the past 10 years, the leader of Türkiye has been Recep Tayyip Erdoğan (pictured below).' Source: Wikipedia.
- 'The people of Turkey consume a lot of popular culture from their region. One of the most popular streaming services in Turkey is BluTV, founded in Istanbul, which features lots of media from Turkey and around the region.' Source: Hollywood Reporter.

J Additional Analyses

J.1 Placebo Tests

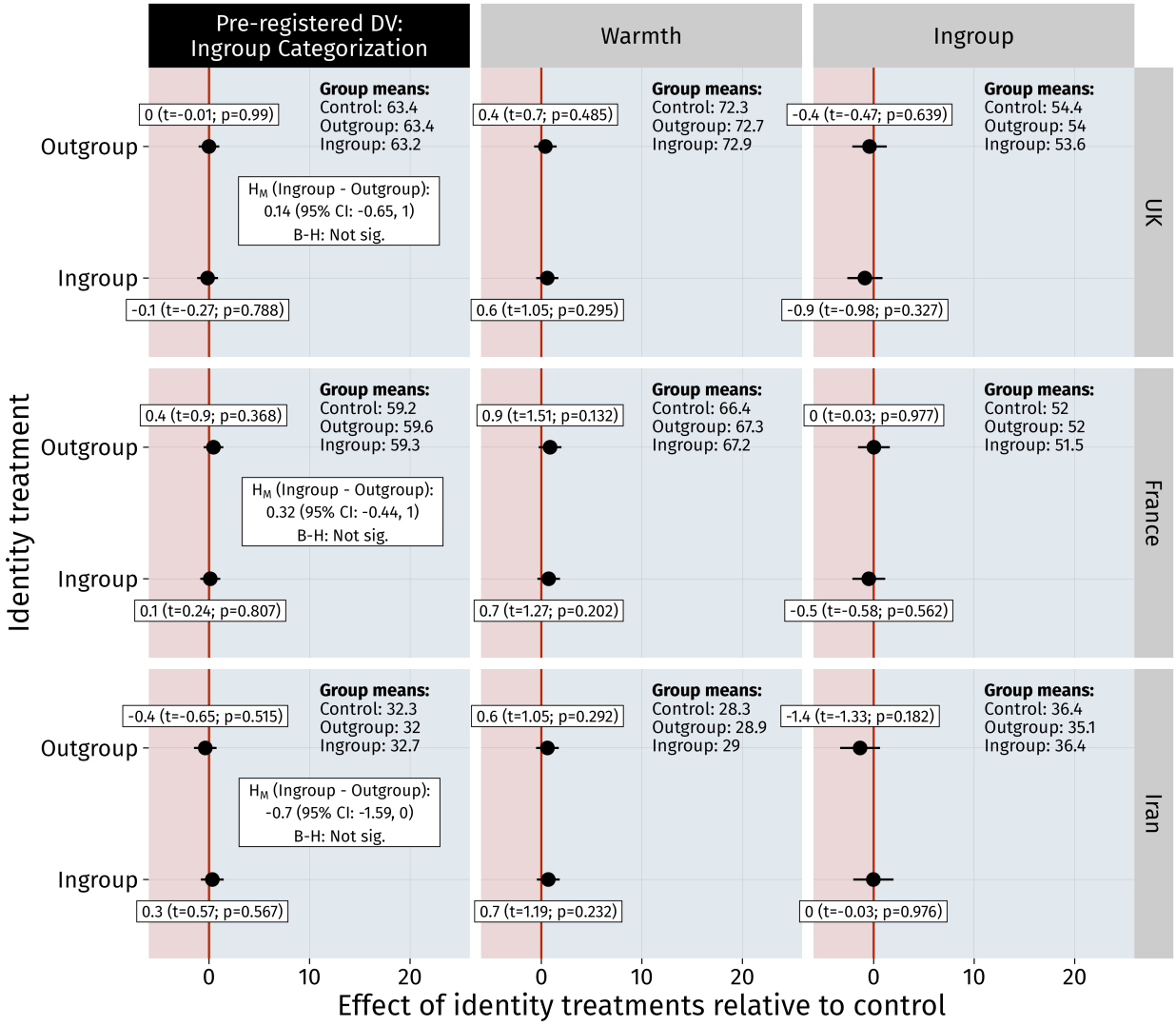
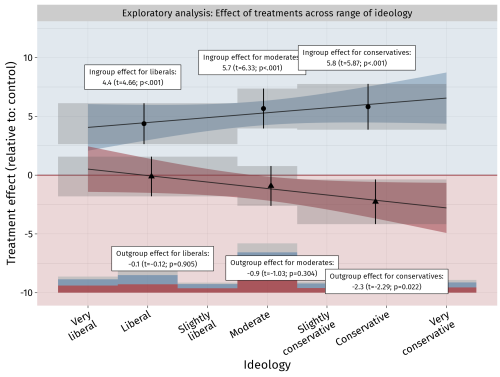
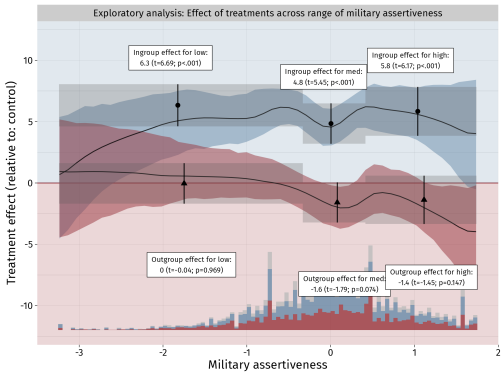


Figure 23: **No evidence of treatment spillover to placebo countries**. Circles represent ATE of treatments relative to control; horizontal bars represent 95% confidence intervals. Dependent variables are all on 0-100 scale with higher values indicating feelings of greater closeness.

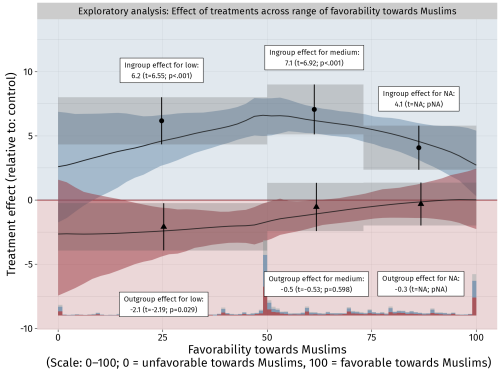
J.2 Treatment Effect Heterogeneity



(a) X Political Ideology



(b) X Military Assertiveness



(c) X Attitudes towards Muslims

Figure 24: **Low Treatment Effect Heterogeneity in our study**: Relative to CONTROL, estimates and 95% CI for INGROUP are shown in blue and for OUTGROUP in red.

J.3 Perceptions of Democracy and Culture

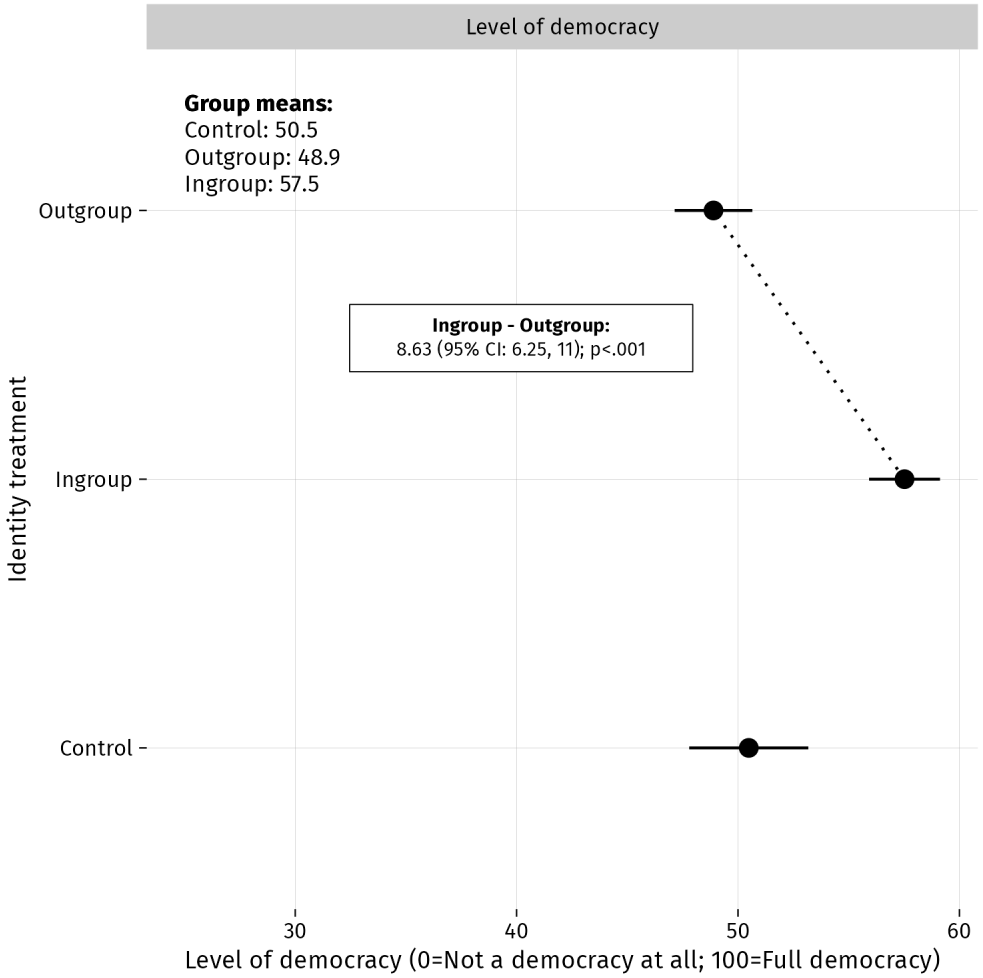


Figure 25: Effect of treatment on perceptions of Turkey's level of democracy.

Table 18: Perceptions of Culture and Democracy for Figures 25 and 26.

	Cultural Superior	Democracy
Identity Manipulation (Baseline: None)		
Ingroup	1.99*	7.04***
	(0.80)	(1.58)
Outgroup	−0.23	−1.59
	(0.80)	(1.57)
Gender (Baseline: Female)		
Male	1.75**	5.01***
	(0.59)	(1.37)
Other	−3.38	2.19
	(4.76)	(8.08)
Region (Baseline: Midwest)		
Northeast	1.55+	0.37
	(0.88)	(1.78)
South and Central	0.43	−2.42
	(0.75)	(1.50)
West	1.02	−3.90*
	(0.89)	(1.77)
Ideology (7pt L-R)	−1.56***	−0.69*
	(0.16)	(0.32)
Age	−0.17***	−0.16**
	(0.02)	(0.05)
Intercept	56.87***	56.49***
	(1.42)	(2.92)
Num.Obs.	5901	1863
R2	0.046	0.050

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

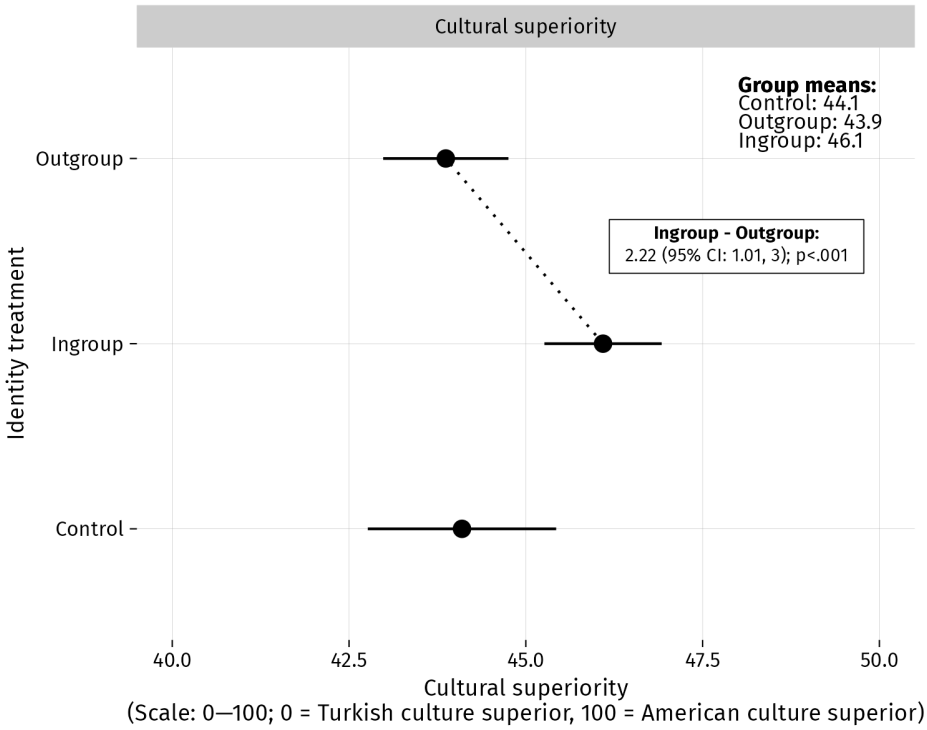


Figure 26: **Identity treatment affects beliefs about cultural superiority**.

J.4 Ingroup Love vs. Outgroup Hate

Which treatment (ingroup or outgroup) was more effective? In probing how and why our treatment worked to affect reputations, one of our key pre-registered exploratory questions was which of our two treatments—ingroup or outgroup—were more impactful? Our main study's design, which included a neutral control condition, allows us to answer this question. In fact, our results—for our manipulation check as well as the outcome measures of reputation and willingness to cooperate—all suggest that our INGROUP manipulation was particularly successful in impacting respondents' beliefs and judgments. Across the pre-registered manipulation check and H1 and H2, our INGROUP treatment was $14x$ more impactful than the OUTGROUP manipulation: the INGROUP treatment moved respondents by an average of 4.4 points (in the positive direction), compared to only 0.3 for the OUTGROUP treatment (in the opposite direction). This is consistent with original and longstanding interpretations of social identity theory in which ingroup and outgroup effects are distinct (not merely mirror images of each other) and in which the former are stronger (Brewer, 1999; for recent empirics in IR consistent with this, see Chu and Lee, 2024).

J.5 How does treatment work?

***How* does the treatment work to change group categorization?** In our conception of how our main study works, our identity manipulation treats group categorization, which affects a host of outcomes related to reputation and likely behavior. To investigate *how* the treatment is working, we first analyze post-treatment questions from the main study that elicit respondents' perceptions of the regime type of Turkey (from a full democracy to fully autocratic). We find evidence that respondents in the INGROUP treatment perceived Turkey as significantly more democratic (by about 8 points on our 0–100 scale) than in the OUTGROUP treatment.[28] This is suggestive of the importance of the language about the leadership of Turkey in the treatment and in accord with our survey results on the importance of regime type as a marker of international identity.[29] However, our results from the REGIME follow-up study confirm that regime type is not solely responsible for our main results. In that study, we fix Turkey's regime at "autocratic" while nudging group categorization via our treatment and still find effects of INGROUP (on reputation) approximately equal to those in the BASELINE follow-up in which regime type varies with identity.

To investigate whether the treatment works through activating racial or ethnic stereotypes, we consider respondents for whom the INGROUP/OUTGROUP treatments were most effective (relative to the CONTROL arm); in effect, treatment effect heterogeneity. Recall that the overall ATE of INGROUP on group categorization (manipulation check) was 5.2 points (while OUTGROUP was -1.1 points). As Figure 24c in Appendix J.2 shows, the INGROUP treatment moved respondents consistently across the entire range of pre-existing attitudes towards Muslims. Conversely, the

[28] 8.63 (95% CI: $6.25, 11, 24$); $p < .001$. See Figure 25 in the Appendix.

[29] Though this potential mechanism—even given the combination of results in our descriptive survey and experiment—should be treated with a healthy dose of caution given recent work demonstrating the issues inherent in the "intermediate outcome" approach to mechanisms. See Blackwell, Ma and Opacic (2024).

OUTGROUP treatment was only effective—and at a much lower magnitude—among those who held the most negative possible views towards Muslims (and wasn't effective at all for the bulk of respondents who were not at the extreme end of the distribution). We interpret this as suggestive evidence that the treatment is unlikely to be working through triggering ethnocentric biases since, presumably, ethnocentric biases would make it tougher for the INGROUP treatment to work among the most biased (which was not the case) and easiest for the OUTGROUP treatment to have a large effect (when it had a comparatively very small effect).

In fact, Figure 24 suggests little evidence of treatment effect heterogeneity more broadly. Across several important moderators—attitudes towards Muslims, military assertiveness and political ideology—the story is the same: the INGROUP treatment works consistently across the spectrum of respondents with different covariate profiles, while the OUTGROUP treatment is barely significantly different from zero. Finally, there are no moderating effects of attention, likely because of the extremely high levels of attentiveness—96% passed the post-treatment check—among the remainers.

J.6 Alternative Estimands

The pre-registered analysis in our Main Study estimates the quantity:

$$(\text{Back down} - \text{Stand firm}|\text{Ingroup}) - (\text{Back down} - \text{Stand firm}|\text{Outgroup}) \tag{1}$$

and answers the question, "Does the magnitude of the difference between back down and stand firm vary between those assigned to ingroup and those assigned to outgroup?" This set-up generates our pre-registered comparison, which was null. While the effect of backing down *did* differ between CONTROL and INGROUP ($\beta = 3.5$, $p = .027$) in our exploratory analysis, the analogous contrast between OUTGROUP and INGROUP was not statistically significant ($\beta = 1.0; p = .45$).

The models we use to estimate H2 also allow us insight into a related question: whether the effects of our group identity treatments were durable *even in the presence* of information about Turkey's past behavior. We can explore this by estimating the following quantities:

$$\text{Ingroup} - \text{Outgroup}|\text{Back down} \tag{2}$$

$$\text{Ingroup} - \text{Outgroup}|\text{Stand firm}. \tag{3}$$

Doing so reveals statistically significant differences in perceived resolve across the INGROUP and OUTGROUP conditions in both the "back down" (-3.80; $p < .001$) as well as the "stand firm" (-2.80; $p = .002$) arms.[30] Thus while we do not find support for H2, we do see that group identity "matters" even in the presence of a very strong behavioral signal of an actor's type.

Our within-subject design also allows us leverage on another potentially relevant set of esti-

[30]Note that the difference is in these two quantities is the pre-registered H2 contrast.

mands:

$$\hat{\Delta}_{T_2 \to T_3}\text{Ingroup}|\text{Back down} \tag{4}$$

$$\hat{\Delta}_{T_2 \to T_3}\text{Outgroup}|\text{Back down} \tag{5}$$

$$\hat{\Delta}_{T_2 \to T_3}\text{Control}|\text{Back down} \tag{6}$$

where $\hat{\Delta}_{T_2 \to T_3}$ is the average within-subject change from T_2 (post-identity treatment, but pre-behavior) to T_3 (post-identity treatment andpost-behavior). These estimands allow us to explore how identity-inflected beliefs about a state change in response to new information about an actor's behavior. The baseline here is the respondent's perception of Turkey's resolve at T_2 conditional on their exposure to one of the group identity conditions (control, ingroup, or outgroup).

Here, we learn that, within the "back down" arm of the main study, the average within-subject difference between T_2 and T_3—i.e., the within subject effect of backing down on reputations—was 11.48 ($p < .001$) for CONTROL, 10.02 ($p < .001$) for OUTGROUP, and 8.07 ($p < .001$) for INGROUP, and the ingroup/outgroup difference was 1.94 ($p = .04$). If we estimate the analogous quantities in the "stand firm" arm, we find that providing information about Turkey's resolved behavior increases perceptions for resolve but the effects are not statistically significant at conventional levels: 1.5 ($p = .39$) for the CONTROL, 2.14 for the OUTGROUP ($p = .19$), and INGROUP by 2.8 ($p = .08$). Given this pattern of results, one might tentatively conclude that negative information about an ingroup member's unresolved behavior does less reputational harm, albeit to a small degree, than similar information about outgroup members. We caution readers, however, that this $T_2 \to T_3$ change analysis is exploratory, not pre-registered, and could be biased by our reliance on the post-group identity baseline (Montgomery, Nyhan and Torres, 2018), but the results are consistent with the broader theory of a positive bias in perceptions for ingroup members.

K Follow-up Studies on Prolific

K.1 Overview

In April, we fielded 3 follow-up studies using the Prolific survey platform. The OSF pre-registration for those studies can be found at: osf.io/swxva.

Our follow-up studies followed a similar pattern to the original studies but differed in ways that allow us to speak to questions of *how* the treatment is working. Our main empirical goal was to investigate whether the direction and significance of the main ingroup/outgroup effect changes as we modified various elements of the treatment. We did this using "purposive variation" in the treatment paired with a sign-generalization test as suggested by Egami and Hartman (2023).

Study 3 (Baseline): Our original studies were fielded on Lucid. The first follow-up study, which we refer to as the **baseline** study, is a replication which establishes a baseline for our studies among respondents recruited on the Prolific platform. This study includes three treatment groups. Respondents are assigned to control with a probability .16 or one of the treatment arms with

probability .42 each. Arms 1-3 are simply the CONTROL, INGROUP and OUTGROUP arms from the initial study. We use the effect size and standard error of the main dependent variables in our initial study to guide our sample size selection of 566. It is nearly identical to our initial study except that it omits the "past behavior" manipulation and the measures that followed in the initial study.

Study 4 (Priming): The second follow-up study, which we refer to as the **priming** study, is a replication of the baseline study but omits the control group and a battery of pre-treatment measures of the DV that participants in all other studies complete. We omit these questions to investigate the potential bias that might result if we were priming respondents to think about politics in terms of identity. We use the unadjusted effect size and standard error of the main dependent variables in our initial study to guide our sample size selection of $N = 1,380$.[31]

Study 5 (Treatment): The third follow-up study, which we refer to as the **treatment** study, will be similar to our baseline study but will omit the CONTROL group and selectively fix portions of the treatment in order to learn about what is "doing the work" in our experimental manipulations. Again, we use the effect size and standard error of the main dependent variables in our initial study to guide our sample size selection of $N = 1,718$.

[31] "Unadjusted" here means that we do not use pre-treatment measures of DV or demographic covariates to improve statistical efficiency.

K.2 Demographics

Table 19: Demographic summary for follow up studies with 2022 Cooperative Election Study benchmark.

	Baseline (N = 570)	Baseline (no error) (N = 559)	No images (N = 995)	Priming (N = 458)	Regime (N = 1,056)	CCES Benchmark
Gender						
Male	283 (49.65%)	275 (49.19%)	483 (48.54%)	223 (48.69%)	513 (48.58%)	48.12%
Female	278 (48.77%)	277 (49.55%)	506 (50.85%)	232 (50.66%)	531 (50.28%)	51.07%
Other/Prefer not to say	9 (1.58%)	7 (1.25%)	6 (0.60%)	3 (0.66%)	12 (1.14%)	0.81%
Total	570	559	995	458	1,056	
Income						
Less than $30,000	98 (17.19%)	87 (15.56%)	121 (12.16%)	64 (13.97%)	166 (15.72%)	25.76%
Between $30,000 and $59,999	150 (26.32%)	145 (25.94%)	277 (27.84%)	126 (27.51%)	249 (23.58%)	25.24%
Between $60,000 and $149,999	252 (44.21%)	261 (46.69%)	423 (42.51%)	201 (43.89%)	496 (46.97%)	31.91%
More than $150,000	60 (10.53%)	53 (9.48%)	156 (15.68%)	61 (13.32%)	130 (12.31%)	8.26%
Prefer not to say	10 (1.75%)	13 (2.33%)	18 (1.81%)	6 (1.31%)	15 (1.42%)	8.82%
Total	570	559	995	458	1,056	
Race						
Asian	44 (7.72%)	44 (7.87%)	63 (6.33%)	34 (7.42%)	58 (5.49%)	4.77%
Black or African American	78 (13.68%)	76 (13.60%)	140 (14.07%)	63 (13.76%)	147 (13.92%)	13.22%
Hispanic	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	9.08%
Indigenous	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0.75%
White	393 (68.95%)	389 (69.59%)	715 (71.86%)	316 (69.00%)	770 (72.92%)	68.87%
Other	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3.31%
Total	570	559	995	458	1,056	
Age						
18–29	113 (19.82%)	115 (20.57%)	202 (20.30%)	86 (18.78%)	208 (19.70%)	21%
30–39	122 (21.40%)	115 (20.57%)	190 (19.10%)	104 (22.71%)	229 (21.69%)	15.61%
40–49	83 (14.56%)	86 (15.38%)	155 (15.58%)	72 (15.72%)	167 (15.81%)	14.78%
50–59	117 (20.53%)	123 (22.00%)	217 (21.81%)	90 (19.65%)	219 (20.74%)	16.41%
60–69	103 (18.07%)	88 (15.74%)	172 (17.29%)	74 (16.16%)	179 (16.95%)	18.89%
70+	32 (5.61%)	32 (5.72%)	59 (5.93%)	32 (6.99%)	54 (5.11%)	13.3%
Total	570	559	995	458	1,056	
Political Party						
Democratic	185 (32.46%)	175 (31.31%)	354 (35.58%)	138 (30.13%)	417 (39.49%)	32.33%
Independent/Other	218 (38.25%)	220 (39.36%)	255 (25.63%)	175 (38.21%)	250 (23.67%)	39.73%
Republican	167 (29.30%)	164 (29.34%)	386 (38.79%)	145 (31.66%)	389 (36.84%)	27.94%
Total	570	559	995	458	1,056	
Region						
Northeast	104 (18.25%)	102 (18.25%)	190 (19.10%)	83 (18.12%)	184 (17.42%)	17.41%
Midwest	116 (20.35%)	95 (16.99%)	188 (18.89%)	97 (21.18%)	181 (17.14%)	21.55%
South and Central	217 (38.07%)	226 (40.43%)	408 (41.01%)	201 (43.89%)	442 (41.86%)	38.23%
West	133 (23.33%)	136 (24.33%)	209 (21.01%)	77 (16.81%)	249 (23.58%)	22.81%
Total	570	559	995	458	1,056	
Education						
No high school diploma	5 (0.88%)	3 (0.54%)	3 (0.30%)	2 (0.44%)	4 (0.38%)	7.83%
High school graduate	68 (11.93%)	62 (11.09%)	84 (8.44%)	47 (10.26%)	95 (9.00%)	28.75%
Some college	110 (19.30%)	123 (22.00%)	152 (15.28%)	99 (21.62%)	180 (17.05%)	20.25%
2-year degree	50 (8.77%)	57 (10.20%)	85 (8.54%)	42 (9.17%)	96 (9.09%)	8.68%
4-year degree	218 (38.25%)	195 (34.88%)	425 (42.71%)	161 (35.15%)	447 (42.33%)	22.08%
Post-grad	119 (20.88%)	119 (21.29%)	246 (24.72%)	107 (23.36%)	234 (22.16%)	12.41%
Total	570	559	995	458	1,056	

K.3 Results

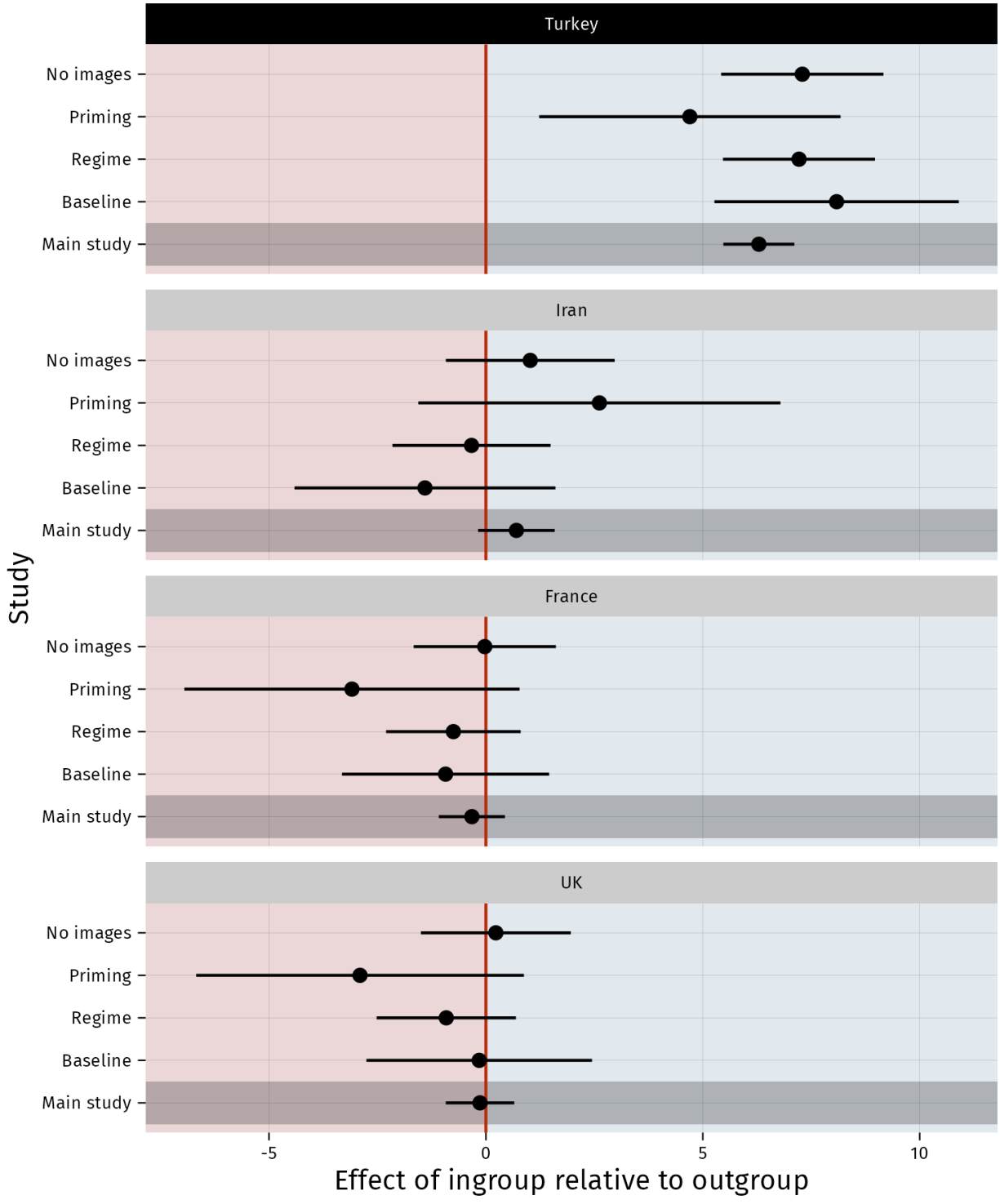


Figure 27: Manipulation checks for follow up studies

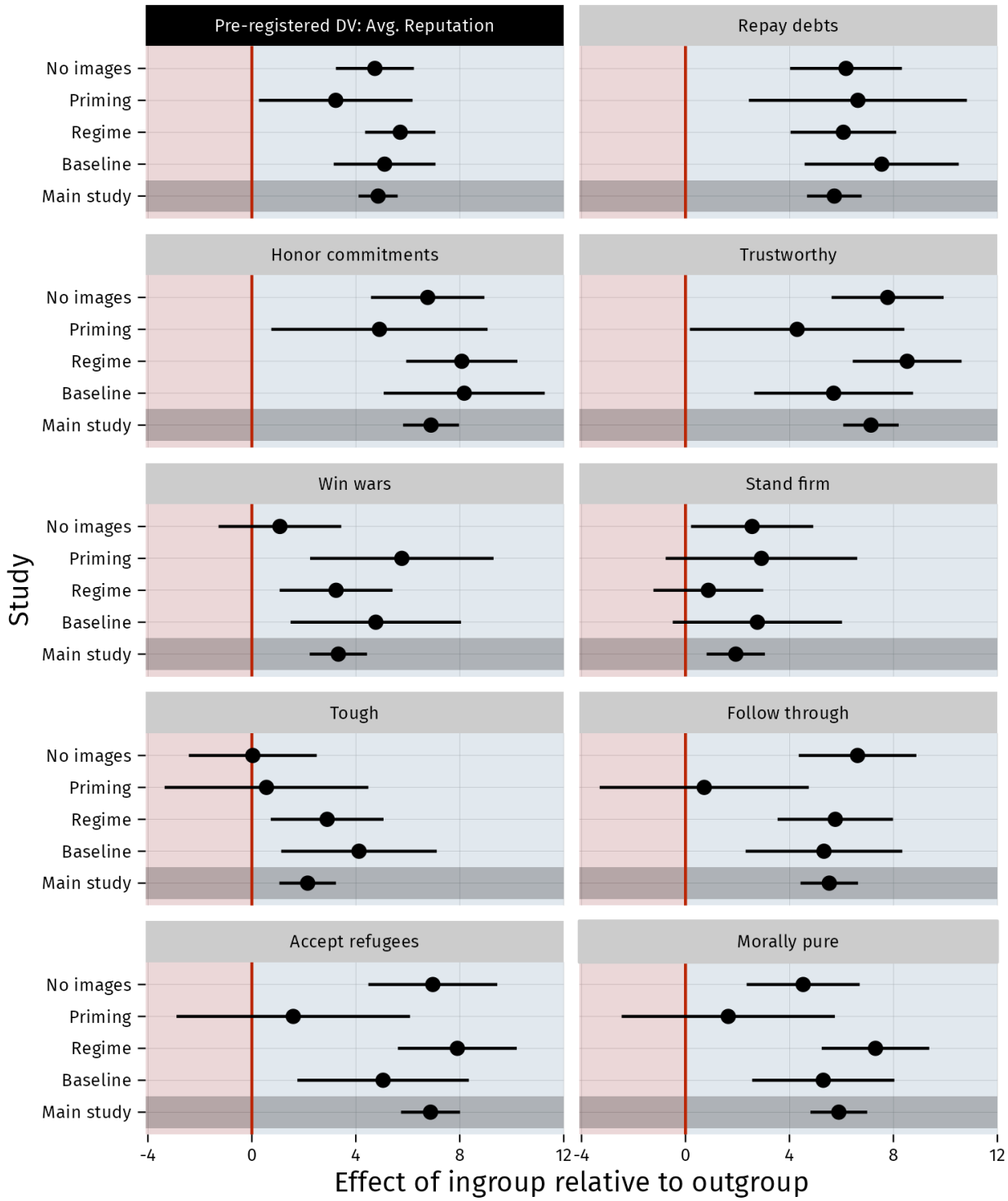


Figure 28: Test of H1 for follow up studies

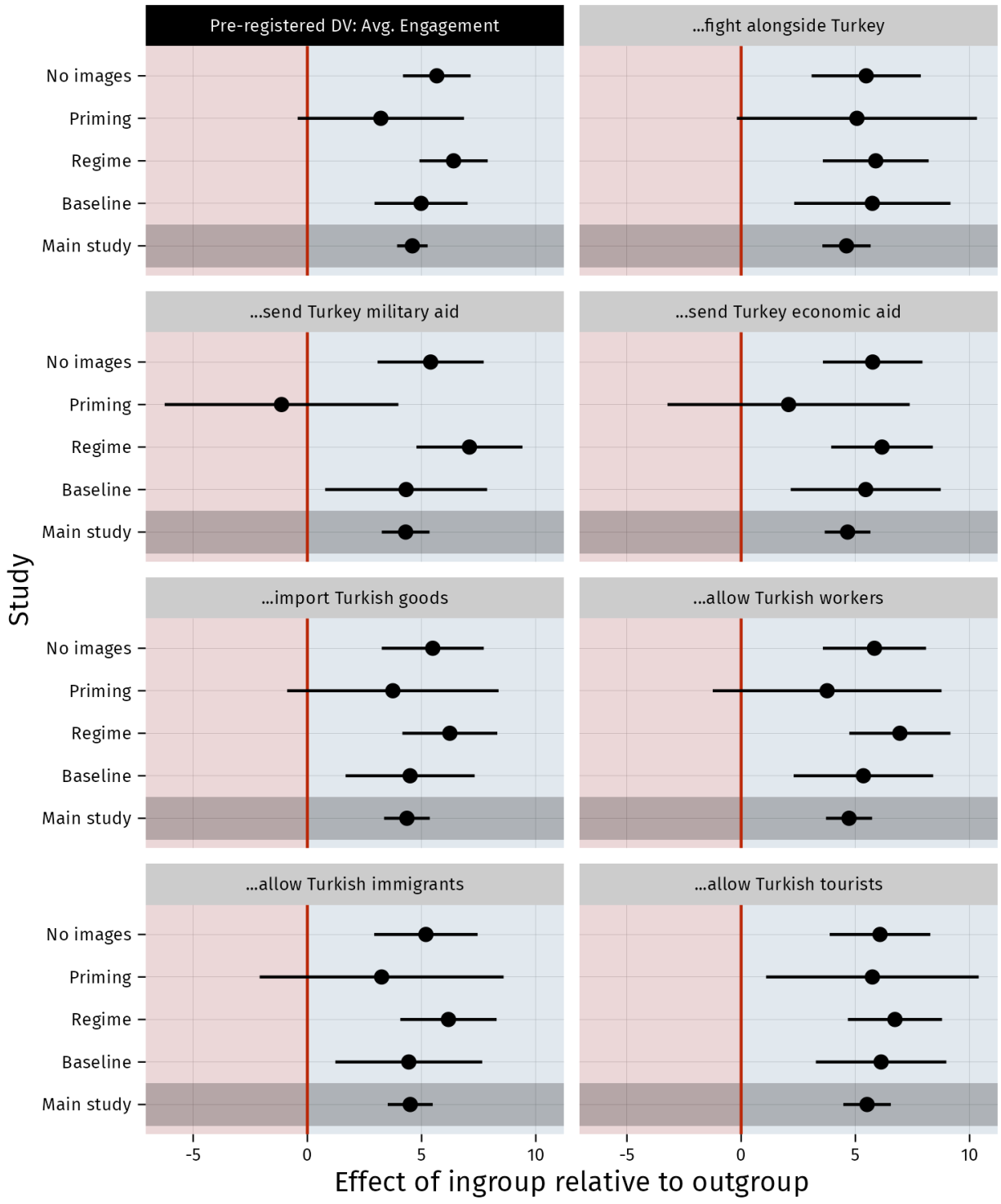


Figure 29: Test of H3 for follow up studies

K.4 Results tables

Table 20: Follow-up: *Baseline* results table for Figure 27.

	Turkey	France	Iran	UK	Turkey	France	Iran	UK	Turkey	France	Iran	UK
Identity Manipulation (Baseline: None)												
Ingroup	8.67***	−0.64	−1.25	1.33	10.92***	2.10	−0.97	0.61	6.41**	−3.39	3.62	−3.12
	(1.56)	(1.45)	(1.55)	(1.76)	(1.80)	(1.63)	(1.56)	(1.47)	(2.37)	(2.43)	(3.14)	(2.80)
Outgroup	2.28	1.01	0.22	−1.26	3.34+	2.02	−0.92	0.59	1.22	−0.01	−1.60	−0.16
	(1.52)	(1.42)	(1.52)	(1.72)	(1.76)	(1.59)	(1.52)	(1.43)	(2.31)	(2.37)	(3.07)	(2.73)
Gender (Baseline: Female)												
Male	−2.82*	0.58	1.27	−3.60**	−3.41**	−0.97	−0.61	−0.46	−2.23	2.14	−6.59**	3.01
	(1.10)	(1.03)	(1.10)	(1.25)	(1.27)	(1.15)	(1.10)	(1.04)	(1.68)	(1.72)	(2.22)	(1.98)
Other	−2.13	−1.06	0.56	2.71	−1.35	3.45	6.64	0.30	−2.91	−5.57	−1.21	0.82
	(4.47)	(4.17)	(4.45)	(5.05)	(5.16)	(4.67)	(4.47)	(4.21)	(6.79)	(6.97)	(9.00)	(8.03)
Region (Baseline: Midwest)												
Northeast	−0.96	−1.29	−0.78	−1.75	−2.26	−3.82*	−2.14	−1.58	0.35	1.24	−1.36	0.01
	(1.76)	(1.64)	(1.75)	(1.99)	(2.03)	(1.84)	(1.76)	(1.66)	(2.68)	(2.74)	(3.55)	(3.16)
South and Central	−3.03*	−1.72	−1.77	0.99	−3.74*	−4.05**	2.12	−4.59**	−2.32	0.61	−0.14	1.06
	(1.50)	(1.40)	(1.49)	(1.69)	(1.73)	(1.56)	(1.50)	(1.41)	(2.27)	(2.33)	(3.01)	(2.69)
West	−3.61*	−1.65	−0.92	−4.34*	−1.05	−3.32+	0.46	−2.93+	−6.17*	0.01	−9.15**	1.10
	(1.66)	(1.55)	(1.65)	(1.88)	(1.92)	(1.73)	(1.66)	(1.56)	(2.52)	(2.59)	(3.35)	(2.98)
Ideology (7pt L-R)	0.07	−0.27	0.20	−0.52	−0.06	−0.50	−0.31	−0.08	0.19	−0.04	−0.74	0.48
	(0.31)	(0.29)	(0.31)	(0.35)	(0.35)	(0.32)	(0.31)	(0.29)	(0.47)	(0.48)	(0.62)	(0.55)
Age	0.02	0.02	0.05	−0.04	0.06	0.05	−0.07*	0.02	−0.02	−0.01	−0.01	0.09
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.06)	(0.07)	(0.06)
Closeness Turkey T1	0.68***	0.09	0.01	0.05	0.25***	0.04	0.04	0.04	1.10***	0.13	0.06	−0.02
	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.09)	(0.10)	(0.13)	(0.11)
Closeness France T1	−0.02	0.60***	0.11	0.02	−0.23**	0.13+	−0.16*	−0.03	0.20+	1.07***	0.19	0.25+
	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.11)	(0.11)	(0.15)	(0.13)
Closeness UK T1	0.06	0.10	0.62***	−0.01	0.06	−0.06	0.09	0.02	0.05	0.27*	−0.11	1.21***
	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.11)	(0.11)	(0.14)	(0.12)
Closeness Iran T1	−0.08	−0.07	−0.02	0.62***	−0.15*	−0.10*	0.11*	−0.07	−0.01	−0.04	1.13***	0.03
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)	(0.10)	(0.09)
Warmth Turkey T1	0.02	−0.07	−0.03	−0.01	0.47***	−0.01	−0.06	0.01	−0.42***	−0.12	0.05	−0.07
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)	(0.10)	(0.09)
Warmth France T1	0.06	0.12*	−0.03	0.04	0.15*	0.66***	0.17**	0.08	−0.02	−0.42***	−0.09	−0.13
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.08)	(0.09)	(0.11)	(0.10)
Warmth Iran T1	0.12**	0.02	−0.01	0.14**	0.25***	0.04	0.79***	0.01	−0.01	0.00	−0.51***	−0.03
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.08)	(0.07)
Warmth UK T1	−0.06	0.02	0.15**	−0.11+	0.05	0.15**	−0.13*	0.78***	−0.17*	−0.11	−0.09	−0.47***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.08)	(0.08)	(0.11)	(0.10)
Intercept	7.92*	10.24**	8.65*	16.66***	3.87	10.29**	8.27*	13.95***	11.97*	10.19+	25.05***	3.36
	(3.55)	(3.31)	(3.53)	(4.01)	(4.10)	(3.71)	(3.55)	(3.34)	(5.39)	(5.53)	(7.15)	(6.37)
Num.Obs.	570	570	570	570	570	570	570	570	570	570	570	570
R2	0.567	0.662	0.618	0.602	0.620	0.687	0.751	0.698	0.432	0.520	0.407	0.486

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 21: Follow-up: *Priming* results table for Figure 27.

	Turkey	France	Iran	UK	Turkey	France	Iran	UK	Turkey	France	Iran	UK
Identity Manipulation (Baseline: Outgroup)												
Ingroup	4.70**	−3.09	−2.90	2.62	6.14**	−0.86	2.44	−1.76	3.26	−5.32*	2.79	−4.05
	(1.77)	(1.96)	(1.91)	(2.09)	(2.15)	(2.27)	(2.19)	(2.08)	(2.46)	(2.69)	(3.25)	(2.91)
Gender (Baseline: Female)												
Male	−1.42	1.41	0.56	−1.06	−1.05	0.26	−0.87	0.50	−1.79	2.57	−1.26	0.62
	(1.76)	(1.95)	(1.91)	(2.08)	(2.14)	(2.26)	(2.18)	(2.07)	(2.45)	(2.68)	(3.24)	(2.90)
Other	−3.43	−13.20	−25.08*	2.40	−6.72	−18.05	3.61	−42.49**	−0.14	−8.35	1.20	−7.67
	(11.02)	(12.19)	(11.91)	(13.00)	(13.36)	(14.14)	(13.63)	(12.95)	(15.29)	(16.72)	(20.22)	(18.10)
Region (Baseline: Midwest)												
Northeast	−0.91	−4.75	−0.86	−1.55	0.26	−6.01+	−1.27	1.42	−2.08	−3.48	−1.83	−3.13
	(2.81)	(3.11)	(3.04)	(3.31)	(3.41)	(3.61)	(3.47)	(3.30)	(3.90)	(4.26)	(5.16)	(4.62)
South and Central	4.56+	0.27	2.48	2.94	5.12+	1.98	3.32	3.07	4.00	−1.43	2.56	1.90
	(2.32)	(2.57)	(2.51)	(2.74)	(2.82)	(2.98)	(2.87)	(2.73)	(3.22)	(3.52)	(4.26)	(3.82)
West	0.48	−2.71	−1.53	2.30	6.89*	−2.21	8.01*	3.25	−5.93	−3.21	−3.41	−6.32
	(2.88)	(3.19)	(3.11)	(3.40)	(3.49)	(3.70)	(3.56)	(3.39)	(4.00)	(4.37)	(5.29)	(4.73)
Ideology (7pt L-R)	−1.01*	−1.75***	−1.23*	−1.51**	−1.29*	−3.14***	−2.46***	−1.49**	−0.72	−0.35	−0.56	−0.97
	(0.47)	(0.52)	(0.51)	(0.56)	(0.57)	(0.61)	(0.58)	(0.56)	(0.66)	(0.72)	(0.87)	(0.78)
Age	0.04	0.17**	0.13*	−0.03	−0.05	0.14+	−0.19**	0.14*	0.13	0.20*	0.13	0.11
	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	(0.10)	(0.09)
Intercept	46.07***	59.72***	64.31***	36.82***	56.06***	72.40***	46.21***	72.29***	36.08***	47.05***	27.42***	56.34***
	(3.76)	(4.15)	(4.06)	(4.43)	(4.55)	(4.82)	(4.64)	(4.41)	(5.21)	(5.70)	(6.89)	(6.17)
Num.Obs.	458	458	458	458	458	458	458	458	458	458	458	458
R2	0.044	0.052	0.040	0.028	0.043	0.073	0.074	0.050	0.034	0.027	0.011	0.020

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 22: Follow-up: *No images* results table for Figure 27.

	Turkey	France	Iran	UK	Turkey	France	Iran	UK	Turkey	France	Iran	UK
Identity Manipulation (Baseline: Outgroup)												
Ingroup	7.29***	−0.03	0.23	1.02	8.49***	0.45	−0.12	0.26	6.09***	−0.50	2.16	0.20
	(0.95)	(0.84)	(0.87)	(0.99)	(1.11)	(0.90)	(1.02)	(0.82)	(1.35)	(1.34)	(1.68)	(1.52)
Gender (Baseline: Female)												
Male	0.52	0.29	0.08	−0.70	−0.06	0.30	−0.83	−0.19	1.09	0.27	−0.57	0.35
	(0.96)	(0.85)	(0.89)	(1.01)	(1.13)	(0.92)	(1.04)	(0.84)	(1.38)	(1.37)	(1.71)	(1.54)
Other	3.21	−2.79	−0.50	0.74	0.67	−5.64	−0.39	−4.48	5.76	0.05	1.86	3.49
	(6.15)	(5.43)	(5.65)	(6.43)	(7.22)	(5.86)	(6.64)	(5.36)	(8.78)	(8.74)	(10.91)	(9.86)
Region (Baseline: Midwest)												
Northeast	−1.57	−0.40	−1.77	−0.61	−1.08	0.03	1.29	−2.17	−2.06	−0.82	−2.51	−1.36
	(1.54)	(1.36)	(1.42)	(1.61)	(1.81)	(1.47)	(1.66)	(1.34)	(2.20)	(2.19)	(2.74)	(2.47)
South and Central	−0.22	−0.36	−1.26	−0.43	−0.18	0.33	−1.34	−0.78	−0.26	−1.04	0.48	−1.74
	(1.32)	(1.17)	(1.21)	(1.38)	(1.55)	(1.26)	(1.43)	(1.15)	(1.89)	(1.88)	(2.34)	(2.12)
West	0.24	−0.30	−1.12	−0.89	−1.44	−0.79	0.12	−1.51	1.93	0.18	−1.90	−0.73
	(1.51)	(1.33)	(1.38)	(1.58)	(1.77)	(1.43)	(1.63)	(1.31)	(2.15)	(2.14)	(2.67)	(2.41)
Ideology (7pt L-R)	−0.40	−0.22	−0.06	−0.33	−0.49	−0.77**	−0.66*	−0.37+	−0.32	0.33	0.01	0.25
	(0.26)	(0.23)	(0.24)	(0.27)	(0.30)	(0.25)	(0.28)	(0.22)	(0.37)	(0.37)	(0.46)	(0.41)
Age	0.01	0.04	0.04	−0.03	0.00	0.06*	−0.05	0.07*	0.03	0.03	−0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.05)
Closeness Turkey T1	0.49***	0.04	0.00	0.00	0.13*	0.00	−0.11*	−0.09*	0.84***	0.07	0.11	0.10
	(0.05)	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)	(0.07)	(0.07)	(0.09)	(0.08)
Closeness France T1	0.07	0.55***	0.02	0.11+	−0.02	0.05	0.00	0.06	0.15+	1.06***	0.23*	−0.02
	(0.06)	(0.05)	(0.05)	(0.06)	(0.07)	(0.05)	(0.06)	(0.05)	(0.08)	(0.08)	(0.10)	(0.09)
Closeness UK T1	0.05	0.24***	0.72***	−0.08	−0.05	0.01	−0.08	0.01	0.15+	0.46***	−0.07	1.44***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.08)	(0.08)	(0.09)	(0.09)
Closeness Iran T1	0.07+	0.05	−0.01	0.62***	−0.13**	−0.05	0.08+	−0.05	0.27***	0.16**	1.15***	0.02
	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)	(0.06)	(0.06)	(0.07)	(0.06)
Warmth Turkey T1	0.11**	−0.02	0.01	0.02	0.57***	0.00	0.13**	0.06+	−0.36***	−0.05	−0.10	−0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.07)	(0.07)
Warmth France T1	−0.01	0.05	0.04	−0.09+	0.00	0.59***	−0.05	0.03	−0.02	−0.48***	−0.12	0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.07)	(0.07)
Warmth Iran T1	0.00	−0.04	−0.04	0.11***	0.15***	0.05	0.72***	−0.02	−0.15**	−0.13**	−0.49***	−0.06
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.06)	(0.05)
Warmth UK T1	−0.02	−0.04	0.01	0.01	0.08	0.15***	0.00	0.70***	−0.12*	−0.24***	0.02	−0.68***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.07)	(0.06)
Intercept	9.10**	10.14***	15.77***	12.62***	13.54***	13.77***	16.16***	20.71***	4.65	6.51	9.07+	10.84*
	(3.03)	(2.68)	(2.79)	(3.17)	(3.56)	(2.89)	(3.27)	(2.64)	(4.33)	(4.31)	(5.38)	(4.86)
Num.Obs.	995	995	995	995	995	995	995	995	995	995	995	995
R2	0.439	0.571	0.578	0.529	0.505	0.597	0.634	0.633	0.353	0.514	0.382	0.495

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 23: Follow-up: *Regime type* results table for Figure 27.

	Turkey	France	Iran	UK	Turkey	France	Iran	UK	Turkey	France	Iran	UK
Identity Manipulation (Baseline: Outgroup)												
Ingroup	7.22***	−0.75	−0.91	−0.33	8.17***	−0.86	−0.83	−1.24	6.26***	−0.64	0.17	−0.59
	(0.89)	(0.78)	(0.80)	(0.92)	(1.04)	(0.85)	(0.89)	(0.79)	(1.27)	(1.26)	(1.57)	(1.36)
Gender (Baseline: Female)												
Male	−0.07	0.52	−0.20	−1.29	0.40	1.01	−1.15	0.47	−0.54	0.02	−1.42	−0.88
	(0.90)	(0.79)	(0.81)	(0.93)	(1.06)	(0.86)	(0.90)	(0.80)	(1.29)	(1.28)	(1.59)	(1.38)
Other	−8.21+	−0.33	1.99	−9.05*	−7.95	0.45	−2.68	3.15	−8.48	−1.10	−15.41*	0.84
	(4.25)	(3.74)	(3.84)	(4.41)	(5.00)	(4.06)	(4.25)	(3.77)	(6.08)	(6.02)	(7.51)	(6.53)
Region (Baseline: Midwest)												
Northeast	−3.18*	−2.52+	0.00	−2.37	−3.02+	−1.63	−3.75*	0.29	−3.33	−3.42	−1.00	−0.28
	(1.51)	(1.33)	(1.36)	(1.56)	(1.78)	(1.44)	(1.51)	(1.34)	(2.16)	(2.14)	(2.67)	(2.32)
South and Central	−1.87	−1.73	−0.55	−0.31	−0.96	−1.77	−1.43	0.02	−2.78	−1.69	0.82	−1.12
	(1.28)	(1.13)	(1.16)	(1.33)	(1.51)	(1.23)	(1.28)	(1.14)	(1.84)	(1.82)	(2.27)	(1.97)
West	−2.73+	−2.32+	−1.85	−0.62	−3.26*	−2.27+	−3.01*	−0.89	−2.19	−2.36	1.76	−2.82
	(1.41)	(1.24)	(1.27)	(1.46)	(1.66)	(1.34)	(1.41)	(1.25)	(2.01)	(2.00)	(2.49)	(2.16)
Ideology (7pt L-R)	−0.15	−0.36+	−0.08	−0.07	0.06	−0.41+	0.15	0.09	−0.37	−0.32	−0.30	−0.26
	(0.24)	(0.21)	(0.21)	(0.25)	(0.28)	(0.23)	(0.24)	(0.21)	(0.34)	(0.34)	(0.42)	(0.36)
Age	−0.03	0.00	0.02	−0.09**	−0.02	−0.01	−0.13***	0.01	−0.04	0.01	−0.05	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Closeness Turkey T1	0.49***	−0.07	−0.01	−0.06	0.10+	−0.06	−0.13**	−0.02	0.88***	−0.07	0.00	0.00
	(0.05)	(0.04)	(0.04)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)	(0.07)	(0.07)	(0.09)	(0.07)
Closeness France T1	−0.01	0.62***	0.03	0.12*	−0.07	0.07	0.01	−0.07	0.05	1.16***	0.23*	0.13
	(0.06)	(0.05)	(0.05)	(0.06)	(0.07)	(0.05)	(0.06)	(0.05)	(0.08)	(0.08)	(0.10)	(0.09)
Closeness UK T1	0.08	0.15***	0.73***	−0.11*	−0.11+	0.01	−0.03	0.13**	0.27***	0.30***	−0.19*	1.33***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)	(0.09)	(0.08)
Closeness Iran T1	0.06	0.03	−0.02	0.65***	−0.11*	0.00	0.08*	−0.03	0.23***	0.06	1.21***	−0.01
	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)	(0.06)	(0.06)	(0.07)	(0.06)
Warmth Turkey T1	0.11**	0.02	−0.02	0.04	0.54***	0.03	0.08*	−0.02	−0.31***	0.02	0.01	−0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.07)	(0.06)
Warmth France T1	−0.04	0.08*	0.02	−0.11*	0.05	0.67***	0.00	0.10*	−0.12*	−0.51***	−0.21**	−0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.08)	(0.07)
Warmth Iran T1	0.04	−0.04	−0.02	0.11***	0.20***	−0.03	0.76***	0.00	−0.12*	−0.05	−0.54***	−0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.06)	(0.05)
Warmth UK T1	−0.03	0.02	0.07+	0.02	0.11*	0.14***	−0.02	0.69***	−0.16**	−0.10+	0.07	−0.54***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.07)	(0.06)
Intercept	14.22***	12.60***	13.31***	18.74***	13.17***	14.42***	15.22***	15.84***	15.27***	10.79**	22.25***	10.77*
	(2.80)	(2.46)	(2.53)	(2.90)	(3.29)	(2.67)	(2.79)	(2.48)	(4.00)	(3.96)	(4.94)	(4.30)
Num.Obs.	1056	1056	1056	1056	1056	1056	1056	1056	1056	1056	1056	1056
R2	0.449	0.613	0.622	0.538	0.523	0.639	0.673	0.637	0.371	0.511	0.404	0.531

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 24: Follow-up: *Baseline (No Error)* results table for Figure 27.

	Turkey	France	Iran	UK	Turkey	France	Iran	UK	Turkey	France	Iran	UK
Identity Manipulation (Baseline: None)												
Ingroup	5.23**	−0.57	−0.41	−0.31	8.41***	0.68	0.39	0.57	2.04	−1.82	−1.01	−1.38
	(1.76)	(1.56)	(1.65)	(1.94)	(1.97)	(1.56)	(1.95)	(1.62)	(2.50)	(2.55)	(3.27)	(2.81)
Outgroup	−2.86+	0.36	−0.25	1.10	−1.06	−0.46	1.59	−0.06	−4.66+	1.18	0.60	−0.45
	(1.73)	(1.53)	(1.62)	(1.91)	(1.94)	(1.53)	(1.92)	(1.59)	(2.46)	(2.51)	(3.22)	(2.76)
Gender (Baseline: Female)												
Male	1.83	0.19	1.08	−0.01	0.09	−0.81	−1.26	−0.43	3.57*	1.18	1.25	2.58
	(1.22)	(1.08)	(1.15)	(1.35)	(1.37)	(1.08)	(1.36)	(1.13)	(1.74)	(1.77)	(2.28)	(1.95)
Other	5.48	1.03	1.47	4.67	4.87	2.77	8.39	−0.51	6.09	−0.72	0.94	3.45
	(5.57)	(4.94)	(5.23)	(6.14)	(6.24)	(4.94)	(6.18)	(5.13)	(7.92)	(8.07)	(10.37)	(8.89)
Region (Baseline: Midwest)												
Northeast	−0.08	0.49	0.40	0.59	−1.96	−2.27	0.84	−1.75	1.79	3.26	0.33	2.54
	(2.05)	(1.82)	(1.92)	(2.26)	(2.29)	(1.82)	(2.27)	(1.89)	(2.91)	(2.97)	(3.82)	(3.27)
South and Central	−0.84	0.45	0.66	−0.63	−0.94	−2.39	−0.78	0.43	−0.73	3.29	−0.48	0.88
	(1.78)	(1.58)	(1.67)	(1.96)	(1.99)	(1.58)	(1.97)	(1.64)	(2.53)	(2.58)	(3.31)	(2.84)
West	0.44	0.67	−2.81	−0.44	−0.84	0.38	−1.02	−1.09	1.73	0.96	0.13	−4.54
	(1.93)	(1.71)	(1.81)	(2.12)	(2.16)	(1.71)	(2.14)	(1.77)	(2.74)	(2.79)	(3.59)	(3.07)
Ideology (7pt L-R)	0.15	0.00	−0.26	−1.07**	−0.24	−0.56+	−1.24**	−0.59+	0.54	0.56	−0.90	0.07
	(0.35)	(0.31)	(0.33)	(0.39)	(0.40)	(0.31)	(0.39)	(0.33)	(0.50)	(0.51)	(0.66)	(0.56)
Age	0.00	−0.02	0.04	0.02	0.06	0.06+	−0.02	0.12**	−0.06	−0.10+	0.06	−0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.08)	(0.06)
Closeness Turkey T1	0.55***	0.09	0.05	0.07	−0.04	−0.02	−0.17*	−0.07	1.14***	0.19+	0.31*	0.17
	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.06)	(0.08)	(0.06)	(0.10)	(0.10)	(0.13)	(0.11)
Closeness France T1	−0.08	0.41***	0.07	0.01	−0.13	0.05	−0.09	−0.05	−0.03	0.78***	0.12	0.18
	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.07)	(0.11)	(0.11)	(0.14)	(0.12)
Closeness UK T1	0.21**	0.25***	0.67***	−0.02	0.17*	0.01	0.08	0.18**	0.24*	0.48***	−0.13	1.16***
	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.06)	(0.08)	(0.07)	(0.10)	(0.11)	(0.14)	(0.12)
Closeness Iran T1	0.06	−0.06	−0.07	0.61***	0.07	−0.11*	0.19**	0.04	0.05	−0.01	1.02***	−0.17*
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.08)	(0.08)	(0.10)	(0.09)
Warmth Turkey T1	0.12*	−0.09+	−0.07	0.00	0.79***	0.01	0.22***	0.07	−0.55***	−0.19*	−0.22*	−0.21*
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.08)	(0.08)	(0.11)	(0.09)
Warmth France T1	0.00	0.31***	0.11*	−0.11+	0.12+	0.78***	0.03	0.18**	−0.12	−0.16+	−0.26*	0.04
	(0.06)	(0.05)	(0.06)	(0.07)	(0.07)	(0.05)	(0.07)	(0.05)	(0.08)	(0.09)	(0.11)	(0.10)
Warmth Iran T1	0.04	0.07+	0.08+	0.12*	−0.01	0.06	0.64***	−0.05	0.09	0.08	−0.40***	0.20**
	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.07)	(0.08)	(0.07)
Warmth UK T1	−0.06	−0.07	0.02	0.11+	−0.13*	0.07	−0.05	0.55***	0.00	−0.22**	0.27*	−0.52***
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.08)	(0.08)	(0.11)	(0.09)
Intercept	6.40+	4.92	9.91**	8.36*	6.18	8.59*	8.55*	10.95**	6.63	1.25	8.17	8.86
	(3.75)	(3.33)	(3.52)	(4.14)	(4.20)	(3.33)	(4.16)	(3.46)	(5.33)	(5.44)	(6.99)	(5.99)
Num.Obs.	559	559	559	559	559	559	559	559	559	559	559	559
R2	0.512	0.671	0.636	0.534	0.589	0.745	0.651	0.690	0.402	0.500	0.365	0.497

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 25: Follow-up: *Baseline* results table for Figure 28.

	Avg. Rep.	Stand Firm	Tough	Honor Commit	Trust	Pure	Win	Follow through	Repay debts	Accept refugees
Identity Manipulation (Baseline: Outgroup)										
Ingroup	2.95*	−1.30	−2.35	5.77**	3.66+	4.29*	0.28	1.16	6.87***	5.53*
	(1.20)	(2.18)	(2.04)	(1.95)	(1.90)	(1.95)	(2.05)	(1.98)	(1.91)	(2.28)
Outgroup	−0.54	−2.71	0.10	−0.18	−1.37	0.88	−2.14	−0.20	0.96	0.00
	(1.18)	(2.15)	(2.01)	(1.91)	(1.86)	(1.92)	(2.02)	(1.95)	(1.87)	(2.24)
Gender (Baseline: Female)										
Male	−1.33	−0.69	−1.20	−1.83	−1.43	−1.09	−1.44	0.35	−2.48+	−0.72
	(0.85)	(1.55)	(1.44)	(1.38)	(1.34)	(1.38)	(1.46)	(1.40)	(1.35)	(1.62)
Other	−1.26	0.69	1.11	4.30	−0.84	1.27	−4.27	1.93	3.09	1.77
	(3.45)	(6.26)	(5.86)	(5.58)	(5.46)	(5.60)	(5.91)	(5.69)	(5.48)	(6.54)
Region (Baseline: Midwest)										
Northeast	−2.68*	−0.77	−1.12	−1.17	−3.49	−2.32	−3.28	−6.61**	−1.32	−1.86
	(1.35)	(2.47)	(2.31)	(2.20)	(2.14)	(2.20)	(2.32)	(2.24)	(2.15)	(2.57)
South and Central	−0.87	0.07	−0.55	0.06	−1.06	−0.58	−1.47	−3.94*	0.11	0.62
	(1.16)	(2.11)	(1.97)	(1.88)	(1.83)	(1.88)	(1.98)	(1.91)	(1.84)	(2.20)
West	−0.92	2.72	0.96	−1.79	−2.95	−0.85	−1.16	−3.59+	−1.48	0.96
	(1.28)	(2.33)	(2.18)	(2.08)	(2.03)	(2.08)	(2.19)	(2.12)	(2.04)	(2.43)
Ideology (7pt L-R)	−0.30	−0.46	0.22	−0.62+	−0.28	0.05	−0.27	−0.89*	−0.87*	−0.79+
	(0.23)	(0.42)	(0.39)	(0.37)	(0.36)	(0.37)	(0.39)	(0.38)	(0.36)	(0.43)
Age	0.04	0.01	0.02	0.09*	0.02	0.02	−0.02	0.01	0.03	0.03
	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
DV at T_1	0.92***	0.60***	0.71***	0.74***	0.79***	0.79***	0.63***	0.70***	0.73***	0.75***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Intercept	6.09*	30.86***	19.09***	12.13**	13.27***	8.80*	25.59***	24.44***	13.46***	12.61**
	(2.43)	(4.18)	(3.79)	(3.68)	(3.47)	(3.54)	(3.63)	(3.78)	(3.55)	(4.14)
Num.Obs.	570	570	570	570	570	570	570	570	570	570
R2	0.705	0.353	0.469	0.533	0.570	0.571	0.410	0.470	0.543	0.483

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 26: Follow-up: *Priming* results table for Figure 28.

	Avg. Rep.	Stand Firm	Tough	Honor Commit	Trust	Pure	Win	Follow through	Repay debts	Accept refugees
Identity Manipulation (Baseline: Outgroup)										
Ingroup	3.23*	2.92	0.56	4.91*	4.30*	1.65	5.77**	0.72	6.63**	1.59
	(1.51)	(1.91)	(1.98)	(2.11)	(2.10)	(2.08)	(1.83)	(2.04)	(2.13)	(2.31)
Gender (Baseline: Female)										
Male	−0.03	1.17	−0.16	−1.86	−0.76	−2.05	0.72	0.44	0.02	2.22
	(1.50)	(1.90)	(1.98)	(2.11)	(2.10)	(2.07)	(1.82)	(2.03)	(2.12)	(2.30)
Other	−9.28	−6.22	−8.20	−10.26	−10.97	−24.95+	−10.35	0.01	−5.60	−7.01
	(9.40)	(11.88)	(12.34)	(13.15)	(13.09)	(12.95)	(11.36)	(12.68)	(13.27)	(14.38)
Region (Baseline: Midwest)										
Northeast	0.29	2.04	1.37	0.39	−1.28	−1.34	2.66	−0.92	−1.45	1.19
	(2.40)	(3.03)	(3.15)	(3.35)	(3.34)	(3.30)	(2.90)	(3.23)	(3.38)	(3.67)
South and Central	4.34*	6.38*	4.21	4.74+	4.18	2.84	6.98**	3.12	2.14	4.51
	(1.98)	(2.50)	(2.60)	(2.77)	(2.76)	(2.73)	(2.39)	(2.67)	(2.80)	(3.03)
West	3.95	2.13	4.27	3.70	6.27+	3.52	3.95	2.80	3.94	4.93
	(2.46)	(3.11)	(3.23)	(3.44)	(3.42)	(3.39)	(2.97)	(3.32)	(3.47)	(3.76)
Ideology (7pt L-R)	−0.21	−0.28	−0.95+	0.38	−0.76	0.06	−0.67	0.10	−0.08	0.33
	(0.40)	(0.51)	(0.53)	(0.56)	(0.56)	(0.56)	(0.49)	(0.54)	(0.57)	(0.62)
Age	0.02	0.15*	0.25***	−0.02	−0.06	−0.16*	0.05	0.02	0.03	−0.13+
	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)
Intercept	51.28***	58.73***	50.04***	54.01***	52.64***	52.15***	42.50***	54.67***	50.14***	46.65***
	(3.20)	(4.05)	(4.20)	(4.48)	(4.46)	(4.41)	(3.87)	(4.32)	(4.52)	(4.90)
Num.Obs.	458	458	458	458	458	458	458	458	458	458
R2	0.028	0.037	0.048	0.026	0.030	0.029	0.048	0.007	0.028	0.017

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

Table 27: Follow-up: *No images* results table for Figure 28.

	Avg. Rep.	Stand Firm	Tough	Honor Commit	Trust	Pure	Win	Follow through	Repay debts	Accept refugees
Identity Manipulation (Baseline: Outgroup)										
Ingroup	4.73***	2.56*	0.04	6.77***	7.78***	4.53***	1.08	6.62***	6.18***	6.96***
	(0.76)	(1.20)	(1.26)	(1.11)	(1.10)	(1.11)	(1.19)	(1.15)	(1.10)	(1.26)
Gender (Baseline: Female)										
Male	−0.51	−1.36	1.37	−0.06	−0.79	0.11	−0.64	0.12	0.00	−1.49
	(0.77)	(1.21)	(1.27)	(1.12)	(1.12)	(1.13)	(1.21)	(1.17)	(1.11)	(1.28)
Other	2.21	7.76	0.06	3.52	0.82	−4.29	−6.30	4.65	4.51	10.09
	(4.96)	(7.77)	(8.16)	(7.20)	(7.16)	(7.21)	(7.76)	(7.48)	(7.14)	(8.18)
Region (Baseline: Midwest)										
Northeast	−0.84	−2.20	0.06	−0.95	−2.67	−1.73	0.65	−1.57	−1.07	2.09
	(1.24)	(1.94)	(2.04)	(1.80)	(1.79)	(1.80)	(1.94)	(1.87)	(1.78)	(2.05)
South and Central	−0.01	−0.56	0.30	−0.01	−0.82	−0.61	2.10	0.17	0.70	1.63
	(1.07)	(1.67)	(1.75)	(1.55)	(1.54)	(1.55)	(1.67)	(1.61)	(1.53)	(1.76)
West	0.22	−1.52	0.60	−0.19	−1.35	0.99	0.79	0.02	0.07	2.49
	(1.21)	(1.90)	(1.99)	(1.76)	(1.75)	(1.76)	(1.89)	(1.82)	(1.74)	(2.00)
Ideology (7pt L-R)	−0.45*	−0.76*	−0.93**	−0.58+	−0.45	−0.21	−0.93**	−0.22	−0.36	−0.19
	(0.21)	(0.32)	(0.34)	(0.30)	(0.30)	(0.30)	(0.32)	(0.31)	(0.30)	(0.34)
Age	0.03	−0.01	0.08*	0.06+	0.04	0.05	−0.03	0.03	0.06+	0.00
	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
DV at T_1	0.85***	0.56***	0.61***	0.70***	0.74***	0.77***	0.65***	0.64***	0.72***	0.67***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
Intercept	7.81***	33.68***	23.55***	13.72***	11.47***	8.16**	23.89***	17.58***	11.19***	14.19***
	(2.13)	(3.16)	(3.21)	(2.91)	(2.83)	(2.84)	(2.94)	(3.03)	(2.87)	(3.18)
Num.Obs.	995	995	995	995	995	995	995	995	995	995
R2	0.601	0.305	0.329	0.465	0.516	0.523	0.383	0.393	0.490	0.416

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 28: Follow-up: *Treatment (No regime type)* results table for Figure 28.

	Avg. Rep.	Stand Firm	Tough	Honor Commit	Trust	Pure	Win	Follow through	Repay debts	Accept refugees
Identity Manipulation (Baseline: Outgroup)										
Ingroup	5.71***	0.88	2.90**	8.08***	8.53***	7.31***	3.24**	5.76***	6.07***	7.91***
	(0.69)	(1.07)	(1.10)	(1.09)	(1.06)	(1.05)	(1.10)	(1.13)	(1.03)	(1.17)
Gender (Baseline: Female)										
Male	0.11	−1.25	−1.45	−0.37	1.45	1.74	−0.24	0.50	0.02	0.75
	(0.69)	(1.08)	(1.11)	(1.10)	(1.07)	(1.06)	(1.11)	(1.14)	(1.04)	(1.18)
Other	−10.34**	−5.42	−15.29**	−15.95**	−8.84+	−11.89*	−11.04*	−8.28	−12.15*	−13.03*
	(3.29)	(5.11)	(5.26)	(5.19)	(5.08)	(5.00)	(5.24)	(5.42)	(4.91)	(5.59)
Region (Baseline: Midwest)										
Northeast	−0.71	0.14	−1.09	1.38	0.50	−0.64	−4.10*	−1.83	−1.18	−0.63
	(1.17)	(1.82)	(1.87)	(1.85)	(1.81)	(1.78)	(1.87)	(1.93)	(1.75)	(1.99)
South and Central	0.02	−1.86	−0.51	2.44	0.94	1.19	−0.28	−1.83	−0.22	0.13
	(0.99)	(1.54)	(1.59)	(1.57)	(1.53)	(1.51)	(1.58)	(1.63)	(1.48)	(1.69)
West	−1.16	−0.69	−1.20	1.23	−0.16	−1.66	−2.70	−2.04	−1.96	−0.93
	(1.09)	(1.69)	(1.75)	(1.72)	(1.68)	(1.66)	(1.74)	(1.80)	(1.63)	(1.85)
Ideology (7pt L-R)	−0.28	−0.58*	−0.30	−0.63*	−0.44	0.06	0.19	−0.28	−0.74**	−0.15
	(0.18)	(0.28)	(0.29)	(0.28)	(0.28)	(0.27)	(0.29)	(0.30)	(0.27)	(0.31)
Age	−0.03	−0.03	−0.04	−0.01	−0.06+	−0.06	−0.05	−0.04	0.01	−0.06
	(0.02)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
DV at T_1	0.86***	0.60***	0.67***	0.69***	0.70***	0.76***	0.64***	0.64***	0.73***	0.71***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Intercept	8.19***	32.96***	24.89***	13.53***	12.82***	8.15**	22.63***	22.02***	14.57***	13.13***
	(1.92)	(2.84)	(2.81)	(2.79)	(2.70)	(2.61)	(2.71)	(2.93)	(2.64)	(2.95)
Num.Obs.	1056	1056	1056	1056	1056	1056	1056	1056	1056	1056
R2	0.633	0.370	0.404	0.458	0.483	0.539	0.390	0.375	0.497	0.460

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

Table 29: Follow-up: *Baseline (no error)* results table for Figure 28.

	Avg. Rep.	Stand Firm	Tough	Honor Commit	Trust	Pure	Win	Follow through	Repay debts	Accept refugees
Identity Manipulation (Baseline: Outgroup)										
Ingroup	4.56***	−0.88	3.91*	9.54***	6.64**	5.54**	1.40	1.79	5.86**	7.98***
	(1.27)	(2.13)	(1.99)	(2.03)	(2.02)	(1.86)	(2.11)	(2.02)	(1.97)	(2.14)
Male	−0.10	2.05	1.09	0.55	−0.84	−0.43	−1.60	−0.35	−0.83	0.84
	(0.88)	(1.48)	(1.39)	(1.41)	(1.41)	(1.29)	(1.47)	(1.41)	(1.37)	(1.49)
Gender (Baseline: Female)										
Other	0.36	−3.60	0.27	8.76	−0.31	−2.00	−1.01	−0.97	1.56	0.75
	(4.00)	(6.74)	(6.30)	(6.43)	(6.39)	(5.88)	(6.66)	(6.39)	(6.22)	(6.77)
Northeast	2.01	2.36	1.57	2.22	4.84*	2.31	−0.13	1.94	5.99**	1.39
	(1.48)	(2.48)	(2.32)	(2.37)	(2.35)	(2.17)	(2.45)	(2.36)	(2.29)	(2.50)
Region (Baseline: Midwest)										
South and Central	0.44	3.11	2.00	0.45	1.18	0.60	−0.56	−0.64	3.57+	−1.21
	(1.28)	(2.15)	(2.01)	(2.05)	(2.04)	(1.88)	(2.12)	(2.04)	(1.99)	(2.17)
West	1.39	2.84	3.70+	−1.73	3.60	1.53	−1.30	−1.08	3.23	2.12
	(1.38)	(2.32)	(2.18)	(2.21)	(2.20)	(2.02)	(2.29)	(2.20)	(2.14)	(2.33)
Ideology (7pt L-R)	−0.14	−0.31	−0.16	−0.22	−0.21	0.02	−0.65	0.27	−0.51	−0.45
	(0.24)	(0.41)	(0.39)	(0.39)	(0.39)	(0.36)	(0.40)	(0.39)	(0.38)	(0.41)
Age	0.04	0.15**	−0.02	0.05	0.06	−0.04	0.01	0.00	0.04	0.00
	(0.03)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
DV at T_1	0.92***	0.62***	0.75***	0.70***	0.74***	0.79***	0.72***	0.71***	0.72***	0.70***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Intercept	3.20	19.91***	13.99***	11.15**	6.49+	8.75**	22.83***	17.98***	11.07**	12.41**
	(2.53)	(4.08)	(3.71)	(3.80)	(3.74)	(3.32)	(3.72)	(3.82)	(3.62)	(3.81)
Num.Obs.	559	559	559	559	559	559	559	559	559	559
R2	0.682	0.403	0.521	0.507	0.516	0.615	0.440	0.473	0.515	0.487

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Robust standard errors in parentheses.

Table 30: Follow-up: *Baseline (no error)* results table for Figure 29.

	Avg. Willingness	Goods	Tourists	Immigrants	Workers	Econ Aid	Mil Aid	Fight For
Identity Manipulation (Baseline: Outgroup)								
Ingroup	3.79***	5.50***	4.89**	3.25*	2.24	2.09	2.44	3.86*
	(0.90)	(1.57)	(1.51)	(1.47)	(1.37)	(1.45)	(1.57)	(1.53)
None	−1.69	−1.31	0.64	0.82	−2.35	−2.57	−2.02	−3.77+
	(1.14)	(1.99)	(1.91)	(1.87)	(1.75)	(1.83)	(1.99)	(1.95)
Gender (Baseline: Female)								
Male	−0.07	−2.24	−0.44	−0.15	−0.62	0.64	0.80	1.39
	(0.82)	(1.44)	(1.37)	(1.34)	(1.26)	(1.32)	(1.43)	(1.40)
Other	4.05	5.56	6.13	1.77	−0.10	4.87	10.83+	6.51
	(3.33)	(5.83)	(5.57)	(5.45)	(5.09)	(5.35)	(5.81)	(5.71)
Region (Baseline: Midwest)								
Northeast	−2.65*	−1.82	−4.40*	−3.93+	−2.48	−2.63	−0.27	−0.81
	(1.31)	(2.29)	(2.19)	(2.14)	(2.01)	(2.12)	(2.29)	(2.24)
South and Central	−3.05**	−4.41*	−3.88*	−2.90	−2.16	−2.06	−3.16	−1.24
	(1.12)	(1.96)	(1.87)	(1.83)	(1.71)	(1.81)	(1.95)	(1.91)
West	−3.75**	−2.16	−4.21*	−3.14	−3.13+	−5.69**	−3.77+	−1.71
	(1.24)	(2.17)	(2.08)	(2.03)	(1.90)	(2.00)	(2.16)	(2.12)
Ideology (7pt L-R)	0.29	−0.06	−1.05**	−0.37	−0.96**	0.30	0.91*	0.46
	(0.24)	(0.42)	(0.39)	(0.40)	(0.37)	(0.37)	(0.39)	(0.38)
Age	0.01	0.03	0.01	0.08+	−0.07+	−0.03	0.05	−0.02
	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
DV at T_1	0.95***	0.86***	0.82***	0.89***	0.88***	0.88***	0.85***	0.86***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Intercept	1.98	9.84*	17.56***	5.90	17.21***	5.14+	−0.07	1.34
	(2.30)	(4.12)	(3.87)	(3.61)	(3.41)	(3.12)	(3.30)	(3.31)
Num.Obs.	570	570	570	570	570	570	570	570
R2	0.832	0.670	0.695	0.767	0.783	0.735	0.688	0.692

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

Table 31: Follow-up: *Treatment (No images)* results table for Figure 29.

	Avg. Willingness	Goods	Tourists	Immigrants	Workers	Econ Aid	Mil Aid	Fight For
Identity Manipulation (Baseline: Outgroup)								
Ingroup	5.67***	5.49***	6.08***	5.19***	5.84***	5.76***	5.40***	5.47***
	(0.75)	(1.13)	(1.11)	(1.15)	(1.14)	(1.11)	(1.17)	(1.22)
Gender (Baseline: Female)								
Male	0.44	1.40	0.88	−0.52	−0.11	0.11	0.95	2.22+
	(0.76)	(1.15)	(1.12)	(1.16)	(1.16)	(1.13)	(1.19)	(1.24)
Other	4.34	0.48	7.69	6.12	5.29	3.81	3.56	9.50
	(4.89)	(7.38)	(7.21)	(7.45)	(7.43)	(7.22)	(7.63)	(7.94)
Region (Baseline: Midwest)								
Northeast	−1.99	−2.04	−2.49	−3.33+	−2.34	−0.81	−0.95	−0.53
	(1.22)	(1.84)	(1.80)	(1.86)	(1.86)	(1.81)	(1.91)	(1.99)
South and Central	−1.91+	−0.98	0.01	−1.33	−3.45*	−2.64+	−2.21	−1.80
	(1.05)	(1.58)	(1.55)	(1.60)	(1.59)	(1.56)	(1.64)	(1.71)
West	−3.13**	−1.96	−0.51	−0.82	−3.41+	−4.44*	−4.70*	−5.43**
	(1.19)	(1.80)	(1.76)	(1.82)	(1.81)	(1.76)	(1.86)	(1.94)
Ideology (7pt L-R)	0.03	−0.35	−0.70*	−0.37	−0.59+	−0.03	0.07	−0.03
	(0.21)	(0.32)	(0.31)	(0.32)	(0.32)	(0.30)	(0.32)	(0.33)
Age	0.05*	0.05	0.08*	0.09*	0.04	0.01	0.08*	−0.02
	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
DV at T_1	0.92***	0.78***	0.82***	0.83***	0.82***	0.83***	0.81***	0.79***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Intercept	0.37	12.46***	8.16**	6.61*	11.51***	5.94*	1.72	5.51+
	(2.10)	(3.18)	(3.13)	(3.03)	(3.09)	(2.68)	(2.80)	(2.97)
Num.Obs.	995	995	995	995	995	995	995	995
R2	0.741	0.578	0.626	0.671	0.658	0.670	0.638	0.599

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

Table 32: Follow-up: *Baseline (no error)* results table for Figure 29.

	Avg. Willingness	Goods	Tourists	Immigrants	Workers	Econ Aid	Mil Aid	Fight For
Identity Manipulation (Baseline: Outgroup)								
Ingroup	4.98***	4.50**	6.13***	4.44**	5.35***	5.45***	4.32*	5.74***
	(0.99)	(1.46)	(1.49)	(1.60)	(1.51)	(1.60)	(1.74)	(1.64)
None	0.07	1.11	−0.37	2.17	1.07	−1.38	−1.79	0.01
	(1.28)	(1.89)	(1.93)	(2.07)	(1.96)	(2.08)	(2.25)	(2.12)
Gender (Baseline: Female)								
Male	0.67	2.27+	0.78	1.63	0.23	−0.96	1.18	0.36
	(0.91)	(1.33)	(1.36)	(1.46)	(1.38)	(1.47)	(1.60)	(1.50)
Other	0.73	7.99	4.73	4.32	5.49	−1.62	−7.69	−3.15
	(4.12)	(6.08)	(6.19)	(6.68)	(6.31)	(6.68)	(7.22)	(6.82)
Region (Baseline: Midwest)								
Northeast	2.09	1.92	0.18	−0.44	2.54	3.39	2.83	3.48
	(1.52)	(2.23)	(2.28)	(2.45)	(2.31)	(2.46)	(2.66)	(2.51)
South and Central	0.83	0.92	−1.22	0.66	2.61	0.01	2.05	1.11
	(1.31)	(1.94)	(1.97)	(2.12)	(2.01)	(2.13)	(2.30)	(2.18)
West	1.08	2.61	−0.39	1.68	2.29	−2.05	4.30+	−0.72
	(1.42)	(2.09)	(2.13)	(2.29)	(2.16)	(2.29)	(2.49)	(2.35)
Ideology (7pt L-R)	−0.33	−1.34***	−0.83*	−0.96*	−0.71+	0.01	−0.61	0.07
	(0.26)	(0.39)	(0.39)	(0.42)	(0.40)	(0.41)	(0.44)	(0.42)
Age	−0.03	0.04	0.00	−0.05	0.00	−0.05	−0.04	−0.09+
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
DV at T_1	0.90***	0.80***	0.79***	0.82***	0.84***	0.83***	0.81***	0.82***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Intercept	4.30+	13.34***	15.54***	13.75***	8.70*	7.75*	5.53	5.05
	(2.50)	(3.75)	(3.80)	(3.72)	(3.69)	(3.51)	(3.64)	(3.55)
Num.Obs.	559	559	559	559	559	559	559	559
R2	0.774	0.649	0.644	0.704	0.720	0.664	0.624	0.659

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

Table 33: Follow-up: *Treatment (No regime type)* results table for Figure 29.

	Avg. Willingness	Goods	Tourists	Immigrants	Workers	Econ Aid	Mil Aid	Fight For
Identity Manipulation (Baseline: Outgroup)								
Ingroup	6.40***	6.24***	6.73***	6.18***	6.95***	6.17***	7.10***	5.89***
	(0.75)	(1.05)	(1.05)	(1.07)	(1.13)	(1.13)	(1.15)	(1.16)
Gender (Baseline: Female)								
Male	2.02**	3.00**	2.38*	2.43*	1.32	0.10	2.69*	2.59*
	(0.76)	(1.06)	(1.06)	(1.08)	(1.15)	(1.14)	(1.17)	(1.17)
Other	−0.45	1.59	−2.07	1.26	−1.00	−3.63	−6.42	5.06
	(3.60)	(5.02)	(4.99)	(5.10)	(5.42)	(5.39)	(5.50)	(5.53)
Region (Baseline: Midwest)								
Northeast	−0.63	−2.19	2.14	0.00	2.14	−2.12	−1.19	−2.98
	(1.28)	(1.79)	(1.78)	(1.82)	(1.93)	(1.92)	(1.96)	(1.97)
South and Central	−0.46	−2.45	−0.21	0.55	0.69	−0.45	−0.26	−0.50
	(1.08)	(1.51)	(1.51)	(1.54)	(1.63)	(1.63)	(1.66)	(1.67)
West	−1.75	−2.35	−1.53	−1.33	0.77	−2.40	−2.21	−3.42+
	(1.19)	(1.67)	(1.66)	(1.69)	(1.80)	(1.79)	(1.82)	(1.83)
Ideology (7pt L-R)	−0.34	−1.03***	−1.12***	−1.09***	−1.25***	−0.09	0.47	0.60+
	(0.21)	(0.28)	(0.29)	(0.29)	(0.31)	(0.30)	(0.30)	(0.30)
Age	0.00	0.04	0.02	0.01	0.01	−0.07*	−0.05	−0.01
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
DV at T_1	0.83***	0.76***	0.78***	0.83***	0.78***	0.75***	0.76***	0.74***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Intercept	5.53**	16.05***	13.78***	9.20**	13.41***	10.02***	2.98	1.96
	(2.14)	(2.93)	(2.94)	(2.84)	(3.08)	(2.75)	(2.73)	(2.79)
Num.Obs.	1056	1056	1056	1056	1056	1056	1056	1056
R2	0.690	0.595	0.632	0.694	0.630	0.589	0.582	0.565

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

Table 34: Follow-up: *Treatment (Priming)* results table for Figure 29.

	Avg. Willingness	Goods	Tourists	Immigrants	Workers	Econ Aid	Mil Aid	Fight For
Identity Manipulation (Baseline: Outgroup)								
Ingroup	3.22+	3.74	5.75*	3.25	3.76	2.08	−1.13	5.07+
	(1.85)	(2.35)	(2.41)	(2.70)	(2.54)	(2.67)	(2.61)	(2.65)
Gender (Baseline: Female)								
Male	3.22+	3.20	4.18+	2.84	2.38	0.11	5.22*	4.60+
	(1.85)	(2.34)	(2.40)	(2.69)	(2.54)	(2.66)	(2.60)	(2.64)
Other	−11.05	−3.08	11.51	0.37	0.05	−37.07*	−35.90*	−13.23
	(11.54)	(14.63)	(14.98)	(16.81)	(15.84)	(16.62)	(16.27)	(16.49)
Region (Baseline: Midwest)								
Northeast	−1.80	−3.89	−3.50	−2.59	1.54	−0.05	−2.05	−2.06
	(2.94)	(3.73)	(3.82)	(4.29)	(4.04)	(4.24)	(4.15)	(4.20)
South and Central	4.27+	0.86	−2.19	0.61	3.73	10.86**	7.78*	8.21*
	(2.43)	(3.08)	(3.16)	(3.54)	(3.34)	(3.50)	(3.43)	(3.48)
West	−0.91	−2.52	−1.85	−7.32+	−1.44	3.31	0.44	3.02
	(3.02)	(3.83)	(3.92)	(4.40)	(4.14)	(4.34)	(4.25)	(4.31)
Ideology (7pt L-R)	−4.11***	−4.18***	−4.67***	−6.44***	−6.24***	−2.88***	−2.22**	−2.13**
	(0.49)	(0.63)	(0.64)	(0.72)	(0.68)	(0.71)	(0.70)	(0.71)
Age	−0.02	−0.01	0.02	0.00	−0.06	−0.07	0.07	−0.10
	(0.06)	(0.07)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
Intercept	67.27***	86.89***	88.16***	84.61***	85.86***	44.61***	37.71***	43.06***
	(3.93)	(4.99)	(5.11)	(5.73)	(5.40)	(5.66)	(5.54)	(5.62)
Num.Obs.	458	458	458	458	458	458	458	458
R2	0.159	0.105	0.128	0.173	0.181	0.072	0.059	0.058

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: Robust standard errors in parentheses.

L Programming Mistake in Initial Study

In the main study fielded on Lucid, subjects in the OUTGROUP and ingroup conditions were meant to view the following photos and captions:

Figure 30: Intended programming for main study - photo of Erdoğan.

Due to a programming error, respondents in the INGROUP condition also were shown a caption for the photo in small print that read: "Pictured: President Erdoğan, the leader of Turkey meeting with President Biden and President Trump. Photo credit: CNN/Emin Sansar/Anadolu Agency/Getty Images."

Mentioning Erdoğan in conjunction with two U.S. Presidents fits the intention of the treatment, which was to induce respondents to see Turkey as part of the U.S. ingroup. Thus, it is probably best thought of as an additional element to the treatment, though one we did not anticipate.

However, to eliminate any concerns that this small bit of text was responsible for any of our results, we fielded an extra follow-up study on Prolific in Spring 2025. That follow-up, with the error included on purpose, can be compared to the BASELINE pre-registered follow-up study (with no error) to show that there are no substantive differences induced by the actual text.