When Contact Causes a Public-Private Divide: How Interventions Affect Personal Beliefs, Public Expression, and Social Change

Laura Paler*

January 8, 2026

Abstract

When do interventions that change personal beliefs succeed—or fail—to influence public expression? Public expression is central to social change, yet few studies consider that interventions altering personal beliefs may fail to affect public expression when social costs are high, yielding preference falsification and slowed progress. Focusing on intergroup contact, I develop a framework showing how interventions operating through personal beliefs versus perceived norms produce asymmetric effects on private versus public outcomes. I test parasocial contact effects on white American attitudes toward Black Americans using a novel virtual enumerator experiment to measure private attitudes versus public expression. The intervention produced private attitude change for individuals facing high social costs, yet public expression of those attitudes and willingness to engage in public action were substantially attenuated. By clarifying how interventions can induce, deepen, or reverse preference falsification, this paper reveals their potential—and limits—to achieve social change.

Word Count: 11,559

*Provost Associate Professor, Department of Government, School of Public Affairs, American University. I would like to thank Nicole Dubrow, Ellen Kamalyan, Pauline Sow, and Karl VonZaburn for their research assistance, and Kyle Johnsen for editing the video treatments. I am especially grateful to Kyle Gray for his substantial contributions to the early stages of this project, including data collection and project design. I also thank Jake Bowers, Ruth Ditlmann, Don Green, Guy Grossman, Chryl Laird, Lucy Martin, Beatrice Montano, Noory Okthariza, Betsy Levy Paluck, Egon Tripodi, Rebecca Wolfe, and participants at EGAP 2024 and the 2025 Randomized Intergroup Interventions Workshop for their comments. This study was pre-registered on OSF (https://osf.io/kcgbx/) and approved by American University's IRB (no: 2023-354).

1 Introduction

When do interventions that change personal beliefs succeed—or fail—to change public expression? More broadly, under what conditions do such interventions generate change across both private and public domains, and when do they instead cause private and public outcomes to diverge?

Interventions that aim to promote social change often seek to shift personal beliefs about socially and politically consequential issues and groups. Yet such interventions also frequently aim to change public expression—what individuals are willing to say or do in the presence of others. Public expression is central to processes of social change that involve persuasion, social signaling, norm change, and social cascades (Granovetter, 1978; Lohmann, 1994; de Mesquita, Shadmehr et al., 2023; Bursztyn, González and Yanagizawa-Drott, 2020; Bursztyn, Egorov and Fiorin, 2020). Yet, precisely because it is visible to others, public expression can entail social costs. A key—but often overlooked—implication follows: interventions that primarily shift personal beliefs may fail to generate public expression when social costs are high or when interventions fail to reduce them directly. Interventions could thus actually induce or deepen preference falsification—a divergence between private and public outcomes (Kuran, 1995)—slowing or stalling the processes of social change they often hope to encourage.

This paper examines these dynamics with a focus on intergroup contact, one of the most promising and widely studied approaches to prejudice reduction (Paluck et al., 2021). Since the foundational work by Allport (1954), contact has been studied in many forms, including face-to-face (Scacco and Warren, 2018; Mousa, 2020), through media (Paluck, 2009; Alrababa'h et al., 2021), in naturally occurring settings (Weiss, 2021), and coupled with perspective-taking (Broockman and Kalla, 2016). Regardless of form, contact is widely viewed as having the potential change personal beliefs and prejudice toward outgroups by increasing empathy, reducing stereotypes, or lowering anxiety (Allport, 1954; Pettigrew and Tropp, 2011).

Yet contact interventions also frequently aim to influence public expression—what individuals are willing to do or say in the presence of other ingroup members. This is reflected in outcomes including visible intergroup cooperation (Mousa, 2020), public signaling within networks (Larson and Lewis, 2025; Siddique, Vlassopoulos and Zenou, 2024), and participation in collective action (Hässler et al., 2020). Moreover, because many studies rely on enumerator-administered surveys

(Paluck et al., 2021), even reported attitudes may constitute public expression (Blair, Coppock and Moor, 2020). Yet the conditions under which contact produces asymmetric effects on private and public outcomes, and the consequences of those effects for preference falsification, remain largely unexamined.

I develop a framework showing how contact and other interventions can produce asymmetric effects on private and public outcomes depending on their operative mechanisms, with implications for inducing, deepening, or reversing preference falsification. The framework distinguishes interventions by their operative mechanisms: those that primarily shift personal beliefs versus those that primarily shift perceived social norms. When an intervention primarily shifts personal beliefs without changing perceived norms, it should have larger effects on private attitudes and behaviors than on public ones, particularly for individuals facing high social costs. Conversely, when an intervention primarily shifts perceived norms without altering personal beliefs, it can induce public expression while having little effect on private attitudes or behaviors. The framework further shows how such asymmetric intervention effects can induce, deepen, or reduce preference falsification depending on baseline levels of divergence between private and public outcomes. By systematically linking mechanisms of change to patterns of preference falsification, the framework clarifies when interventions may or may not drive social change through public expression, and when social change may instead be more likely to occur through private behaviors.

I then test a core prediction of the framework: interventions that primarily shift personal beliefs without changing perceived norms should produce larger effects on private than public outcomes, particularly among individuals facing high social costs. I test these predictions use a large-scale online media experiment that examines the effects of parasocial contact on white American attitudes towards Black Americans. Parasocial contact—positive, one-sided exposure to outgroup members through media (Schiappa, Gregg and Hewes, 2005, 2006)—provides an ideal test as a common form of outgroup exposure with demonstrated social and political consequences (Goldman, 2012; Alrababa'h et al., 2021; Kim and Patterson Jr, 2025). Moreover, it is theorized to operate primarily by fostering affective ties and shifting personal beliefs, rather than by changing perceived social norms through depictions of ingroup-outgroup interaction.

I implement this experiment using curated content from the reality television show *Married at First Sight* (MAFS), randomly assigning respondents to watch approximately 25 minutes featuring

either a successful Black couple or a comparable white couple navigating married life. Reality television offers a particularly strong setting for parasocial contact insofar as it features emotionally engaging and positive portrayals of outgroup members that challenge persistent negative stereotypes (Ramasubramanian, 2015). Additionally, *MAFS*'s production emphasizes couples' personal stories, consistent with the kind of personal belief-change interventions that should generate asymmetric effects on private and public outcomes (Banas, Bessarabova and Massey, 2020).

Testing these predictions requires distinguishing treatment effects on private attitudes from effects on public expression—a measurement challenge for research on preference falsification and asymmetric intervention effects. I address this by developing a "virtual enumerator" experiment in which participants are cross-randomized to report attitudes either via a standard online survey (*private* condition) or to a virtual ingroup interviewer (*public* condition).[1] This design offers a methodological advance: by manipulating only the perceived visibility of responses while holding all other features constant, it enables a clean test of whether interventions affect private attitudes, public expression, or both. The approach leverages well-documented enumerator and social desirability effects (Adida et al., 2016; Blair, Coppock and Moor, 2020) but applies them experimentally within a single survey instrument, allowing individual-level measurement across multiple outcomes—an advantage over existing indirect methods like list experiments (Blair, Coppock and Moor, 2020). This provides a direct test of the pre-registered prediction that parasocial contact should produce larger effects on privately versus publicly reported attitudes, particularly among individuals facing high social costs—those embedded in homogeneous ingroup networks who perceive ingroup norms as condoning prejudice. I then extend the analysis to behavioral intentions and actual behavior. While the virtual enumerator holds outcome features constant, behaviors vary along multiple dimensions—including perceived benefits and effort costs—providing a complementary but noisier test of whether contact primarily shifting personal beliefs changes private but not public behavior.

The results reveal a pattern consistent with predictions. The treatment fostered affective ties

[1]This virtual enumerator design was adapted from Krysan and Couper (2003); a similar version has been used in Paler and Khan (2025) to study the effects of a campaign ad on public–private candidate preferences.

with the outgroup couple and reduced negative stereotypes, with little corresponding change in perceived ingroup norms—conditions under which the framework predicts asymmetric effects. The virtual enumerator experiment provides direct evidence of these asymmetries with respect to one of three key outcomes: stereotype reduction. Among respondents facing high social costs, the treatment substantially increased positive evaluations of Black Americans when responses were private (0.36 SD, $p = .007$), but this effect was fully attenuated when responses were public (–0.02 SD, $p = .867$), yielding a statistically significant negative interaction (–0.38 SD, $p = .042$). Behavioral intentions mirror this pattern: although the treatment increased willingness to take public action overall, these effects were entirely driven by individuals facing low social costs, with no detectable effects among those facing high social costs. Finally, while the treatment produced no detectable effects on actual behavior, the design's separation of private and public behavioral outcomes helps distinguish null effects driven by effort costs alone from those arising from the combined barriers of effort and social costs. These results demonstrate that parasocial contact can change private attitudes without enabling corresponding changes in public expression when social costs are high, clarifying when attitude change fails to generate the behaviors that drive social change.

This paper contributes to research on contact and prejudice reduction, and to broader debates about when interventions promote—or fail to promote—meaningful social change. By demonstrating the existence and implications of asymmetric treatment effects, it highlights the need for greater attention to mechanisms, social costs, and the public or private nature of outcomes in contact studies. Critically, many contact studies do not theorize or test both personal and social mechanisms, even when both are plausible (e.g., Scacco and Warren, 2018; Mousa, 2020; Alrababa'h et al., 2021). Similarly, many studies leave unspecified whether attitudinal or behavioral outcomes are visible to ingroup members and potentially constrained by social costs (e.g. Murrar and Brauer, 2018; Mousa, 2020; Hässler et al., 2020). This is consequential as researchers often study contact interventions in settings where social costs to public expressions of tolerance are plausibly high (Paluck, 2009; Scacco and Warren, 2018; Mousa, 2020; Siddique, Vlassopoulos and Zenou, 2024). Clarifying mechanisms, the public or private nature of outcomes, and the link between them can improve measurement, facilitate interpretation of treatment effects, and potentially reduce inconsistencies across and within studies.

More broadly, this paper clarifies the potential and limits of contact and other interventions to

achieve social change. Central to theories of social change is the relationship between private and public outcomes—and specifically, when they converge versus diverge. One influential line of work acknowledges the possibility of gaps between the two but treats them as largely transitory, emphasizing that personal beliefs and perceived norms—and thus private and public outcomes—will ultimately converge, either because belief change induces norm change (de Mesquita, Shadmehr et al., 2023) or because norms recursively shape personal beliefs (Terry and Hogg, 1996; Bicchieri, 2005; Tankard and Paluck, 2016). Recent work by Larson and Lewis (2025) demonstrates one pathway to convergence in showing how attitude change from a prejudice reduction intervention diffuses through social networks. However, such network spillovers depend on individuals being willing to express changed attitudes publicly in the first place. This paper examines the conditions under which social costs inhibit such public expression and potentially give rise to preference falsification instead.

Attention to this possibility is critical as research on preference falsification emphasizes that gaps between private and public outcomes can be pervasive, persistent, and detrimental to social change (Kuran, 1995; Paler, Marshall and Atallah, 2018; Hale and Colton, 2017; Valentim, 2024). Preference falsification can result in misperceived social norms (Bursztyn and Yang, 2022), making dominant positions appear more accepted than they are and dampening momentum for social change. Recent studies have shown the efficacy of interventions that correct such misperceived norms (Bursztyn, González and Yanagizawa-Drott, 2020; Bursztyn, Egorov and Fiorin, 2020), yet this work typically treats underlying beliefs as fixed, aiming to bring public expression in line with existing private preferences. By contrast, this paper highlights the consequences of interventions that shift those underlying preferences, thereby creating the conditions for misperceived norms to emerge or persist. It also underscores an alternative pathway to social change often overlooked in approaches that center on public expression: change that proceeds through the aggregation of private behaviors—such as voting, suppressing discriminatory impulses, or donating to social causes—even when public expression remains constrained.

2 Personal Beliefs, Social Norms, and the Public-Private Gap

Contact—like many interventions that seek to promote social change—can operate through two main mechanisms: by shifting personal beliefs and shifting perceived ingroup norms. This section introduces a framework clarifying how these mechanisms generate asymmetric effects on private versus public outcomes, and when such asymmetries induce, deepen, or reduce preference falsification.

Differentiating Mechanisms of Change

Contact has long been viewed as a means of shaping *personal beliefs*—how individuals think and feel about outgroup members (Allport, 1954; Pettigrew and Tropp, 2011). Such change can operate through multiple specific pathways, including empathy and perspective-taking, reduced stereotyping, and lower intergroup anxiety (Pettigrew and Tropp, 2011; Banas, Bessarabova and Massey, 2020), though evidence that contact interventions reliably produce such change is mixed (Scacco and Warren, 2018; Mousa, 2020; Lowe, 2024).

Contact can also operate by shifting *perceived ingroup norms*—beliefs about what other ingroup members view as socially acceptable attitudes or behaviors toward the outgroup (Paluck, 2009; Pettigrew and Tropp, 2011; Banas, Bessarabova and Massey, 2020; Littman, Scacco and Weiss, 2023). Contact provides opportunities to observe how ingroup members engage with the outgroup, facilitating social learning (Bandura, 1986; Grady, Iannantuoni and Winters, 2021). In group-based interventions, individuals observe ingroup peers interacting with outgroup members; in media-based contact, viewers observe cross-group interaction or friendship—commonly termed *vicarious* contact—both of which can reshape perceptions of acceptable ingroup behavior (Vezzali et al., 2014; Banas, Bessarabova and Massey, 2020).

Although contact interventions may plausibly operate through shifts in personal beliefs, perceived ingroup norms, or both, both mechanisms need not be operative in practice. Clarifying which mechanisms are operative is essential for anticipating whether interventions generate symmetric or asymmetric effects across private and public outcomes. I formalize these distinctions in a simple model showing how belief-based and norm-based interventions generate different effects as a function of social costs. The formalization is presented in Appendix A and summarized below.

Asymmetric Intervention Effects

Individuals decide whether to adopt more favorable attitudes or behaviors toward an outgroup. These outcomes are defined as public or private based on their visibility to other ingroup members. *Private attitudes* encompass affective evaluations (warmth, empathy), cognitive beliefs (stereotypes), and political and policy preferences that individuals hold internally; *public attitudes* are those attitudes as expressed to others. *Private behaviors* are actions not visible to ingroup members, such as voting or suppressing discriminatory impulses, whereas *public behaviors* are visible actions, such as confronting prejudiced remarks or attending protests.

Adoption occurs when perceived benefits exceed perceived costs. Two components shape this calculus. First, outcomes vary in effort costs: expressing an attitude requires minimal effort, whereas political actions such as attending a protest are more effortful. Second, public outcomes have social consequences that depend on perceived ingroup norms. When norms are perceived as inclusive, public support for the outgroup may be socially rewarded; when norms are perceived as prejudiced, such expressions can entail social costs. Individuals also differ in their sensitivity to these consequences, reflecting personality-based differences and features of social environments—e.g. network homogeneity, which generates conformity pressure.

As described above, contact interventions can shift incentives by altering personal beliefs, perceived ingroup norms, or both. Interventions that shift personal beliefs increase the perceived benefits of adoption, regardless of whether outcomes are private or public. By contrast, interventions that shift perceived norms alter the social consequences of adoption, reducing social costs—or increasing social benefits—only for public outcomes, with no direct effect on private ones.

Asymmetric treatment effects depend on which mechanism is operative. I distinguish three ideal-typical cases: interventions that primarily shift personal beliefs but not perceived norms; interventions that primarily shift perceived norms but not personal beliefs; and interventions that affect both. When interventions primarily shift personal beliefs, they increase the perceived benefits of adoption, making private outcomes—attitudes and behaviors—more likely. Public outcomes, however, require that perceived benefits exceed both effort and social costs. Because belief-based interventions do not reduce social costs, they are less likely to enable public adoption unless social costs are sufficiently low.

This logic yields the central insight tested in the paper. Holding baseline benefits and effort costs constant, contact that primarily shifts personal beliefs will have larger effects on private attitudes and behaviors than on public outcomes, with this gap widening as social costs increase. Even when belief-based contact generates benefits sufficient to overcome both effort and social costs, its effects on public outcomes remain smaller than on private outcomes whenever social costs are present. The framework also implies a converse possibility: interventions that primarily shift perceived norms without altering personal beliefs may change public attitudes and behaviors while leaving private outcomes unchanged. While I do not test this implication here, it highlights that asymmetric effects can run in either direction. The framework thus clarifies why contact that primarily shifts personal beliefs is likely to generate asymmetric effects and why public change is demanding: it requires benefits sufficient to clear both effort and social cost barriers.

Implications for Preference Falsification and Social Change

The patterns described above concern asymmetric intervention effects—how different types of contact shift incentives for adopting private versus public outcomes. Asymmetry alone, however, does not determine whether contact induces, deepens, or overcomes preference falsification. Preference falsification occurs when the adoption threshold is crossed in one domain but not the other; it is mitigated when both domains cross their respective thresholds.

Whether belief-based contact reduces or exacerbates preference falsification depends on baseline conditions—where individuals begin relative to the private and public adoption thresholds. Three scenarios are particularly relevant. First, individuals may lie above the public threshold but below the private threshold at baseline. This configuration is likely in contexts where antiprejudice norms are widespread and public expressions of tolerance or outgroup support are socially rewarded, masking privately held prejudiced attitudes and behaviors (Blinder, Ford and Ivarsflaten, 2013). In such settings, interventions that shift personal beliefs can move private adoption past its threshold, aligning private and public outcomes and thereby *reducing* preference falsification. Second, individuals may lie above the private threshold but below the public threshold, consistent with existing preference falsification. In this case, belief-based interventions that leave social costs unchanged can *deepen* preference falsification by further increasing private adoption while public expression remains constrained. Third, individuals may lie below both thresholds at baseline. Here,

belief-based interventions may *induce* preference falsification by moving private adoption past its threshold while public expression remains below it, generating divergence where none previously existed.

The Public-Private Hypotheses

The framework yields testable implications for asymmetric effects on private versus public outcomes. I focus on predictions related to contact interventions that primarily shift personal beliefs, consistent with determining when contact that changes private attitudes also yields public expression. As pre-registered, the framework predicts that such interventions will have:

H1 a positive effect on *private* outcomes.

H2 a small positive or null effect on *public* outcomes.

H3 a smaller effect on *public* than on *private* outcomes.

These predictions are conditional on social costs, which determine when divergence should be most pronounced. The paper therefore examines exploratory hypotheses about heterogeneity by social costs.[2] Social costs sharpen H3, yielding the prediction that contact will have:

H3-E1 a smaller effect on public than on private outcomes *for those facing high social costs.*

Social costs also refine H2. When costs are low, contact that changes beliefs can translate into public outcomes; when costs are high, effects are expected to be attenuated or null. This yields the prediction that contact will have:

H2-E1 a smaller effect on public outcomes *for those facing high versus low social costs.*

3 The Parasocial Contact Experiment

I test these hypotheses by embedding a media experiment in an online survey to examine the effects of parasocial contact on white Americans' attitudes towards Black Americans and support for social

[2]These hypotheses were not formally pre-registered, although the logic motivating these predictions is articulated throughout the pre-analysis plan (see Appendix O).

change, using content from the reality television series *Married at First Sight* (MAFS).

Parasocial Contact as a Personal Belief-Based Intervention

Parasocial contact is among the most prominent forms of media-based contact (Schiappa, Gregg and Hewes, 2005; Banas, Bessarabova and Massey, 2020). The parasocial contact hypothesis holds that positive exposure to outgroup characters can reduce prejudice and improve intergroup attitudes (Schiappa, Gregg and Hewes, 2005, 2006). Political scientists have increasingly drawn on this framework to examine how exposure to outgroup exemplars—such as Barack Obama (Goldman, 2012) or Mo Salah (Alrababa'h et al., 2021)—shapes attitudes towards outgroups as a whole. The same logic underpins the logic of edutainment interventions, which leverage entertaining media narratives to promote social change (Paluck, 2009; Grady, Iannantuoni and Winters, 2021; Siddique, Vlassopoulos and Zenou, 2024).

Parasocial contact is distinct from general media exposure, which may be positive or negative and need not engage viewers or foster emotional ties. By contrast, parasocial contact involves positive exposure, affective attachment to characters, and generalization from depicted individuals to the broader outgroup (Schiappa, Gregg and Hewes, 2005; Bond, 2020). Although some accounts emphasize the importance of sustained exposure, recent evidence suggests that even brief encounters can reduce prejudice (Banas, Bessarabova and Massey, 2020; Joyce and Harwood, 2014; Siddique, Vlassopoulos and Zenou, 2024).

Most importantly, parasocial contact is primarily theorized to operate by shifting personal beliefs rather than perceived ingroup norms (Banas, Bessarabova and Massey, 2020). Considered the media-based analogue of direct contact, parasocial contact is believed to operate through the same personal-belief based pathways, such as empathy and stereotype reduction (Banas, Bessarabova and Massey, 2020; Schiappa, Gregg and Hewes, 2005). This distinguishes parasocial contact from vicarious contact, which operates primarily through social learning and norm change (Joyce and Harwood, 2014; Banas, Bessarabova and Massey, 2020). Parasocial contact therefore has the potential to produce the kind of asymmetric treatment effects predicted by the framework.

At the same time, whether parasocial contact operates primarily through personal belief change is an empirical question. It may indirectly affect perceived norms by prompting viewers to infer others' reactions, signaling popularity, or encouraging discussion (Papa et al., 2000; Chwe, 2013;

Tankard and Paluck, 2016). Consequently, I measure both personal beliefs and perceived norms to test which mechanisms are operative. If the treatment shifts both, the framework predicts symmetric rather than asymmetric effects on private and public outcomes (see Appendix A).

The Context: Racial Prejudice and Social Costs in the United States

The United States provides an informative context for studying parasocial contact and racial prejudice. Private racial prejudice persists: many White Americans hold negative views of Black Americans, and many Black Americans report experiencing discrimination (Pew Research Center, 2015, 2019; Morgan, 2022). Racial attitudes also remain politically consequential, shaping policy preferences and political behavior (Tesler, 2013; Engelhardt et al., 2019).

At the same time, social norms governing the public expression of prejudice vary widely. In some settings, strong anti-prejudice norms render public bias socially unacceptable (Pickup et al., 2023; Hopkins and Washington, 2020). In others, recent political shifts have produced an "emboldening effect," increasing the acceptability of expressing prejudiced views (Newman et al., 2021). Where racial identity is politically salient, individuals may face incentives to express rather than suppress bias (Jardina, 2019). Politically sorted and affectively polarized networks amplify this heterogeneity (Mason, 2018; Jardina and Ollerenshaw, 2022). As a result, the social costs of publicly signaling tolerance vary substantially (Crandall, Miller and White, 2018); for those embedded in more prejudiced networks, such expression may carry meaningful costs.

Moreover, sustained cross-group interaction remains limited. Social networks in the United States are highly racially segregated (McPherson, Smith-Lovin and Cook, 2001; Enos, 2017), constraining opportunities for direct contact and making media a primary source of outgroup exposure. While media portrayals of Black Americans have long reinforced negative stereotypes (Gilliam Jr and Iyengar, 2000; Dixon, 2008*a*,*b*; Collins, 2000), more positive representations have been on the rise in recent years. This shift is evident in reality television—a genre viewed by over 80 percent of U.S. adults (Nussbaum, 2024). Unlike fictional characters or highly salient exemplars, reality television features "ordinary" individuals in intimate settings. Relationship-based shows such as *Married at First Sight* may be especially impactful, depicting outgroup members in relatable, emotionally engaging, and stereotype challenging roles (Post, 2023).

Why *Married at First Sight?*

The reality television show *Married at First Sight* provides a well-suited test of parasocial contact for three reasons.[3] First, its portrayal of successful Black couples fosters parasocial engagement and challenges enduring stereotypes. Each season features three to five couples, including at least one Black couple and one white couple.[4] Couples meet for the first time at the altar and spend eight weeks navigating married life before deciding whether to remain together. Despite its sensational premise, the show centers on relatable individuals confronting familiar relational challenges in ways that encourage viewer attachment while offering counter-stereotypical portrayals of Black Americans as committed and relationship-oriented (Ramasubramanian, 2015).

Second, the show's structure enables a tightly controlled experimental comparison. Couples receive comparable airtime and narrative focus, with their stories unfolding in parallel but separate arcs. This allows the treatment group to follow a successful Black couple while the control group follows a highly similar white couple. Holding content largely constant while varying only the couple's race addresses a core challenge in parasocial contact research: establishing a credible counterfactual (Banas, Bessarabova and Massey, 2020).

Third, *MAFS* enables a clean test of parasocial contact by minimizing vicarious contact—observing interactions between ingroup and outgroup members—which could shift perceived norms. Although some episodes include cross-group friendships, these were peripheral and were excluded from the parasocial contact content shown to participants. This isolates contact effects that operate primarily through personal beliefs rather than perceived norms.

The Treatments

Participants in both conditions watched a 25-minute "mini-season" of *Married at First Sight*. Treatment participants viewed a storyline featuring a successful Black couple, while control participants

[3] *MAFS* first aired in 2014 and has since completed 18 seasons, with recent seasons drawing approximately two million viewers.

[4] This study does not focus on interracial couples; viewing an interracial relationship would blur the distinction between parasocial and vicarious contact, and consequently contact that primarily shifts personal beliefs versus perceived ingroup norms.

viewed a comparable white couple. The content was extensively piloted and edited to ensure comparability, foster positive engagement, and maintain viewer interest.

Respondents were randomly assigned with equal probability to treatment or control and, within each condition, to one of two couples (see Appendix D). Control participants viewed Couple A or C, while treatment participants viewed Couple B or D (see Figure 1). Including multiple couples allows for testing whether effects are driven by specific couples or generalize across them. Each treatment–control pair (A/B and C/D) comes from the same season, enabling season fixed effects to account for production differences.

Figure 1: Parasocial Contact Experiment Couples

Couples A and B

Couples C and D

Notes: Ingroup couples appear on the left and outgroup couples on the right. Control participants were randomly assigned to view content featuring either Couple A or Couple C; treatment participants viewed content featuring either Couple B or Couple D.

4 Empirical Strategy

Sample, Recruitment, and Randomization

A sample of 1,537 white Americans was recruited in October 2023 via CloudResearch Connect using soft quotas to balance respondents by gender and baseline outgroup attitudes.[5] Eligibility was restricted to white, non-Hispanic U.S. citizens aged 18–50 with no prior exposure to *Married at First Sight*. All participants provided informed consent. Following a pre-treatment survey and attention check, respondents were randomly assigned with equal probability to one of three conditions: ingroup parasocial contact (control), outgroup parasocial contact (treatment), or vicarious contact (Appendix D).[6] While all analyses use the full sample to maximize precision, the substantive focus is on the comparison between ingroup and outgroup parasocial contact. Of 1,118 respondents assigned to these conditions, 1,037 completed the study.

Appendix Table E.1 reports baseline descriptive statistics. The data indicate meaningful scope for parasocial contact to affect tolerance and support for social change. Although reported prejudice was generally low, it was not trivial: 19 percent of respondents expressed a preference for white over Black neighbors at baseline. Using the same measure fielded by CloudResearch three days prior to study launch, this share rises to 31 percent.

Outcome Measures and Public-Private Variation

Main Outcomes and Measures

Participants viewed their assigned *MAFS* content immediately after the pre-treatment survey and then completed a post-treatment survey measuring reactions to the content, followed by the main outcomes and mechanisms. Summary statistics appear in Appendix Table E.2.

The primary outcomes capture general positive attitudes toward Black Americans and support for social change. The tolerance index combines feeling thermometers toward Black and white Americans, desire for more Black and white friends, and support for interracial marriage, with

[5] See Appendix Figure B.1 for an overview of the empirical strategy.

[6] The vicarious contact condition was pre-registered as a separate experiment but did not generate detectable first-stage effects and is not analyzed here.

higher values indicating reduced prejudice and greater acceptance of the outgroup. Two versions are reported: one capturing attitudes toward the outgroup only, and one capturing differences between outgroup and ingroup attitudes. The outgroup-only measure provides the cleanest test of parasocial contact, while the difference measure captures more generalized attitudinal change.[7] Support for social change is captured by a single-item measure reflecting a more explicitly political orientation toward racial equality: "When it comes to ensuring that Black people have the same rights and opportunities as White people, do you think this country has gone too far, been about right, or not gone far enough?"

The survey also measures specific pathways associated with personal belief change, including stereotype reduction, empathy, perceived similarity, and comfort with intergroup interaction. Stereotypes are measured by asking respondents to estimate the percentage of Black and white Americans who possess positive traits. As with the tolerance index, two stereotype indices are constructed: one capturing positive attributes of the outgroup and one capturing outgroup–ingroup differences.[8] To evaluate social norms change as a possible operative mechanism, the survey also includes three related measures of perceived ingroup norms—whether respondents believe that white people in general, their white friends, and their white family members want more close friends who are Black.

The Virtual Enumerator Experiment: Private Attitudes vs. Public Expression

This paper uses a novel "virtual enumerator" experiment to test whether the contact treatment has asymmetric effects on attitudes that are privately held versus publicly expressed. Participants were cross-randomized to either a *private* condition, in which they completed a standard online survey, or a *public* condition, in which they were primed to believe they were reporting to a virtual ingroup

[7] While the difference measure was pre-registered as the primary index, it can reflect changes in ingroup attitudes due to ingroup parasocial contact, making it less preferred. See Appendix O for deviations from the pre-analysis plan.

[8] A fifth stereotype item is excluded because it primarily reflects ingroup attitude shifts. See Appendix J.1 for details.

enumerator.[9] In the public condition, respondents viewed an enumerator image and brief chat-style introduction near the start of the post-treatment survey; the image remained visible throughout.

Responses in the private condition reflect respondents' private attitudes, consistent with evidence that self-administered surveys elicit more candid responses than enumerator-administered ones (Tourangeau and Yan, 2007; Kreuter, Presser and Tourangeau, 2008; Krumpal, 2013; Bursztyn et al., 2025). Responses in the public condition capture attitudes respondents are willing to express to an ingroup stranger. The design exploits social desirability pressures and enumerator effects, drawing on evidence that respondents treat interviewers—real or virtual (Krysan and Couper, 2003)—as social referents and adjust their responses to conform to perceived norms (Huddy and Feldman, 2009; Adida et al., 2016; Blair, Coppock and Moor, 2020). Rather than attempting to eliminate social desirability bias—typically treated as methodological nuisance—this study leverages it as a "social microscope" to observe how social pressures shape public expression (Krysan and Couper, 2003, 375).

The design offers three methodological advantages. First, the sole difference between conditions is response visibility, enabling a clean test of whether the treatment affects private attitudes, public expression, or both, while holding constant all other outcome attributes specified in the framework (perceived benefits and effort costs). Second, unlike list experiments or indirect questioning techniques—which focus on one or a small number of sensitive items (Blair, Coppock and Moor, 2020)—it allows measurement of a full range of outcomes at the individual level. Third, because the manipulation mirrors common survey administration modes, the design permits direct assessment of whether response visibility conditions treatment effect detectability—an important consideration given widespread reliance on survey outcomes in contact research (Paluck et al., 2021; Lowe, 2024).

Manipulation checks confirm that respondents noticed the virtual enumerator: 96 percent of participants in the public condition identified having a white enumerator, compared to just 3 percent in the private condition. Additional checks indicate that the effects reflect the intended ingroup priming rather than increased attentiveness (Appendix G).

[9]Participants in the public condition were further randomized to one of eight enumerators varying by age and gender (Appendix C.1).

Private vs. Public Behaviors

To examine whether the contact treatment affects private and public behavior, the survey measures both behavioral intentions and actual behavior. Behavioral intentions are captured using eight items: four explicitly private (voting for a Black or white candidate supporting racial justice, reading an article, signing an anonymous petition) and four explicitly public (sharing an article on social media, confronting a friend or family member, signing a public petition, attending a protest). These items are used to create pre-registered indices of private and public behavioral intentions.

Actual behavior is assessed using a contest experiment administered at the end of the survey (Appendix C.2). Participants could increase their chances of winning a gift card by completing up to eight actions—four related to racial justice and four unrelated. Participants were independently randomized to one of two contest versions: a private version where none of the actions were observable to one's social network (e.g. reading an article) and a public version where all of the actions required visibility to network members (e.g., sharing an article on social media).[10] The design provides a behavioral test of whether parasocial contact increases private and/or public engagement with race (relative to non-race) related actions.

Because behavior outcomes are defined as private versus public, they permit tests of H1 (positive effects on private outcomes) and H2 (small or null effects on public outcomes) but—unlike the virtual enumerator experiment—not a direct test of H3 (asymmetric effects). The behavior measures also differ from the enumerator experiment in substantively important ways. Whereas the enumerator experiment captures public expression to an ingroup stranger, these measures involve visibility of actions to respondents' actual social networks, plausibly closer to behaviors associated with social change. At the same time, this constitutes a noisier test: behaviors vary along multiple dimensions, including perceived benefits and effort costs, making it harder to isolate the role of social costs in generating asymmetric effects. Moreover, if public actions entail both higher effort and higher social costs, it is not possible to determine whether these are driven by social costs and suggested by the framework; the results discussion takes steps to disentangle these. Moreover, because public actions may involve both greater effort and higher social costs than private ones,

[10]Since participants were assigned to either private or public contest versions, this design splits the sample; it enables a test of the effects of contact on private and public actions separately.

observed differences could reflect either mechanism; the analysis below takes steps to disentangle these costs to confirm the role of social costs in driving null results, consistent with the framework.

Data Quality Checks

As shown in Appendix F, covariate balance was achieved across experiments. The parasocial contact experiment exhibits some chance imbalance, addressed by controlling for all covariates in main estimations. There is no evidence of differential attrition or inattention due either to the contact or virtual enumerator experiments (Appendix G). Participants assigned to the contact treatment were, however, more likely to infer the study's purpose, raising potential demand effect concerns. These concerns are mitigated by treatment effects being driven almost entirely by one outgroup couple (Appendix J.2); if demand effects were driving results, similar responses would be expected across both couples.

Estimation

Average treatment effects are estimated using the pre-registered specification:

$$Y_i = \alpha + \delta Contact_i + X'i\psi + \mu j + \epsilon_i \tag{1}$$

where Y_i is the outcome of interest, coded so that higher values indicate more positive outcomes. $Contact_i$ equals 1 if assigned to outgroup parasocial contact and 0 if assigned to ingroup contact. X_i includes pre-specified covariates and an indicator for vicarious contact assignment, and μ_j denotes wave–season fixed effects (Appendix D). Standard errors are robust. When Y_i refers to private behaviors, H1 predicts $\delta > 0$; when Y_i refers to public behaviors, H2 predicts a small positive or null δ. For attitudinal outcomes, this equation estimates average treatment effects pooled across measurement contexts; effects on private versus public attitudes are tested using the interaction model below.

To assess whether treatment effects differ between privately held and publicly expressed attitudes, the following interaction model is estimated:

$$Y_i = \alpha + \beta_1 Contact_i + \beta_2 Pub_i + \beta_3 (Contact_i \times Pub_i) + X_i'\psi + \mu_j + \epsilon_i \tag{2}$$

where Pub_i equals 1 for respondents assigned to the public condition in the virtual enumerator experiment. All other terms are defined as above. Marginal effects are evaluated when $\text{Vic}_i = 0$. A negative interaction ($\beta_3 < 0$) indicates weaker effects on public relative to private attitudes, consistent with H3.

Heterogeneous Effects by Social Cost

Exploratory analyses examine whether treatment effects vary by social costs. Following the pre-analysis plan, respondents are classified as facing high social costs when they both perceive prejudiced ingroup norms and have ingroup-homogeneous social networks.[11] Perceived norms capture the direction of social pressure, while network homogeneity proxies sensitivity to social pressure, consistent with more homogeneous networks generating stronger conformity pressures (McPherson, Smith-Lovin and Cook, 2001). Social costs are therefore considered high when both conditions are present, which characterizes about 40 percent of respondents. Detailed coding procedures and robustness checks are in Appendix H.

To assess whether the public–private gap varies by social cost, the interaction model is extended to include a three-way interaction between parasocial contact, public assignment, and high social cost (Appendix K.2). For behaviors, two-way interactions between contact and social cost are estimated separately for private and public actions.

5 Treatment Check: Parasocial Engagement

The analysis first assesses whether the media treatments generated parasocial engagement. As shown in Appendix I, viewers responded favorably: 89 percent were engaged; 73 percent cared about the couples' outcomes; 73 percent liked the couples overall. Responses were positive for both ingroup and outgroup couples, with slightly more favorable reactions to outgroup couples. Participants also responded positively to both male and female outgroup characters. Most found them likable and relatable, expressed interest in their lives, and perceived them as sharing values. While fewer saw them as potential friends, friendship ratings were similar across conditions.

[11] The operationalization deviates slightly from the pre-analysis plan to improve measurement precision. See Appendix O for details.

Open-ended responses reinforce these findings. One viewer noted: "*I really liked how both Amani and Woody [Couple B] went into this situation with an open mind and a willingness to compromise... I see a lot of my own marriage reflected back at me...*" Another praised Greg and Deonna [Couple D]: "*They had many admirable qualities—high achievers, rooted in faith, pleasant personalities, home owners, valued family... They remind me somewhat of myself and my husband.*" Several respondents referenced race explicitly, revealing that the treatment engaged with and potentially challenged negative stereotypes. One respondent contrasted the portrayal with "*how media presents most Blacks as poor, uneducated or thugs*" while another expressed: *"I like that they are a working-class Black couple... There are so few Black people (at least in the Midwest) who hold down jobs, have long-term commitments, and who make responsible choices."* These comments suggest the treatment provided counter-stereotypical portrayals that surprised respondents holding negative baseline views.

In addition, 67 (74) percent of respondents viewed the male (female) outgroup character as typical of the broader outgroup, and a majority viewed the content as authentic. Taken together, these patterns indicate scope for generalizing from exposure to these characters to attitudes toward the outgroup more broadly.

6 Effects on Private Attitudes and Public Expression

This section examines the effects of the contact treatment on attitudes toward the outgroup, beginning with average treatment effects and then turning to differences between privately held and publicly expressed attitudes using the virtual enumerator experiment.

Average Treatment Effects

Table 1 presents average treatment effects on main attitudinal outcomes. Panel A shows the contact treatment produced positive effects on both tolerance indices: 0.12 SD for the outgroup index ($p = .011$) and 0.10 SD for the difference index ($p = .024$). The largest component effect was on desire for Black friends, which increased by 0.11 units ($p = .004$) with no effect on desire for white friends, consistent with viewing couples as relatable peers. The treatment also increased support for social change by 0.11 units ($p = .034$).

Table 1: Average Treatment Effects

	Control	Contact Effects			N	
	mean	coef	s.e.	p-val	Full	Para
Panel A: General Attitudinal Outcomes						
1a. Warmth index (outgroup)	0.00	0.12**	(0.05)	0.011	1543	1037
1b. Warmth index (diff)	0.00	0.10**	(0.05)	0.024	1543	1037
Thermometer						
Outgroup	7.07	0.03	(0.10)	0.734	1543	1037
Ingroup	7.07	-0.10	(0.11)	0.358	1543	1037
Diff	0.00	0.13	(0.10)	0.202	1543	1037
Want more [. . .] friends						
Outgroup	2.75	0.11***	(0.04)	0.004	1543	1037
Ingroup	2.68	0.01	(0.04)	0.844	1543	1037
Diff	0.07	0.10**	(0.04)	0.012	1543	1037
Support intermarriage	0.00	0.08	(0.05)	0.116	1543	1037
2. Support social change	0.00	0.11**	(0.05)	0.034	1542	1037
Panel B: Personal Belief Mechanisms						
3a. Positive attributes (outgroup)	0.00	0.11**	(0.05)	0.045	1540	1035
3b. Positive attributes (diff)	0.00	0.09*	(0.05)	0.082	1539	1034
Hard working						
Outgroup	6.34	0.22**	(0.11)	0.042	1540	1035
Ingroup	6.29	0.16	(0.10)	0.114	1540	1035
Diff	0.05	0.06	(0.10)	0.537	1539	1034
Intelligent						
Outgroup	6.51	0.13	(0.10)	0.210	1540	1035
Ingroup	6.28	0.03	(0.10)	0.786	1540	1035
Diff	0.23	0.11	(0.09)	0.224	1539	1034
Family values						
Outgroup	6.22	0.20	(0.12)	0.101	1540	1035
Ingroup	5.99	0.09	(0.12)	0.467	1540	1035
Diff	0.21	0.13	(0.13)	0.349	1539	1034
Want commitment						
Outgroup	6.34	0.16	(0.12)	0.189	1540	1035
Ingroup	6.72	-0.02	(0.11)	0.824	1540	1035
Diff	-0.39	0.19*	(0.10)	0.064	1539	1034
4. Similarity index	0.00	0.03	(0.06)	0.579	1540	1035
5. Empathy index	0.00	0.02	(0.05)	0.645	1539	1034
6. Comfort around (diff)	-0.17	0.05	(0.04)	0.198	1542	1037
Panel C: Social Norms Mechanisms						
7. Perceived ingroup norms index	0.00	0.07	(0.05)	0.205	1540	1035

Notes: *p<.10, ** p<.05, *** p<.01. Table reports the effects of the contact treatment on pre-registered outcomes and mechanisms. Indices labeled "(outgroup)" are based on outgroup ratings only; measures labeled "diff" represent the difference between outgroup and ingroup ratings. All indices are inverse-covariance weighted. All measures are presented in their original scales, as summarized in Appendix Table E.1.

Panel B reports effects on specific pathways associated with personal belief change (see Appendix J.1 for detailed results). The clearest evidence emerges for stereotype reduction. The treatment produced a 0.11 standard deviation increase in the outgroup positive attributes index ($p = .045$), with positive shifts across all individual attributes. For example, the treatment increased perceptions of the outgroup's desire for commitment by 0.16 units while having little effect on the same perception for the ingroup, yielding a 0.19 unit difference ($p = .064$). This pattern is noteworthy because "desire for commitment" was the main attribute on which Black Americans were initially rated less favorably than whites, suggesting that the treatment countered a subtle

stereotype. There is little evidence that the treatment affected other personal belief mechanisms such as perceived similarity, empathy, or comfort interacting with outgroup strangers.[12]

Finally, as shown in Panel C, the treatment had no effect on perceived ingroup norms regarding the desirability of outgroup friendships. That the treatment increased respondents' personal desire for outgroup friendships without shifting perceptions of ingroup members' preferences for more outgroup friendships is consistent with the conclusion that it operated primarily through changes in personal beliefs. Taken together, the results suggest the treatment worked primarily by changing personal rather than social beliefs, validating the conditions for testing the public-private hypotheses.[13]

Evaluating the Public-Private Hypotheses

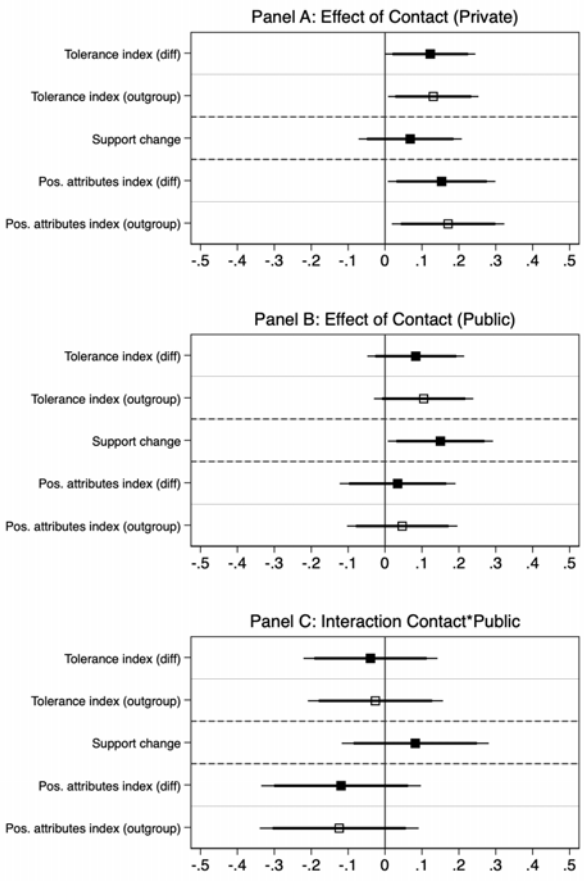
Having established that the treatment operated through personal belief change without shifting perceived norms, I now test whether this generates asymmetric effects on private versus public attitudes. If contact primarily changes personal beliefs rather than perceived norms, its effects should be larger on private than public attitudes. Figure 2 presents the tests of H1–H3 using the virtual enumerator experiment, reporting standardized treatment effects on the two general attitudinal outcomes and the positive attributes index—the primary supported mechanism (see Appendix K.1 for detailed results).

The results reveal important heterogeneity across outcomes while providing suggestive evidence for asymmetric effects for one key outcome: positive attributes (stereotype reduction). Panel A shows effects in the private condition: the treatment produced clear positive effects on both the tolerance and positive attributes indices, and suggestively positive effects on support for social change, consistent with H1. Panel B presents effects in the public condition and reveals partial divergence: While effects on tolerance and support for social change remain positive, effects on perceived positive attributes are clearly null, consistent with H2. Panel C reports the corresponding interaction effects, which are suggestively negative for the positive attributes indices, consistent with

[12] As pre-registered, effects across all specific pathways were not expected; evidence of change in at least one pathway is interpreted as support for personal belief change.

[13] See Appendix J.3 for multiple-hypothesis testing corrections.

Figure 2: Effects of Contact on Private vs. Public Attitudes



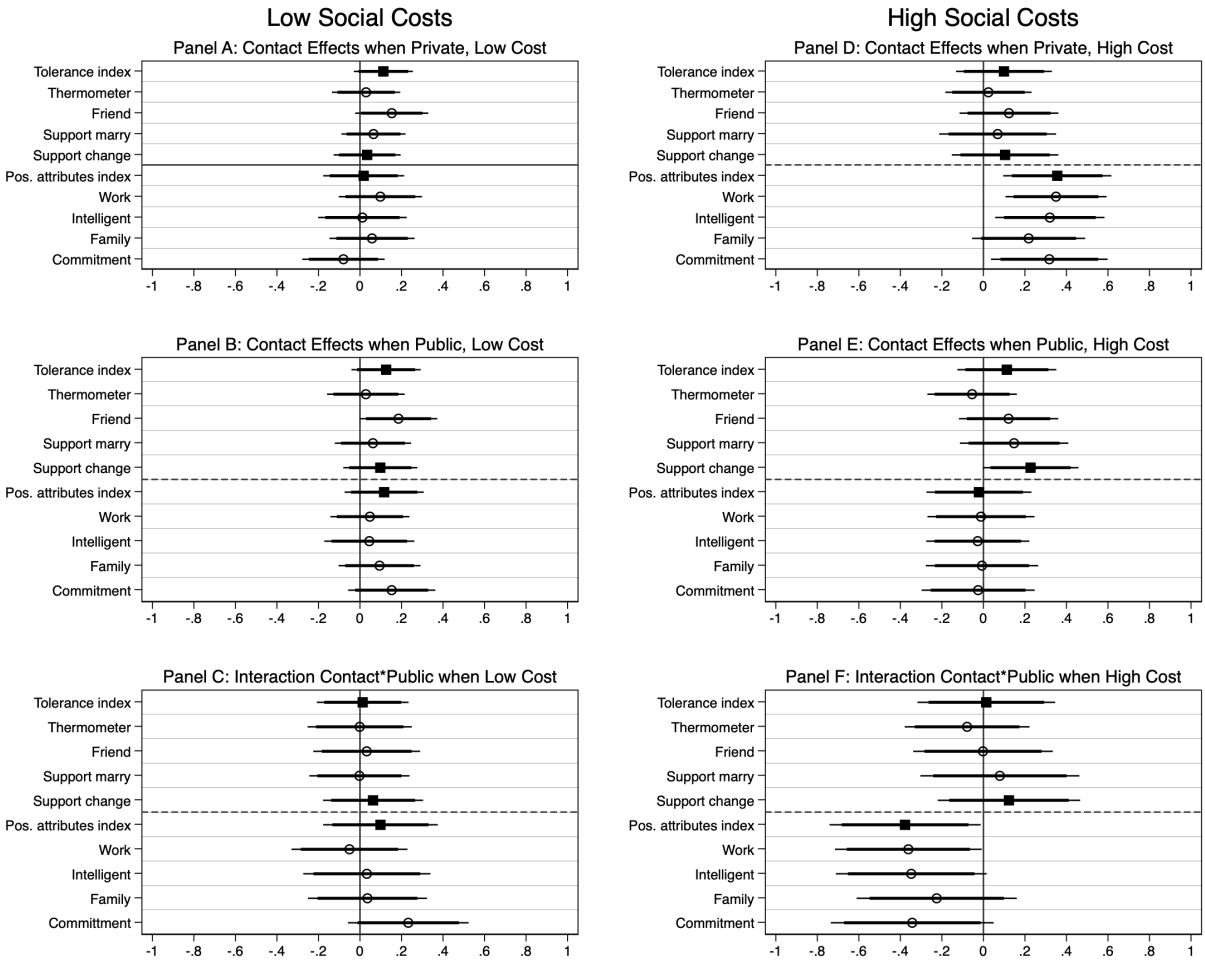
Notes: The figure shows parasocial contact treatment effects on tolerance, support for social change, and perceived positive attributes. Outcomes are standardized to the control group. Panel A shows effects in private (β_1), Panel B shows calculated marginal effects for the public condition ($\beta_1 + \beta_3$) and Panel C shows interaction effects (β_3). Bars show 90% and 95% confidence intervals. Hollow squares reflect outgroup-only measures; black squares reflect outgroup–ingroup differences.

H3. The similarity between the "difference" and "outgroup-only" estimates indicates that these gains reflect improved perceptions of the outgroup rather than shifts in ingroup evaluations.

Although the average interaction effects on perceived positive attributes are suggestively negative, the framework predicts that such public–private divergence should be concentrated among individuals for whom public expression is most socially costly. I therefore test the exploratory hypothesis that the public–private gap varies with social costs (H3-E1). Figure 3 presents marginal effects from the triple-interaction model, plotted separately for respondents facing low versus high social costs.[14]

[14] "Difference" outcomes are excluded because they incorporate ingroup evaluations conditioned on perceived norms toward the outgroup, complicating interpretation of the visibility and social-cost

Figure 3: Effects of Contact on Private vs. Public Attitudes, by Social Cost



Notes: The figure shows the effects of the contact treatment on tolerance, support for social change, and perceived positive attributes. The left column displays results for respondents with low social costs, and the right column shows those with high social costs, estimated using a triple interaction regression. The top row presents effects in the private condition, the middle row in the public condition, and the bottom row shows the interaction between contact and the public condition. Bars represent 90% and 95% confidence intervals.

The central result is clear evidence of the predicted public–private asymmetry for one of the three main outcomes: perceived positive outgroup attributes (reduced stereotyping). Among respondents facing high social costs (right column of Figure 3), the positive average treatment effect is driven entirely by the private condition. The treatment produced a large and statistically significant increase in perceived positive attributes when responses were private (0.36 SD, $p = .007$), but this effect was fully attenuated when responses were public (–0.02 SD, $p = .867$), yielding a negative

mechanisms.

and statistically significant interaction (–0.38 SD, $p = .042$) (see Appendix K.2).[15] Effects are consistent in direction across individual index items, while inter-item correlations are only moderate ($r = 0.38 - 0.68$), indicating the aggregate effect reflects heterogeneous updating across specific attributes rather than uniform or mechanical responding. In contrast, among respondents facing low social costs (left panel), there is little evidence of divergence between private and public effects on positive attributes, consistent with the framework.[16]

Interestingly, the predicted divergence does not extend to the other two outcomes—tolerance or support for social change. That the pattern emerges most clearly for stereotypes is theoretically informative. Stereotype reduction was the primary mechanism through which the treatment operated. Moreover, stereotypes may be particularly subject to social costs: explicitly rating Black Americans on attributes like "work ethic" or "commitment" requires respondents to make group-level judgments that directly invoke racial categories, potentially triggering greater social desirability concerns than more diffuse measures of tolerance or policy support. This interpretation is consistent with evidence that privacy effects vary by the type of racial attitude measured (Krysan, 1998) and with work on aversive racism, which shows that individuals often endorse general egalitarian principles while hesitating to express positive evaluations of outgroup members in social contexts (Dovidio, Gaertner et al., 2004).

Even if the pattern does not generalize across all outcomes, its presence on stereotype reduction is meaningful. Negative stereotyping is among the most widely theorized and empirically supported mechanisms through which contact reduces prejudice (Allport, 1954; Pettigrew and Tropp, 2011). For respondents facing high social costs, estimated treatment effects on this central pathway were

[15] By focusing on within–cost-group comparisons, this analysis avoids confounding from baseline differences between low- and high-cost respondents.

[16] It is also worth noting that treatment effects in the private condition are concentrated among high-cost rather than low cost respondents (Panels A vs. D), suggesting the former had greater scope for belief change. Although the analysis controls for baseline prejudice—which is correlated with perceiving ingroup norms as prejudiced—the larger private effects among high-cost respondents likely reflect residual differences in outgroup attitudes, indicating greater latent capacity for change in privately held attitudes.

highly sensitive to whether attitudes were measured privately or publicly. In effect, whether the treatment "worked" on this outcome depended on how it was measured—highlighting not only when contact shifts private beliefs without public expression, but also how outcome visibility can condition the ability to detect treatment effects at all.

7 Effects on Private and Public Behaviors

I next turn to behavior outcomes. I begin by analyzing treatment effects on intentions to take private and public actions, with results reported in Table 2. Two findings stand out. First, contrary to expectations, the contact treatment did not significantly affect the private behavior index (0.07 SD, $p = .168$). In this case, the null masks notable item-level heterogeneity: while the treatment had no effect on either voting item, it had positive and significant effects on willingness to read an article and sign a private petition. The null effects on voting outcomes likely reflect that electoral choices represent high-stakes decisions involving multiple considerations that brief parasocial contact cannot easily shift. By contrast, lower-stakes actions such as reading an article or signing a petition may be more responsive to such an intervention.

Second, and contrary to H2, the treatment significantly increased the public action index by 0.11 SD ($p = .017$). At the item level, the treatment had positive effects on all four public behaviors, with statistically significant effects on confronting a friend and signing a public petition ($p = .040$ and $p = .008$, respectively).

Table 2: Effects of Contact Treatment on Intentions to Take Private vs. Public Actions

		Control	**Contact Effects**			**N**	
	Prediction	**mean**	coef	se	p-val	Full	Para
Panel A: Private actions							
Private action index	+	0.00	0.07	(0.05)	0.168	1542	1037
Vote Black candidate	+	3.77	-0.01	(0.07)	0.900	1542	1037
Vote white candidate	+	3.73	-0.04	(0.07)	0.599	1542	1037
Read an article	+	3.93	0.12*	(0.07)	0.085	1542	1037
Sign a private petition	+	3.72	0.16**	(0.08)	0.038	1542	1037
Panel B: Public actions							
Public action index	0/+	0.00	0.11**	(0.05)	0.017	1542	1037
Confront friend	0/+	3.75	0.14**	(0.07)	0.040	1542	1037
Attend a public protest	0/+	2.81	0.13	(0.08)	0.117	1542	1037
Share an article	0/+	2.95	0.08	(0.08)	0.300	1542	1037
Sign a public petition	0/+	3.59	0.20***	(0.08)	0.008	1542	1037

Notes: * p<.10, ** p<.05, *** p<.01. Table shows the effects of the contact treatment (estimated using Equation 1) where each row is a different outcome. Individual behaviors measured on 1-5 scale with 5 being very likely and 1 being very unlikely. Indices are z-scores produced using inverse covariance weighting.

Given the mixed pattern across items, I assess whether the behavior measures distinguish private and public actions as intended. Factor analysis largely confirms this distinction (Appendix L.1): the voting items load cleanly on one factor, and the four public behaviors load on another. Two items—reading an article and signing a private petition—cross-load, suggesting respondents may perceive these actions as occupying an intermediate space between private and public.

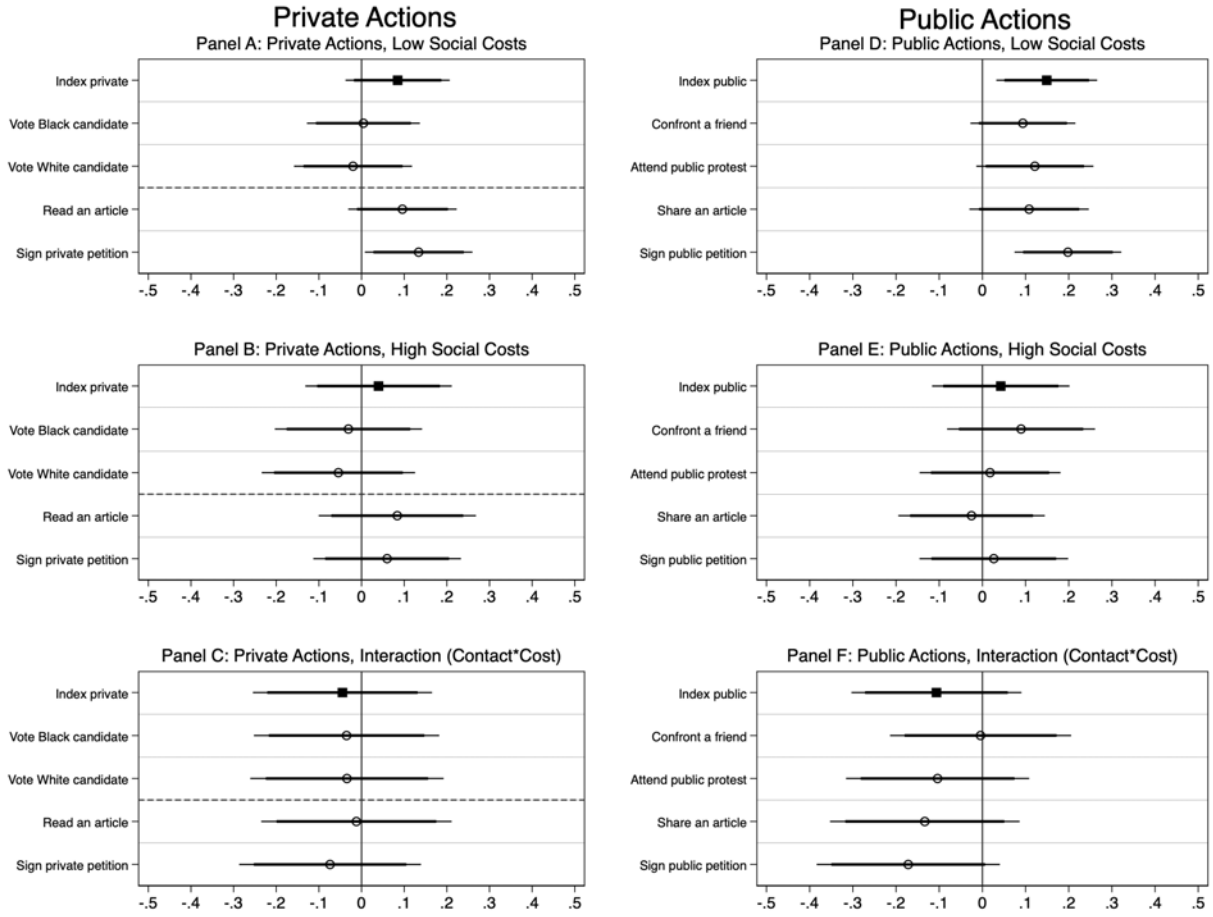
Building on the factor analysis, I next test the exploratory hypothesis that contact effects on private behavior will be unconditional on social costs, while effects on public behavior may be positive for those with low costs and attenuated for those with high costs (H2-E1). These patterns are expected for behaviors that load cleanly as private or public; for intermediate behaviors, ex ante predictions are less clear but the results can indicate whether they function more like private or public actions in practice.

The results, shown in Figure 4 (see Appendix L.2), are strongly suggestive of the predicted pattern for public behaviors (right column). The positive average effect on the public action index is driven almost entirely by respondents facing low social costs: for this group (Panel D), contact produces a sizable increase in public action, with positive effects across all component items. By contrast, among respondents facing high social costs (Panel E), these effects are nonexistent. Consistent with this, the interaction between contact and high social cost is suggestively negative (Panel F), with three of the four public behaviors exhibiting negative point estimates. Although the interaction is not statistically significant, the pattern aligns with expectations and mirrors the virtual enumerator results, which show that contact has substantially weaker effects on public expression when social costs are high.

In contrast, no comparable pattern emerges for private behaviors (left column). The voting items show consistently null effects for both high and low cost respondents (Panels A and B), suggesting the treatment did not meaningfully affect voting preferences. For the intermediate behaviors—reading an article and signing a private petition—there are suggestive positive effects in both cost groups, yielding null interactions (Panel C). The absence of social cost heterogeneity suggests these items may function more like private behaviors.

Comparing treatment effects across similar private and public actions—most notably, reading versus sharing an article and signing a private versus public petition—yields further insights. Among respondents facing low social costs, the treatment produced positive effects on both reading

Figure 4: Effects of Contact Treatment on Behavioral Intentions, by Social Cost



Notes: The figure displays the effects of the contact treatment on behavioral intentions, disaggregated by social cost. Panel A presents effects for individuals with low social costs, while Panel B shows effects for those with high social costs. Panel C displays the interaction effect. Each figure is organized by action type: the top section includes clearly private actions, the middle section presents intermediate actions, and the bottom section includes clearly public actions. A dashed line separates these categories. All outcome variables are standardized to the control group's mean and standard deviation. Squares represent the main pre-registered outcomes, while circles denote key components. Bars represent 90% and 95% confidence intervals.

and sharing an article, as well as on both private and public petition signing (Panels A and D). This pattern is consistent with the prediction that contact should have similar effects on private and public behaviors when social costs are low. These comparisons also help rule out an effort-based explanation for the attenuation of public behavior: if public actions were systematically more effortful than private ones, treatment effects would be uniformly smaller on public actions even among low-cost respondents, which is not what the data show.

Among respondents facing high social costs, however, this symmetry breaks down (Panels B

and E). The treatment effect on reading an article remains positive, while the effect on sharing an article is attenuated and slightly negative (see also Appendix Table L.2). A similar, though less pronounced, pattern appears for petition signing, with smaller effects on public than private petitions. Taken together, these patterns—involving actions that are plausibly comparable in effort—are suggestive of the prediction that contact increases private action regardless of social cost but yields weak or null effects on public action when social costs are high.

Finally, despite clear effects on behavioral intentions, there is no evidence that the contact treatment affected actual behavior—public or private—as measured by contest participation (Appendix L.3). The treatment had no effect on taking private or public race-related (relative to non-race-related) actions. This null result is consistent with H2 (null effects on public behavior) but inconsistent with H1 (positive effects on private behavior).

Crucially, the design allows these null results to yield useful insights nonetheless. Because private behaviors entail effort but not social costs, the absence of treatment effects indicates that the belief change induced by the intervention was insufficient to overcome the effort required for behavior change. By contrast, public behaviors entail both effort and social costs. As a result, they constitute a tougher test for an intervention that primarily shifts personal beliefs—potentially making null effects more likely—but, on their own, provide little insight into why those nulls arise. Had this study only measured public behavior, it would not be possible to discern whether null effects reflected effort costs, social costs, or both. Distinguishing between these barriers matters: it clarifies whether effective intervention design—and the initiation of broader social change—requires strengthening belief change, reducing effort barriers, or coupling belief change with norm-shifting interventions. Thus, as research on contact increasingly relies on incentivized behavioral outcomes because they are costly and less susceptible to demand effects (Lowe, 2024), it is important to consider which costs such outcomes require treatments to overcome—and whether that is plausible given the mechanisms the intervention actually operates through.

8 Conclusion

This paper develops a theoretical framework and provides evidence showing that contact interventions that primarily shift personal beliefs can generate asymmetric effects on private versus

public outcomes—and that these asymmetries depend critically on the social costs of public expression. Using a parasocial contact intervention that changes personal beliefs without shifting perceived ingroup norms, the results show clear private attitude change toward the outgroup, including among individuals facing high social costs. Yet for these same individuals, public expression of those attitudes and willingness to engage in public action are substantially attenuated. Together, these findings demonstrate both the potential for belief change without public expression and the demanding conditions under which private change translates into visible social behavior.

Methodologically, this paper demonstrates how experimentally manipulating measurement contexts can reveal asymmetric treatment effects that would otherwise remain hidden. The virtual enumerator design—which cross-randomizes treatment assignment with perceived visibility of responses—provides a flexible tool for studying when social costs constrain public expression. Critically, the results show that the ability to detect treatment effects can depend on the mode of measurement: effects visible in private contexts may disappear in public ones. This has important implications for interpreting null findings in intervention research, particularly in studies relying on enumerator-administered surveys or other public measurement modes where social costs may bind. More broadly, accurately detecting intervention effects requires matching measurement approaches to operative mechanisms. Belief-based interventions may show effects when measured privately but not publicly if social costs remain high; the framework predicts that norm-shifting interventions should show effects in public measurement even when private attitudes are unchanged. Researchers should therefore carefully consider whether their outcome measures capture private or public expression, and interpret findings accordingly.

The findings have important substantive implications for both theory and practice, and they point to several directions for future research. By documenting asymmetric treatment effects, the results clarify when private and public outcomes converge or diverge. A central implication is the importance of understanding baseline alignment: whether private beliefs and public expression are already aligned or whether a private–public gap exists ex ante. Such divergence is not uncommon. In some settings, individuals may privately hold prejudiced beliefs but behave more tolerantly due to strong anti-prejudice norms (Blinder, Ford and Ivarsflaten, 2013). In others, people may reject prejudice privately but conform publicly to intolerant norms, as may have occurred in Rwanda prior to the genocide (Bhavnani, 2006; Paluck, 2009). Few studies explicitly account for these baseline

conditions. Yet doing so is essential for understanding when interventions are likely to deepen, mitigate, or reverse preference falsification.

This has direct implications for intervention design. When the goal is convergence between private and public outcomes, identifying the binding constraint is critical. Contexts with private prejudice but tolerant norms call for interventions that shift personal beliefs. Conversely, contexts where individuals privately reject prejudice but perceive intolerant norms require interventions that correct misperceived social norms (Bursztyn, González and Yanagizawa-Drott, 2020; Bursztyn, Egorov and Fiorin, 2020). When both personal beliefs and perceived norms are aligned—individuals are both privately and publicly prejudiced—interventions targeting only one mechanism may be insufficient. Effective change may require complementary strategies to shift both beliefs and norms, or a greater focus on the sequencing of how these shape each other over time. This requires better systematic tracking of intervention effects on private and public outcomes over time; without such measurement, researchers risk misinterpreting whether and how change occurs.

Future research should also examine the mirror-image case: contact that primarily shifts perceived norms—rather than personal beliefs—should produce stronger effects on public than private outcomes. When perceived norms change, the social costs of public expression decline even if private beliefs remain unchanged. This logic is consistent with work on misperceived norms, which emphasizes correcting norm perceptions to close gaps between private and public outcomes (Bursztyn, González and Yanagizawa-Drott, 2020; Bursztyn, Egorov and Fiorin, 2020). The framework extends this logic to other interventions, including contact, that may operate through norm-shifting channels.

This framework also offers a new perspective on the longstanding "attitude–behavior gap" (Paluck et al., 2021). Classic theories posit that attitude change precedes behavioral change (Ajzen, 1991; Harmon-Jones, Armstrong and Olson, 2019), yet much of the prejudice-reduction literature documents behavioral change in the absence of corresponding attitude change (Scacco and Warren, 2018; Mousa, 2020; Siddique, Vlassopoulos and Zenou, 2024). This paper suggests that what appears to be an attitude–behavior gap may instead reflect a private–public gap. When researchers compare private attitudes to public behaviors, asymmetric treatment effects of the sort documented here can create the appearance of disconnect even when attitudes and behaviors are aligned within their respective domains. Conversely, interventions that primarily shift perceived norms may affect

public attitudes and public behavior while leaving private outcomes unchanged. Progress in understanding how interventions shape the attitude–behavior relationship therefore requires measuring both private and public attitudes and behaviors, and explicitly accounting for the mechanisms that link them.

The implications of this framework extend beyond the contact literature. Many interventions aim to change personal beliefs, including efforts to correct misinformation belief (Badrinathan and Chauchard, 2024), to influence vote choice through campaign advertising (Paler and Khan, 2025), or to encourage vaccine uptake (Bavel et al., 2020; Bicalho, Platas and Rosenzweig, 2021; Bicchieri and Dimant, 2022). Yet such interventions often also seek to influence public expression, whether through social media sharing, interpersonal persuasion, or other visible behavior within social networks. Other interventions instead target perceived norms, raising parallel questions about their effects on privately held beliefs and behaviors. Explicitly distinguishing between personal beliefs and perceived norms—and between private and public outcomes—is therefore essential not only for accurately evaluating intervention effects, but also for designing strategies capable of bridging the private–public divide and producing durable social change.

References

Adida, Claire, Karen Ferree, Daniel Posner and Amanda Robinson. 2016. "Who's Asking? Interviewer Coethnicity Effects in African Survey Data." *Comparative Political Studies* 49(12):1630–1660.

Ajzen, Icek. 1991. "The theory of planned behavior." *Organizational behavior and human decision processes* 50(2):179–211.

Allport, Gordon. 1954. *The Nature of Prejudice*. Addison-Wesley.

Alrababa'h, Ala', William Marble, Salma Mousa and Alexandra Siegel. 2021. "Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes." *American Political Science Review* 115(4):1111–1128.

Anderson, Michael. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention." *Journal of the American Statistical Association* 103(484):1481–1495.

Badrinathan, Sumitra and Simon Chauchard. 2024. ""I Don't Think That's True, Bro!" Social Corrections of Misinformation in India." *The International Journal of Press/Politics* 29(2):394–416.

Banas, Johm, Elena Bessarabova and Zachary Massey. 2020. "Meta-Analysis on Mediated Contact and Prejudice." *Human Communication Research* 46:120–160.

Bandura, Albert. 1986. *Social Foundations of Thought and Action*. Prentice Hall.

Bavel, Jay J Van, Katherine Baicker, Paulo S Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J Crockett, Alia J Crum, Karen M Douglas, James N Druckman et al. 2020. "Using social and behavioural science to support COVID-19 pandemic response." *Nature human behaviour* 4(5):460–471.

Bhavnani, Ravi. 2006. "Ethnic norms and interethnic violence: Accounting for mass participation in the Rwandan genocide." *Journal of peace research* 43(6):651–669.

Bicalho, Clara, Melina R Platas and Leah R Rosenzweig. 2021. ""If we move, it moves with us:" Physical distancing in Africa during COVID-19." *World Development* 142:105379.

Bicchieri, Cristina. 2005. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, Cristina and Eugen Dimant. 2022. "Nudging with care: The risks and benefits of social information." *Public choice* 191(3):443–464.

Blair, Graeme, Alexander Coppock and Margaret Moor. 2020. "When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments." *American Political Science Review* 114(4):1297–1315.

Blinder, Scott, Robert Ford and Elisabeth Ivarsflaten. 2013. "The better angels of our nature: How the antiprejudice norm affects policy and party preferences in Great Britain and Germany." *American Journal of Political Science* 57(4):841–857.

Bond, Bradley. 2020. "The Development and Influence of Parasocial Relationships with Television Characters: A Longitudinal Experimental Test of Prejudice Reduction through Parasocial Contact." *Communication Research* 48(4):573–593.

Broockman, David and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment in door-to-door canvassing." *Science* 352(6282).

Bursztyn, Leonardo, Alessandra L González and David Yanagizawa-Drott. 2020. "Misperceived social norms: Women working outside the home in Saudi Arabia." *American economic review* 110(10):2997–3029.

Bursztyn, Leonardo and David Y Yang. 2022. "Misperceptions about others." *Annual Review of Economics* 14(1):425–452.

Bursztyn, Leonardo, Georgy Egorov and Stefano Fiorin. 2020. "From extreme to mainstream: The erosion of social norms." *American economic review* 110(11):3522–3548.

Bursztyn, Leonardo, Ingar K Haaland, Nicolas Röver and Christopher Roth. 2025. The Social Desirability Atlas. Technical report National Bureau of Economic Research.

Chwe, Michael Suk-Young. 2013. *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press.

Collins, Patricia Hill. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 2 ed. New York: Routledge.

Crandall, Christian S, Jason M Miller and Mark H White. 2018. "Changing norms following the 2016 US presidential election: The Trump effect on prejudice." *Social Psychological and Personality Science* 9(2):186–192.

de Mesquita, Ethan Bueno, Mehdi Shadmehr et al. 2023. "Social norms and social change." *Quarterly Journal of Political Science* 18(3):339–363.

Dixon, Travis. 2008*a*. "Network News and Racial Beliefs: Exploring the Connection Between National Television News Exposure and Stereotypical Perceptions of African Americans." *Journal of Communications* 58(321-337).

Dixon, Travis L. 2008*b*. "Crime News and Racialized Beliefs: Understanding the Relationship Between Local News Viewing and Perceptions of African Americans and Crime." *Journal of Communication* 58(1):106–125.

Dovidio, John F, Samuel L Gaertner et al. 2004. "Aversive racism." *Advances in experimental social psychology* 36:4–56.

Engelhardt, Andrew M et al. 2019. "Trumped by race: Explanations for race's influence on whites' votes in 2016." *Quarterly Journal of Political Science* 14(3):313–328.

Enos, Ryan D. 2017. *The space between us: Social geography and politics*. Cambridge University Press.

Gilliam Jr, Franklin D and Shanto Iyengar. 2000. "Prime suspects: The influence of local television news on the viewing public." *American journal of political science* pp. 560–573.

Goldman, Seth. 2012. "Effects of the 2008 Obama Presidential Campaign on White Racial Prejudice." *Public Opinion Quarterly* 76(4):663–687.

Grady, Christopher, Alice Iannantuoni and Matthew Winters. 2021. "Influencing the means but not the ends: The role of entertainment-education interventions in development." *World Development* 138.

Granovetter, Mark. 1978. "Threshold models of collective behavior." *American journal of sociology* 83(6):1420–1443.

Hale, Henry E and Timothy J Colton. 2017. "Who Defects? Unpacking a Defection Cascade from Russia's Dominant Party 2008–12." *American Political Science Review* 111(2):322–337.

Harmon-Jones, Eddie, Joel Armstrong and James Olson. 2019. *The Handbook of Attitudes*. Routledge chapter The Influence of Behavior on Attitudes.

Hässler, Tabea, Johannes Ullrich, Michelle Bernardino, Nurit Shnabel, Colette Van Laar, Daniel Valdenegro, Simone Sebben, Linda R Tropp, Emilio Paolo Visintin, Roberto González et al. 2020. "A large-scale test of the link between intergroup contact and support for social change." *Nature Human Behaviour* 4(4):380–386.

Hopkins, Daniel J and Samantha Washington. 2020. "The rise of Trump, the fall of prejudice? Tracking white Americans' racial attitudes via a panel survey, 2008–2018." *Public opinion quarterly* 84(1):119–140.

Huddy, Leonie and Stanley Feldman. 2009. "On Assessing the Political Effects of Racial Prejudice." *Annual Review of Political Science* 12:423–47.

Jardina, Ashley. 2019. *White Identity Politics*. Cambridge University Press.

Jardina, Ashley and Trent Ollerenshaw. 2022. "The polls—trends: The polarization of White racial attitudes and support for racial equality in the US." *Public Opinion Quarterly* 86(S1):576–587.

Joyce, Nick and Jake Harwood. 2014. "Improving Intergroup Attitudes through Televised Vicarious Intergroup Contact: Social Cognitive Processing of Ingroup and Outgroup Information." *Communication Research* 41(5):627–643.

Kim, Eunji and SHAWN Patterson Jr. 2025. "The American viewer: Political consequences of entertainment media." *American Political Science Review* 119(2):917–931.

Kreuter, Frauke, Stanley Presser and Roger Tourangeau. 2008. "Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity." *Public opinion quarterly* 72(5):847–865.

Krumpal, Ivar. 2013. "Determinants of social desirability bias in sensitive surveys: a literature review." *Quality & quantity* 47(4):2025–2047.

Krysan, Maria. 1998. "Privacy and the expression of white racial attitudes: A comparison across three contexts." *Public Opinion Quarterly* pp. 506–544.

Krysan, Maria and Mick Couper. 2003. "Race in the Live and Virtual Interview: Racial Deference, Social Desirability, and Activation Effects in Attitude Surveys." *Social Psychology Quarterly* 66(4):364–383.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, MA: Harvard University Press.

Larson, Jennifer M and Janet I Lewis. 2025. "Reducing prejudice toward refugees: Evidence that social networks influence attitude change in Uganda." *American Political Science Review* 119(1):349–367.

Littman, Rebecca, Alexandra Scacco and Chagai Weiss. 2023. Reducing prejudice through intergroup contact interventions. In *Psychological Intergroup Interventions*. Routledge pp. 3–16.

Lohmann, Susanne. 1994. "The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91." *World politics* 47(1):42–101.

Lowe, Matt. 2024. "Has intergroup contact delivered?" *Annual Review of Economics* 17.

Mason, Liliana. 2018. *Uncivil Agreement: How Politics Became Our Identity*. Chicago: University of Chicago Press.

McPherson, Miller, Lynn Smith-Lovin and James Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Morgan, Stephen L. 2022. "Prejudice, Bigotry, and Support for Compensatory Interventions to Address Black-White Inequalities: Evidence from the General Social Survey, 2006-2020." *Sociological science* 9.

Mousa, Salma. 2020. "Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq." *Science* 369(6505):866–870.

Murrar, Sohad and Markus Brauer. 2018. "Entertainment-Education Effectively Reduces Prejudice." *Group Processes and Intergroup Relations* 21(7):1053–1077.

Newman, Benjamin, Jennifer L Merolla, Sono Shah, Danielle Casarez Lemi, Loren Collingwood and S Karthick Ramakrishnan. 2021. "The Trump effect: An experimental investigation of the emboldening effect of racially inflammatory elite communication." *British Journal of Political Science* 51(3):1138–1159.

Nussbaum, Emily. 2024. *Cue the Sun!: The Invention of Reality TV*. Random House.

Paler, Laura, Leslie Marshall and Sami Atallah. 2018. "The Social Costs of Public Political Participation: Evidence from a Petition Experiment in Lebanon." *Journal of Politics* 80(4):1405–1410.

Paler, Laura and Sarah Khan. 2025. "Shy Women Voters: Private Preferences and Public Expression in the 2024 U.S. Presidential Election.".

Paluck, Elizabeth Levy. 2009. "Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda." *Journal of Personality and Social Psychology* 96(3):574–587.

Paluck, Elizabeth Levy, Roni Porat, Chelsey Clark and Donald Green. 2021. "Prejudice Reduction: Progress and Challenges." *Annual Review of Psychology* 72:533–60.

Papa, Michael, Arvind Singhal, Sweety Law, Saumya Pant, Suruchi Sood, Everett Rogers and Corinne Shefner-Rogers. 2000. "Entertainment-Education and Social Change: An Analysis of Parasocial Interaction, Social Learning, Collective Efficacy, and Paradoxical Communication." *Journal of Communication* 50(4):31–55.

Pettigrew, Thomas and Linda Tropp, eds. 2011. *When Groups Meet: The Dynamics of Intergroup Contact*. Psychology Press.

Pew Research Center. 2015. "Exploring Racial Bias Among Biracial and Single-Race Adults: The IAT.". Accessed: 2025-06-10.
URL: *https://www.pewresearch.org/social-trends/2015/08/19/exploring-racial-bias-among-biracial-and-single-race-adults-the-iat/*

Pew Research Center. 2019. "Race in America 2019.". Accessed: 2025-06-10.
URL: *https://www.pewresearch.org/social-trends/2019/04/09/race-in-america-2019/*

Pickup, Mark, Erik O Kimbrough, Eric William Groenendyk and Antoine Jevon Banks. 2023. "What Counts as Racism? American Norms, Partisan Norms, and Personal Beliefs About Racism.".

Post, Washington. 2023. "'Love Is Blind' star Brett knew Tiffany was the one when she did this.". Accessed: 2025-06-10.
URL: *https://www.washingtonpost.com/arts-entertainment/2023/04/14/love-is-blind-tiffany-brett/*

Ramasubramanian, Srividya. 2015. "Using celebrity news stories to effectively reduce racial/ethnic prejudice." *J. Soc. Issues* 71(1):123–138.

Scacco, Alexandra and Shana Warren. 2018. "Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria." *American Political Science Review* pp. 1–24.

Schiappa, Edward, Peter Gregg and Dean Hewes. 2005. "The Parasocial Contact Hypothesis." *Communication Monographs* 72(1):92–115.

Schiappa, Edward, Peter Gregg and Dean Hewes. 2006. "Can One TV Show Make a Difference? Will and Grace and the Parasocial Contact Hypothesis." *Journal of Homosexuality* 51(4):15–37.

Siddique, Abu, Michael Vlassopoulos and Yves Zenou. 2024. "Leveraging edutainment and social networks to foster interethnic harmony." *Available at SSRN 4810983* .

Tankard, Margaret and Elizabeth Levy Paluck. 2016. "Norm Perception as a Vehicle for Social Change." *Social Issues and Policy Review* 10(1):181–211.

Terry, Deborah and Michael Hogg. 1996. "Group Norms and the Attitude-Behavior Relationship: A Role for Group Identification." *Personality and Social Psychology Bulletin* 22(8):776–793.

Tesler, Michael. 2013. "The Return of Old-Fashioned Racism to White Americans' Partisan Preferences in the Early Obama Era." *Journal of Politics* 75(1):110–123.

Tourangeau, Roger and Ting Yan. 2007. "Sensitive questions in surveys." *Psychological bulletin* 133(5):859.

Valentim, Vicente. 2024. "Political Stigma and Preference Falsification: Theory and Observational Evidence." *The Journal of Politics* 86(4):1382–1402.

Vezzali, Loris, Miles Hewstone, Dora Capozza, Dino Giovannini and Ralf Wolfer. 2014. "Improving intergroup relations with extended and vicarious forms of indirect contact." *European Review of Social Psychology* 25(314-389).

Weiss, Chagai M. 2021. "Diversity in health care institutions reduces Israeli patients' prejudice toward Arabs." *Proceedings of the National Academy of Sciences* 118(14):e2022634118.

Appendix

Table of Contents

A Conceptual Framework

This appendix presents a utility-based framework that formalizes how a contact intervention can generate asymmetric effects on private versus public outcomes. The model integrates two mechanisms emphasized in the literature—shifts in personal beliefs and shifts in perceived ingroup norms—and shows how each maps onto distinct outcomes through separate components of utility.

The framework distinguishes two analytically separate questions. First, whether an intervention produces *asymmetric treatment effects*, meaning that private and public outcomes respond differently to the same intervention. Second, whether an intervention generates *convergence or divergence in levels*, meaning whether private and public utilities ultimately lie on the same or opposite sides of their respective adoption thresholds. These distinctions matter because the effect of asymmetric responsiveness on convergence or divergence depends on baseline conditions—most importantly, whether preference falsification is present prior to intervention. Understanding convergence/divergence, in turn, is what allows for linking asymmetric responsiveness to broader theories of social change.

A.1 Setup

To formalize the framework, I model the utility of adopting tolerant attitudes or behaviors in private and public settings as a function of perceived benefits and costs. The key distinction is that public outcomes, unlike private ones, are subject to social costs or social benefits that depend on perceived ingroup norms and individuals' sensitivity to social sanctioning.

In the utility functions, let $B \in \mathbb{R}$ denote the baseline personal benefit an individual derives from adopting a tolerant attitude or behavior, reflecting their underlying valuation of tolerance. Negative values ($B < 0$) correspond to baseline prejudice.

Utility is also shaped by two distinct cost components. The first, $C_E \geq 0$, captures effort costs associated with adoption. Low values of C_E correspond to low-effort outcomes, such as expressing a tolerant attitude on a survey, whereas higher values represent increasingly effortful behaviors, such as signing a petition or attending a protest.

The second cost component, $C_S \in \mathbb{R}$, enters the utility function only for public outcomes and captures the social costs or social benefits of public expression. When $C_S > 0$, public expression entails social sanctioning that reduces utility; when $C_S < 0$, it entails social benefits that increase utility. This parameter reflects the perceived *direction* of social pressure, determined by perceived ingroup norms. The effect of C_S is scaled by a parameter $\sigma \in [0, 1]$ that captures an individual's sensitivity to ingroup sanctioning. This sensitivity may arise from structural features of one's social environment—e.g. network homogeneity, which strengthens conformity incentives (McPherson, Smith-Lovin and Cook, 2001)—or from stable individual differences in responsiveness to social pressure.

Baseline utilities are given by:

$$U^{Priv} = B - C_E, \qquad U^{Pub} = B - C_E - \sigma C_S.$$

A.2 Thresholds and Baseline Convergence/Divergence

Individuals adopt an outcome when its utility exceeds zero. This yields distinct adoption thresholds for private and public outcomes:

$$B > C_E \quad \text{(Private threshold)},$$

$$B > C_E + \sigma C_S \quad \text{(Public threshold)}.$$

The relative positions of private and public utilities with respect to these thresholds determine whether outcomes are *convergent* or *divergent* at baseline. Convergence occurs when private and public utilities lie

on the same side of their respective thresholds, either both above or both below. Divergence occurs when private and public utilities lie on opposite sides of their thresholds, indicating a misalignment between private and public outcomes.

Depending on baseline values of B, C_E, and C_S, individuals therefore fall into one of three qualitatively distinct regions:

1. **Convergence: neither private nor public adoption.** $B < C_E$ and $B < C_E + \sigma C_S$: both private and public utilities lie below their thresholds.
2. **Convergence: both private and public adoption.** $B > C_E + \sigma C_S$: both private and public utilities exceed their thresholds.[1]
3. **Divergence between private and public outcomes.** Private and public utilities lie on opposite sides of their thresholds. Two forms of divergence are possible, depending on the direction of social costs:
   - **Preference falsification** (when $C_S > 0$): $C_E < B < C_E + \sigma C_S$, so private utility exceeds its threshold while public utility does not.
   - **Misrepresentation** (when $C_S < 0$): $C_E + \sigma C_S < B < C_E$, so public utility exceeds its threshold while private utility does not.

The above distinguishes two forms of divergence. When individuals privately support tolerance but refrain from public expression due to social costs, divergence takes the form of classic *preference falsification*, characterized by private tolerance and public silence. When individuals publicly express tolerance despite lacking private support—typically because social incentives reward such expression—divergence instead reflects *misrepresentation*, characterized by public tolerance without private internalization.

Although four private–public outcome combinations are logically possible, the structure of the model implies that only three regions are feasible in any given context. When $C_S > 0$, the public threshold strictly exceeds the private threshold, ruling out misrepresentation and allowing divergence only as preference falsification. When $C_S < 0$, the ordering reverses, ruling out preference falsification and allowing divergence only as misrepresentation. Thus, for any given set of baseline parameters, individuals may exhibit convergence below threshold, convergence above threshold, or a single form of divergence.

A.3 Interventions

Interventions may shift beliefs, norms, or both.

Belief-based interventions increase perceived benefits:

$$B \to B + \delta_B, \qquad \delta_B > 0.$$

Norm-based interventions reduce perceived social costs:

$$C_S \to C_S - \delta_S, \qquad \delta_S > 0.$$

Joint interventions affect both components:

$$(B, C_S) \to (B + \delta_B,\ C_S - \delta_S).$$

Belief shifts increase utilities directly, while norm shifts reduce the social-cost component of public expression. Table A.1 summarizes the resulting utility functions and shows how belief-based and norm-based contact enter private and public utilities differently.

[1] Because the public threshold weakly dominates the private threshold for any C_S, satisfying $B > C_E + \sigma C_S$ necessarily implies $B > C_E$.

Table A.1: Utilities for adopting tolerant private and public outcomes

	U^{Priv}	U^{Pub}
Baseline (control)	$B - C_E$	$B - C_E - \sigma C_S$
Contact affects personal beliefs	$B + \delta_B - C_E$	$B + \delta_B - C_E - \sigma C_S$
Contact affects perceived norms	$B - C_E$	$B - C_E - \sigma(C_S - \delta_S)$
Contact affects both	$B + \delta_B - C_E$	$B + \delta_B - C_E - \sigma(C_S - \delta_S)$

Notes: Table shows theorized utility functions of adopting private and public outcomes. All parameters are indexed by individual i; subscripts are omitted for clarity.

A.4 Asymmetric Treatment Effects

Asymmetric treatment effects concern how utilities respond to intervention, holding adoption thresholds fixed. Because private and public utilities incorporate different cost components, belief-based and norm-based interventions generate distinct patterns of responsiveness.

Belief shifts. Belief-based interventions increase both utilities:

$$U^{Priv} \to B + \delta_B - C_E, \qquad U^{Pub} \to B + \delta_B - C_E - \sigma C_S.$$

The private–public wedge remains:

$$U^{Priv} - U^{Pub} = \sigma C_S.$$

Because this wedge increases monotonically in σC_S, belief-based interventions generate asymmetric responsiveness whenever social costs are positive. Private utilities increase one-for-one with δ_B, whereas public utilities remain dampened by social pressure, implying larger treatment effects on private than on public outcomes as σC_S increases.

Norm shifts. Norm-based interventions operate by shifting the social cost component of public expression, such that $C_S \to C_S - \delta_S$ with $\delta_S > 0$. Private utility is unaffected, while public utility increases:

$$U^{Pub} \to B - C_E - \sigma(C_S - \delta_S).$$

This reduces the private-public wedge to $\sigma(C_S - \delta_S)$. Sufficiently large norm shifts can eliminate social costs entirely or convert them into social benefits, implying that norm change may not only relax constraints on public expression but also actively reward it. As a result, norm-based interventions generate the reverse asymmetry of belief shifts, with larger effects on public than on private outcomes. Although the framework allows for norm reversal in principle, the empirical analysis focuses on settings in which baseline social pressure entails costs ($C_S > 0$) and does not seek to identify reward regimes.

Joint shifts. When contact simultaneously shifts personal beliefs and perceived norms ($\delta_B > 0$ and $\delta_S > 0$), both private and public utilities increase. The direction and magnitude of asymmetry depend on the relative strength of the two mechanisms. Larger belief shifts widen the private-public wedge, whereas larger norm shifts shrink it. Symmetric responsiveness arises only when social pressure is fully neutralized ($\delta_S = C_S$) or when individuals are insensitive to social sanctioning ($\sigma = 0$), in which case private and public utilities respond equally to belief change.

A.5 Convergence and Divergence at Endline

Asymmetric responsiveness describes how private and public utilities respond to intervention; convergence and divergence describe their resulting alignment relative to adoption thresholds (Section A).

At baseline, individuals occupy one of three states: convergence below threshold, convergence above threshold, or divergence between private and public outcomes. Divergence reflects a misalignment between private

and public utilities and takes the form of either preference falsification or misrepresentation, depending on the direction of social pressure.

Belief-based interventions raise both private and public utilities, whereas norm-based interventions affect only public utility by reducing social pressure. Endline convergence or divergence depends on whether these shifts move both utilities across their thresholds. Divergence persists when treatment shifts only one utility across its threshold and disappears when both cross; new divergence can arise when interventions affect one domain but not the other.

When baseline convergence. To illustrate, consider interventions that primarily shift personal beliefs ($\delta_B > 0$, $\delta_S \approx 0$). For individuals who begin in a convergent state below threshold—such that both private and public utilities lie below their respective thresholds—a moderate belief shift may move private utility above its threshold while leaving public utility below threshold due to persistent social costs. In this case, belief change generates private adoption without public expression, producing preference falsification at endline. Only when belief change is sufficiently large to also induce public expression—that is, to overcome both effort and social costs—will belief-based interventions produce convergence.

When baseline preference falsification. Similarly, for individuals who begin in a state of preference falsification—where private utility exceeds its threshold but public utility does not—moderate belief shifts raise private utility further above threshold without enabling public adoption, thereby deepening existing divergence between private support and public expression.

When baseline misrepresentation. Conversely, when individuals begin in a divergent state characterized by misrepresentation—where public utility exceeds its threshold but private utility does not—contact that shifts personal beliefs can promote convergence by allowing private utility to cross its threshold, aligning private and public outcomes.

Although the empirical analysis does not directly observe transitions across these baseline states—because individuals' counterfactual locations relative to thresholds are not observed—this framework clarifies the implications of asymmetric treatment effects for social change. In particular, it shows that belief-based interventions can improve private attitudes without generating public expression, induce or deepen preference falsification, or produce convergence only under specific conditions.

A.7 Summary of Predictions

A.7.1 Predictions Tested Empirically

P1. Belief-based interventions ($\delta_B > 0$, $\delta_S \approx 0$) produce larger effects on private than on public outcomes whenever social costs are positive ($\sigma C_S > 0$).

P2. The private–public effect gap increases with the magnitude of social costs (σC_S)

Note that "outcomes" refers to both attitudes and behaviors. Attitudes and behaviors differ in their effort costs C_E but share the same benefit parameter B and social cost term C_S for public expression. Holding B and C_E constant implies parallel treatment effects for attitudes and behaviors within private or public domains. Differences between attitudes and behaviors arise only when B or C_E differs across them. Thus, the framework predicts divergence across private and public outcomes, not between attitudes and behaviors per se.

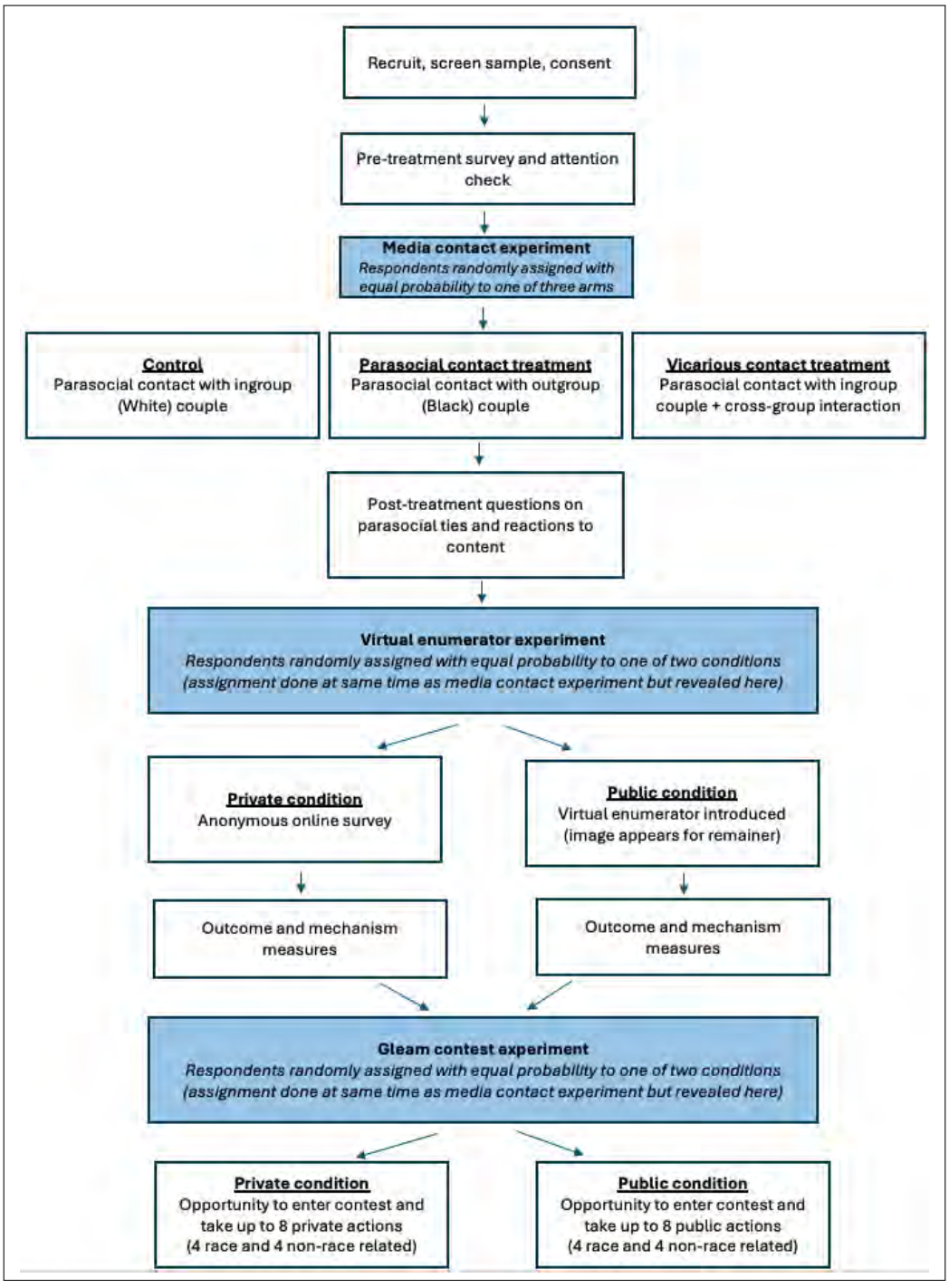
A.7.2 Predictions Not Tested Here

P3. Norm-based interventions ($\delta_S > 0$, $\delta_B \approx 0$) generate reverse asymmetry, with larger effects on public than on private outcomes.

P4. Joint belief-and-norm interventions generate asymmetry whose direction depends on the relative magnitudes of δ_B and δ_S; sufficiently strong shifts in both can eliminate asymmetry.

B Research Design Overview

Figure B.1: Research Design Overview



C The Public/Private Experiments

C.1 The Virtual Enumerator Experiment

As described in Section 4, respondents were randomly assigned with equal probability to either the *private* or *public* condition in the virtual enumerator experiment. Those assigned to the public condition were further randomly assigned to one of eight ingroup enumerators that varied in gender and age, as shown in Figure C.1. Virtual enumerators were initially introduced using images and chat bubbles (see Figure C.2), and then the image appeared throughout the remainder of the post-treatment survey.

Figure C.1: Enumerator Images

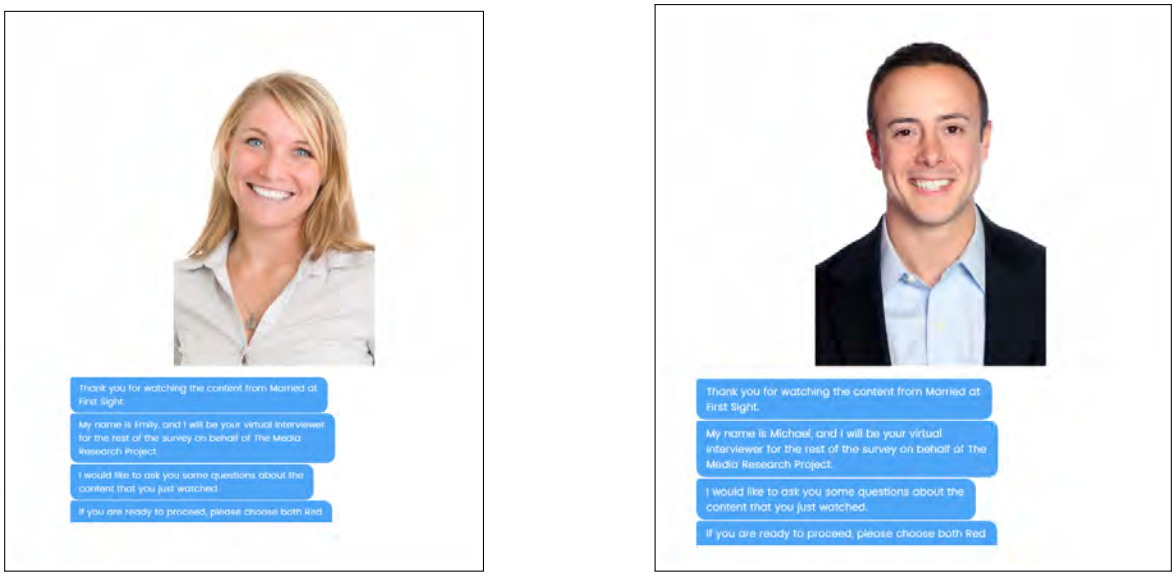


Figure C.2: Examples of virtual ingroup enumerator introductions

C.2 The Gleam Contest Experiment

The incentivized behavioral outcome measure was a novel experiment embedded at the end of the post-treatment survey and framed as a contest to win a $500 Amazon gift card. The contest was implemented via Gleam.io, a marketing platform that allows contest entrants to complete actions in exchange for additional entries that could increase their changes of winning. There were eight possible actions for each contest, with each action worth one additional entry; Gleam automatically verifies each action before awarding the additional entry.

Public/private contests: To test the effects of contact on private and public behavior, two versions of the contest were created: one containing only private actions (e.g., reading an article) and one containing only public actions (e.g., sharing an article on social media). In each version, participants could complete up to eight actions, four of which were racial justice related and four of which were not. Non–race-related actions allowed participants to increase their chances of winning without having to take race-related actions. Table C.1 lists all actions. Figure C.3 shows images of the private versus public versions of the contest.

Minimizing ordering effects: To mitigate concerns about ordering effects (participants simply completing whichever actions appeared first), the study was implemented in 12 waves where the order of the actions for each contest was randomized in each wave. Details on randomization can be found in Appendix D.

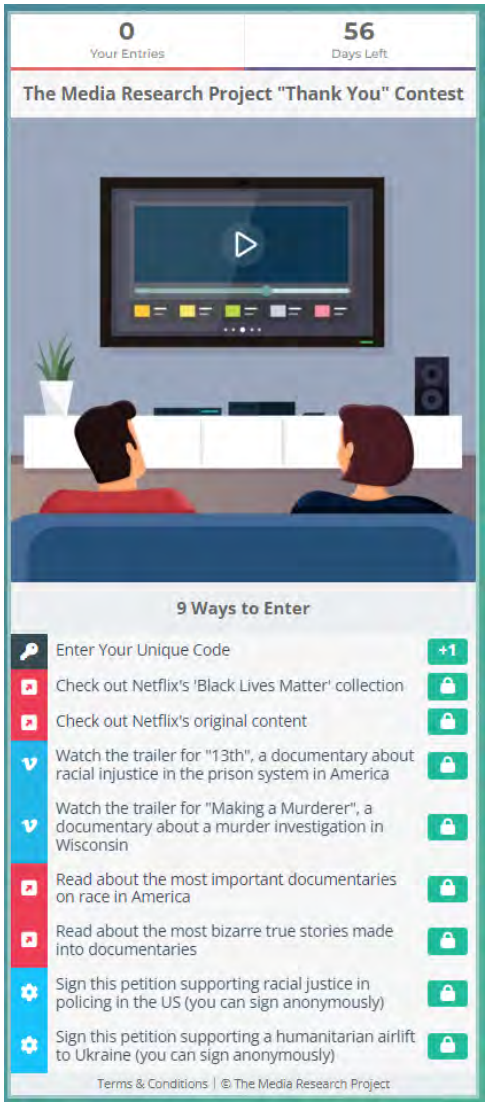
Outcome measures: The primary outcome measures for estimating the average treatment effects of parasocial contact are: (1) whether a participant chose to enter the contest, and (2) the difference between the total number of race-related and non-race-related actions taken. The logic behind the first measure is that participants viewed the list of available actions before deciding whether to enter such that those who saw actions they were more willing to take would be more likely to enter the contest in the first place. To assess effects on private versus public behavior, the analysis focuses on the differences between race-related and non-race-related actions separately within the private and public contest conditions.

Contest winners: Winners in the Gleam.io contest were drawn after data collection was completed and all entries verified. There was one winner selected from each of the private and public contests; each winner received the promised $500 Amazon gift card.

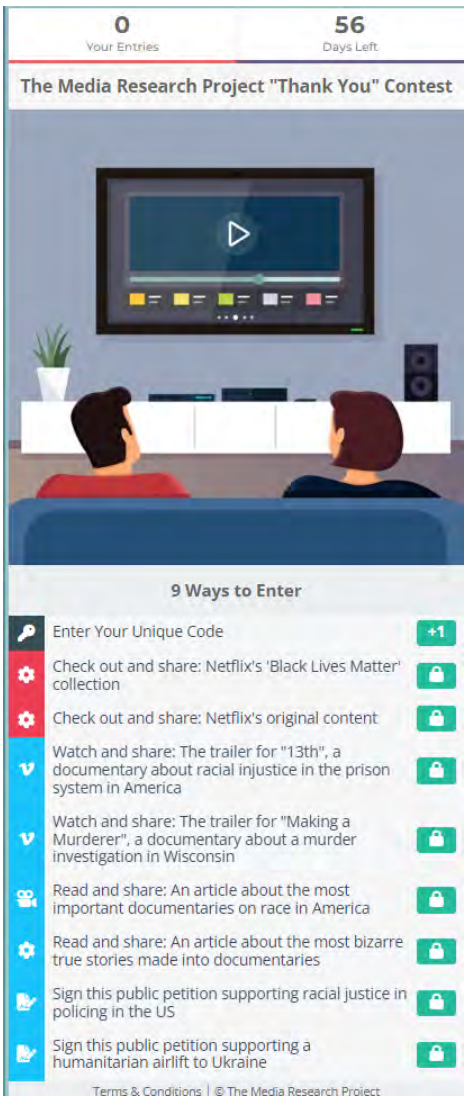
Table C.1: Gleam Contest Measures

	Race-related action	Not race-related action
Private contest	1. Browse Netflix's BLM collection 2. Watch the trailer for "13th" 3. Read about documentaries on race 4. Sign a private petition on racial justice in policing	1. Browse Netflix's general collection 2. Watch the trailer for "Making a Murderer" 3. Read about documentaries on bizarre stories 4. Sign a private petition on a humanitarian airlift to Ukraine
Public contest	1. Share Netflix's BLM collection on social media 2. Share the trailer for "13th" 3. Share an article about documentaries on race 4. Sign a public petition on racial justice in policing	1. Share Netflix's general collection on social media 2. Share the trailer for "Making a Murderer" 3. Share an article about documentaries on bizarre stories 4. Sign a public petition on a humanitarian airlift to Ukraine

Notes: Index is z-score. Regression includes vicarious contact sample.



(a) Private Contest



(b) Public Contest

Figure C.3: Gleam public and private contest images

D Randomization

As shown in Figure B.1, the design involved three cross-cutting experiments: the media contact experiment, the virtual enumerator experiment, and the Gleam contest experiment.

Details on the contact experiment. All respondents were randomly assigned with equal probability to one of three arms: (1) control (ingroup parasocial contact), (2) parasocial contact treatment (outgroup parasocial contact), and (3) vicarious contact treatment (ingroup parasocial contact plus ingroup-outgroup interaction). This top-level randomization is shown in Figure D.1, Panel A, along with realized sample sizes. As discussed in Section 4, respondents assigned to the vicarious contact treatment are retained in the anlaysis but results for this treatment are not analyzed here.

Conditional on initial assignment, participants were also assigned with equal probability to one of two couples as shown in Panel B. Those assigned to the control were randomly assigned to either ingroup Couple A or C; those assignment to parasocial contact treatment were randomly assigned to either outgroup Couple B or D. Finally, participants assigned to Couple B were randomly assigned, with equal probability, to one of two versions that varied in whether they included brief clips in which the outgroup characters explicitly referenced "Black love."[2]

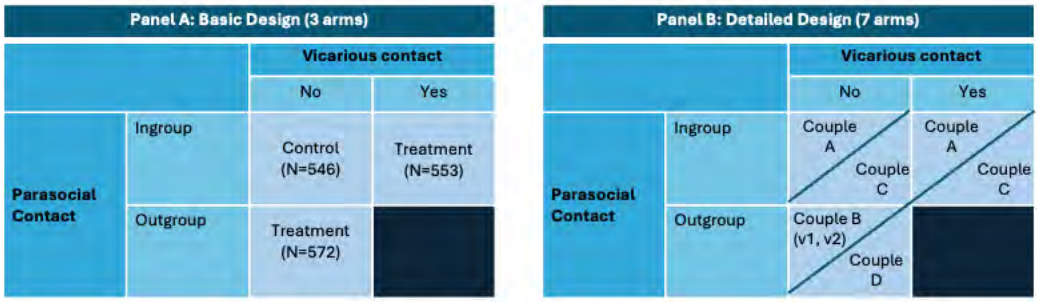


Figure D.1: Basic Experimental Design

Randomization profiles. To obtain optimal balance across the three experiments, 48 unique randomization profiles were created, as shown in Table D.1. This table shows that participants were also assigned with equal probability to public and private conditions in the virtual enumerator experiment and Gleam contest experiment.[3] All assignments were fully orthogonal.

Participants were assigned to profiles using block randomization within survey waves. As each wave (described below) began, incoming participants were assigned to a randomly drawn profile without replacement. Once all profiles had been used, a new cycle began. This approach guaranteed that each wave contained a near-even distribution of all treatment combinations, improving covariate balance and ensuring that treatment effects could be estimated within waves.

Implementation waves. The study was fielded in 12 recruitment waves between October 9–27, 2023, targeting 96–144 participants per wave. The main reason for the waves was to allow the reordering of Gleam contest actions, as discussed in Appendix Section C.2. These targets were large enough to assign each of the 48 profiles to 2–3 participants per wave, while small enough to permit re-randomization of the order in which Gleam actions appeared over the course of implementation. Within each wave, survey launch times were also randomized to ensure participants in different U.S. time zones had a chance to enter before the wave's

[2] This variation reflected the characters' actual storyline, but there was *ex ante* uncertainty about whether explicit references to race would have positive or negative effects on white viewers (Bond, 2020).

[3] There were some additional random assignments related to post-treatment vignette experiments, as specified in our pre-analysis plan. All of these were done orthogonally and are controlled for in the analysis.

quota filled (e.g. starting at 9 a.m. ET could otherwise fill the quota before many West Coast participants began).

Analysis implications. In all analyses, fixed effects for both wave and show season are included. Season fixed effects are important because the parasocial contact treatment featured two outgroup couples per season. This ensures comparisons are made within season (e.g., Couple A vs. Couple B; Couple C vs. Couple D) rather than across seasons, isolating treatment effects from any differences in show content or production. This is relevant as one of the seasons occurred during the Covid-19 lockdown and includes some production differences.

Table D.1: Randomization Profiles

Resp No.	Treat	Para	Vic	Treat	Enum	Gleam
1	Control	0	0	Couple A	0	0
2	Control	0	0	Couple A	0	1
3	Control	0	0	Couple A	0	0
4	Control	0	0	Couple A	0	1
5	Control	0	0	Couple A	1	0
6	Control	0	0	Couple A	1	1
7	Control	0	0	Couple A	1	0
8	Control	0	0	Couple A	1	1
9	Control	0	0	Couple C	0	0
10	Control	0	0	Couple C	0	1
11	Control	0	0	Couple C	0	0
12	Control	0	0	Couple C	0	1
13	Control	0	0	Couple C	1	0
14	Control	0	0	Couple C	1	1
15	Control	0	0	Couple C	1	0
16	Control	0	0	Couple C	1	1
17	Vic	0	1	Couple A	0	0
18	Vic	0	1	Couple A	0	1
19	Vic	0	1	Couple A	0	0
20	Vic	0	1	Couple A	0	1
21	Vic	0	1	Couple A	1	0
22	Vic	0	1	Couple A	1	1
23	Vic	0	1	Couple A	1	0
24	Vic	0	1	Couple A	1	1
25	Vic	0	1	Couple C	0	0
26	Vic	0	1	Couple C	0	1
27	Vic	0	1	Couple C	0	0
28	Vic	0	1	Couple C	0	1
29	Vic	0	1	Couple C	1	0
30	Vic	0	1	Couple C	1	1
31	Vic	0	1	Couple C	1	0
32	Vic	0	1	Couple C	1	1
33	Para	1	0	Couple B (Black love)	0	0
34	Para	1	0	Couple B (Black love)	0	1
35	Para	1	0	Couple B (Black love)	1	0
36	Para	1	0	Couple B (Black love)	1	1
37	Para	1	0	Couple B (No Bl love)	0	0
38	Para	1	0	Couple B (No Bl love)	0	1
39	Para	1	0	Couple B (No Bl love)	1	0
40	Para	1	0	Couple B (No Bl love)	1	1
41	Para	1	0	Couple D	0	0
42	Para	1	0	Couple D	0	1
43	Para	1	0	Couple D	0	0
44	Para	1	0	Couple D	0	1
45	Para	1	0	Couple D	1	0
46	Para	1	0	Couple D	1	1
47	Para	1	0	Couple D	1	0
48	Para	1	0	Couple D	1	1

Notes: Table shows the 48 randomization profiles created to obtain optimal balance across the contact, enumerator, and gleam contest experiments. Respondents were randomly assigned to one profile without replacement; once all profiles were exhausted, a new block was initiated.

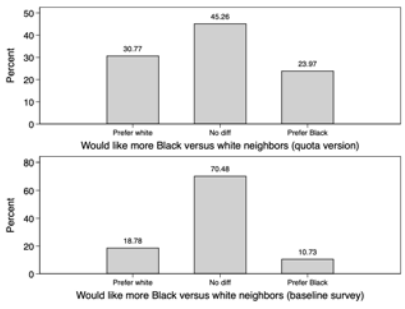
E Summary Statistics

Table E.1: Summary Statistics Baseline Variables

	Min	Max	Mean	SD	N	Notes on scales
I. Demographics						
Age	18	49	34.86	7.37	1118	Numeric
Female (%)	0	1	0.49	0.50	1118	1=female, 0 otherwise
Education completed	1	6	4.39	1.07	1118	1=primary school, 6=graduate school
Income	1	26	18.98	4.91	1118	1=¡$1,000, 26=¿$170,000
Urban (%)	0	1	0.25	0.44	1118	1=urban, 0=rural, semi-urban
Party ID (%)	0	1	0.23	0.42	1118	1=Republican, 0 otherwise
Share of network married	1	5	3.00	1.14	1118	1=almost none, 5=almost all
Like reality tv	1	5	3.35	1.25	1118	1=dislike a lot, 5=like a lot
II. Region						
Region: Southeast (%)	0	1	0.06	0.24	1118	1=living in region, 0 otherwise
Region: New England (%)	0	1	0.04	0.20	1118	1=living in region, 0 otherwise
Region: Northeast (%)	0	1	0.16	0.37	1118	1=living in region, 0 otherwise
Region: Midwest East (%)	0	1	0.18	0.38	1118	1=living in region, 0 otherwise
Region: Midwest West (%)	0	1	0.07	0.25	1118	1=living in region, 0 otherwise
Region: Southatlantic (%)	0	1	0.22	0.41	1118	1=living in region, 0 otherwise
Region: South Central (%)	0	1	0.09	0.29	1118	1=living in region, 0 otherwise
Region: Mountain (%)	0	1	0.07	0.25	1118	1=living in region, 0 otherwise
Region: Pacific (%)	0	1	0.11	0.31	1118	1=living in region, 0 otherwise
III. Social Media						
Index social media	-2.21	1.51	0	1	1118	ICW index
Facebook use	1	7	4.86	2.25	1118	1=never, 7=multiple times a day
Twitter/X use	1	7	4.10	2.37	1118	1=never, 7=multiple times a day
Instagram use	1	7	4.73	2.23	1118	1=never, 7=multiple times a day
IV. Previous media contact						
Index outgroup shows	-0.61	3.43	0.04	1.00	1118	ICW index
Black-ish	-1	1	-0.51	0.85	1118	1=yes, 0=maybe/don't know, -1=no
Dear White People	-1	1	-0.83	0.54	1118	1=yes, 0=maybe/don't know, -1=no
Abbott Elementary	-1	1	-0.57	0.82	1118	1=yes, 0=maybe/don't know, -1=no
V. Tolerance (not prejudiced)						
Like more white neighbors	1	5	3.78	0.89	1118	1=dislike a lot, 5=like a lot
Index like outgroup (excl Black) (z-score)	-3.12	1.48	-0.03	0.99	1118	ICW index
Like more Asian neighbors	1	5	3.78	0.87	1118	1=dislike a lot, 5=like a lot
Like more Latino neighbors	1	5	3.61	0.94	1118	1=dislike a lot, 5=like a lot
Index like outgoup (Black)	-2.95	1.97	-0.02	0.98	1118	ICW index
Like more Black neighbors (quota version)	1	5	3.66	0.98	1118	1=dislike a lot, 5=like a lot
Like more Black neighbors (baseline survey)	1	5	3.59	0.98	1118	1=dislike a lot, 5=like a lot
Like more Black vs. white neighbors (baseline survey)	1	3	1.92	0.54	1118	1=prefer white, 2=neutral, 3=prefer Black
VI. Social						
Share of network that is Black	1	5	1.74	0.73	1118	1=almost none, 5=almost all
Share of network that is white	1	5	4.09	0.84	1118	1=almost none, 5=almost all
Homogeneous ingroup network (%)*	0	1	0.78	0.41	1118	1=more than half of network is white, 0 otherwise
Uncomfortable with ingroup disapproval	1	4	2.48	0.81	1118	1=very comfortable, 4=very uncomfortable
VII. Ingroup norms support prejudice						
Index norms neighbors	-2.37	2.34	0.02	0.02	1118	ICW index
Share of ingroup in general wanting Black neighbors	1	5	3.25	0.85	1118	1=almost all, 5=almost none
Share of white family members wanting Black neighbors	1	5	3.07	1.13	1118	1=almost all, 5=almost none
Share of white friends wanting Black neighbors	1	5	2.67	1.08	1118	1=almost all, 5=almost none
Index norms insensitive comments	-1.27	2.74	0.08	1.05	1118	ICW index
Ingroup (general) attitudes towards racial comments	1	5	2.23	1.17	1118	1=very unacceptable, 5=very acceptable
White family members attitudes towards racial comments	1	5	2.52	1.46	1118	1=very unacceptable, 5=very acceptable
White friends attitudes towards racial comments	1	5	2.31	1.45	1118	1=very unacceptable, 5=very acceptable
Prejudiced norms (%)*	0	1	0.49	0.50	1118	1=above median on factor score combining six components
VIII. High social cost moderator						
High social costs	0	1	0.40	0.49	1118	1=prej. norms + homog. network, 0 otherwise

Notes: The table reports summary statistics (minimum, maximum, mean, standard deviation, and sample size) for all pre-treatment variables in the parasocial contact sample. Variables are presented in their original scales, though some have been recoded for consistency in directionality. Asterisks (*) indicate variables used to construct the main high social cost moderator.

Figure E.1: Personal Prejudice



Notes: Figure shows response distributions for the difference in wanting more Black versus white neighbors as measured for quota sampling (top figure) and then repeated during the baseline survey (bottom figure).

Table E.2: Summary Statistics Outcomes and Mechanisms

	min	max	mean	sd	N	Notes on scales
I. Main Attitudinal Outcomes						
1a. Warmth index (outgroup)	-4.13	1.51	0.02	1.01	1037	ICW based on outgroup variables
1b. Warmth index (diff)	-4.95	2.63	0.01	1.03	1037	ICW index based on 'difference' variables
Thermometer						
Outgroup	0	10	7.02	2.19	1037	0=extremely cold, 10=extremely warm
Ingroup	0	10	7.03	2.07	1037	0=extremely cold, 10=extremely warm
Diff	-10	9	-0.02	2.42	1037	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
Friends						
Want more outgroup friends	1	4	2.79	0.69	1037	4=want a lot, 1=don't want at all
Want more ingroup friends	1	4	2.69	0.71	1037	4=want a lot, 1=don't want at all
Diff	-3	3	0.10	0.80	1037	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
Support intermarriage	1	5	4.54	0.90	1037	5=very willing to support, 1=very unwilling
2. Support social change	1	3	2.47	0.73	1037	3=not gone far enough, 2=been about right, 1=gone too far
II. Mechanisms						
3a. Positive attributes index (outgroup) (excl. 'not angry')	-3.95	2.28	0.02	1.02	1035	ICW based on outgroup variables, excludes 'not angry' component
3b. Positive attributes index (diff) (excl. 'not angry')	-5.23	4.78	0.02	1.04	1034	ICW based on 'difference' variables, excludes 'not angry' component
Positive attributes index (ingroup) (excl. 'not angry')	-3.68	2.77	0.01	1.01	1035	ICW based on ingroup variables, excludes 'not angry' component
Positive attributes index (outgroup) (all)	0.01	1.02	-4.11	2.42	1035	ICW based on outgroup variables, all components
Proportion hard working						
Outgroup	0	10	6.41	1.96	1035	0=none, 10=all
Ingroup	0	10	6.36	1.67	1035	0=none, 10=all
Diff	-8	9	0.04	1.89	1034	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
Proportion intelligent						
Outgroup	0	10	6.54	1.89	1035	0=none, 10=all
Ingroup	0	10	6.29	1.80	1035	0=none, 10=all
Diff	-7	7	0.24	1.62	1034	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
Proportion strong family values						
Outgroup	0	10	6.25	2.20	1035	0=none, 10=all
Ingroup	0	10	6.01	1.98	1035	0=none, 10=all
Diff	-9	10	0.23	2.30	1034	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
Proportion want commitment						
Outgroup	0	10	6.37	2.05	1035	0=none, 10=all
Ingroup	0	10	6.70	1.78	1035	0=none, 10=all
Diff	-8	9	-0.33	1.73	1034	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
Proportion not angry/violent						
Outgroup	0	10	6.06	2.11	1035	0=none, 10=all
Ingroup	0	10	5.84	2.09	1035	0=none, 10=all
Diff	-9	9	0.21	2.12	1034	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
4. Similarity index	-2.94	2.30	-0.01	1.00	1035	ICW index
Perceived similarity (general)	1	5	3.07	0.78	1035	1=strongly disagree, 5=strongly agree
Perceived similarity (values)	1	5	3.50	1.11	1035	5=very similar, 1=very different
5. Empathy index	-3.04	1.73	0.00	0.98	1034	ICW index
Empathy (general)	1	5	3.28	1.03	1034	5=very easy, 1=very hard
Empathy (policing)	1	7	5.25	1.64	1034	5=very easy, 1=very hard
6. Comfort around (diff)	-4	3	-0.17	0.72	1037	Diff: Outgroup-ingroup rating, higher=outgroup rated more positively
Outgroup strangers	1	5	3.79	1.12	1037	1=very uncomfortable, 5=very comfortable
Ingroup strangers	1	5	3.96	1.06	1037	1=very uncomfortable, 5=very comfortable
7. Perceived Ingroup Norms	-3.04	2.55	0.02	1.01	1035	ICW index
General	1	4	2.63	0.59	1035	4=want a lot, 1=don't want at all
Family	1	4	2.80	0.65	1035	4=want a lot, 1=don't want at all
Friends	1	4	2.46	0.73	1035	4=want a lot, 1=don't want at all
III. Behavioral Intentions						
8. Action index	-2.20	1.53	0.02	0.98	1037	ICW index
8a. Private action index	-2.30	0.98	0.01	0.99	1037	ICW index
Vote Black candidate	1	5	3.73	1.39	1037	5=very likely, 1=very unlikely
Vote White candidate	1	5	3.68	1.34	1037	5=very likely, 1=very unlikely
Read an article	1	5	3.96	1.31	1037	5=very likely, 1=very unlikely
Sign a private petition	1	5	3.77	1.46	1037	5=very likely, 1=very unlikely
8b. Public action index	-1.97	1.39	0.03	0.99	1037	ICW index
Confront friends	1	5	2.99	1.50	1037	5=very likely, 1=very unlikely
Public protest	1	5	3.77	1.31	1037	5=very likely, 1=very unlikely
Share an article	1	5	3.65	1.47	1037	5=very likely, 1=very unlikely
Sign a public petition	1	5	2.86	1.50	1037	5=very likely, 1=very unlikely
IV: Actual Behavior (Gleam Contest)						
9. Entered contest	0	1	0.50	0.50	1118	1=entered contest, 0 otherwise
10. Actions difference	-2	3	-0.01	0.26	1118	Difference (Race - non-race)
Total race-related actions	0	4	0.50	1.21	1118	Total race-related actions taken
Total non-race-related actions	0	4	0.51	1.21	1118	Total non-race-related actions taken
10a. Private actions (difference)	-2	3	-0.01	0.33	555	Difference: Private race minus non-race-related actions
Total race-related private actions	0	4	0.90	1.51	555	Total private race-related actions taken
Netflix BLM	0	1	0.24	0.43	555	1=took action, 0 otherwise
Documentary	0	1	0.24	0.43	555	1=took action, 0 otherwise
Trailer	0	1	0.26	0.44	555	1=took action, 0 otherwise
Petition	0	1	0.16	0.37	555	1=took action, 0 otherwise
Total non-race related private actions	0	4	0.91	1.51	555	Total private non-race-related actions taken
Netflix BLM	0	1	0.24	0.43	555	1=took action, 0 otherwise
Documentary	0	1	0.24	0.43	555	1=took action, 0 otherwise
Trailer	0	1	0.25	0.43	555	1=took action, 0 otherwise
Petition	0	1	0.17	0.38	555	1=took action, 0 otherwise
10b. Public actions (difference)	-1	1	0.00	0.14	563	Difference: Public race minus non-race related actions
Total race-related public actions	0	4	0.12	0.58	563	Total public race-related actions taken
Netflix BLM	0	1	0.02	0.16	563	1=took action, 0 otherwise
Documentary	0	1	0.03	0.17	563	1=took action, 0 otherwise
Trailer	0	1	0.04	0.19	563	1=took action, 0 otherwise
Petition	0	1	0.03	0.16	563	1=took action, 0 otherwise
Total non-race related public actions	0	4	0.12	0.59	563	Total public non-race related actions taken
Netflix BLM	0	1	0.03	0.18	563	1=took action, 0 otherwise
Documentary	0	1	0.02	0.16	563	1=took action, 0 otherwise
Trailer	0	1	0.03	0.18	563	1=took action, 0 otherwise
Petition	0	1	0.03	0.17	563	1=took action, 0 otherwise

Notes: Table shows summary statistics (min, max, mean, standard deviation, sample size) for the parasocial contact sample for main outcome and mechanism measures.

F Balance

Table F.1: Balance

		PANEL A: CONTACT EXPERIMENTS									PANEL B: PUBLIC-PRIVATE EXPERIMENTS									
		Control	Parasocial Contact				Vicarious Contact				Control	Enum (Public)				Control	Gleam (Public)			
		mean	b	se	p	Count	b	se	p	Count	mean	b	se	p	Count	Mean	b	se	p	Count
1	Age	34.92	-0.11	0.44	0.804	0	0.49	0.45	0.269	0	34.90	-0.07	0.44	0.877	0	34.98	-0.23	0.44	0.598	0
2	Female (%)	0.53	-0.07	0.03	0.024	1	-0.02	0.03	0.412	0	0.47	0.04	0.03	0.187	0	0.49	0.01	0.03	0.854	0
3	Education	4.38	0.02	0.06	0.702	0	0.08	0.07	0.228	0	4.42	-0.05	0.06	0.461	0	4.40	-0.02	0.06	0.713	0
4	Income	19.08	-0.18	0.29	0.545	0	-0.04	0.30	0.896	0	19.09	-0.22	0.29	0.458	0	19.17	-0.38	0.29	0.197	0
5	Urban	0.26	0.00	0.03	0.859	0	0.01	0.03	0.671	0	0.24	0.04	0.03	0.160	0	0.27	-0.04	0.03	0.130	0
6	Party ID (% Republication)	0.23	0.01	0.03	0.775	0	-0.01	0.03	0.586	0	0.24	-0.02	0.03	0.419	0	0.24	-0.02	0.03	0.401	0
7	Share of network married	3.01	-0.01	0.07	0.832	0	0.07	0.07	0.299	0	3.03	-0.05	0.07	0.434	0	3.03	-0.07	0.07	0.321	0
8	Like reality tv	3.33	0.05	0.07	0.528	0	0.04	0.07	0.595	0	3.33	0.06	0.07	0.459	0	3.37	-0.02	0.07	0.740	0
9	Region: Southeast	0.07	-0.02	0.01	0.238	0	-0.01	0.01	0.667	0	0.06	0.00	0.01	0.841	0	0.06	0.00	0.01	0.853	0
10	Region: New England	0.04	0.01	0.01	0.463	0	0.00	0.01	0.968	0	0.04	0.00	0.01	0.970	0	0.04	0.00	0.01	0.960	0
11	Region: Northeast	0.17	-0.01	0.02	0.551	0	-0.03	0.02	0.211	0	0.18	-0.03	0.02	0.259	0	0.18	-0.03	0.02	0.248	0
12	Region: Midwest East	0.19	-0.02	0.02	0.288	0	-0.01	0.02	0.677	0	0.17	0.03	0.02	0.215	0	0.19	-0.03	0.02	0.201	0
13	Region: Midwest West	0.06	0.02	0.02	0.274	0	0.01	0.02	0.369	0	0.08	-0.02	0.01	0.261	0	0.08	-0.02	0.01	0.255	0
14	Region: Southatlantic	0.22	0.00	0.02	0.851	0	0.02	0.03	0.541	0	0.23	-0.03	0.02	0.184	0	0.19	0.05	0.02	0.047	1
15	Region: South Central	0.09	0.00	0.02	0.869	0	-0.01	0.02	0.627	0	0.09	0.01	0.02	0.576	0	0.10	-0.02	0.02	0.269	0
16	Region: Mountain	0.06	0.02	0.01	0.262	0	0.01	0.01	0.499	0	0.07	0.00	0.01	0.866	0	0.05	0.03	0.01	0.080	0
17	Region: Pacific	0.10	0.01	0.02	0.440	0	0.01	0.02	0.489	0	0.09	0.04	0.02	0.052	0	0.10	0.01	0.02	0.438	0
18	Index social media	0.00	0.06	0.06	0.327	0	0.04	0.06	0.512	0	0.03	0.01	0.06	0.926	0	0.05	-0.03	0.06	0.558	0
19	Like ingroup	3.76	0.03	0.05	0.530	0	0.08	0.05	0.144	0	3.80	-0.04	0.05	0.427	0	3.79	-0.02	0.05	0.657	0
20	Index like outgroup (excl. Black)	0.00	0.01	0.06	0.869	0	0.09	0.06	0.148	0	0.03	-0.04	0.06	0.479	0	0.01	-0.02	0.06	0.767	0
21	Index like outgroup (Black)	0.00	-0.03	0.06	0.663	0	0.03	0.06	0.593	0	-0.01	-0.01	0.06	0.858	0	-0.06	0.09	0.06	0.129	0
22	Share of network Black	1.74	-0.01	0.04	0.833	0	0.02	0.04	0.654	0	1.70	0.07	0.04	0.121	0	1.74	0.00	0.04	0.938	0
23	Share of network white	4.09	0.01	0.05	0.871	0	-0.01	0.05	0.869	0	4.14	-0.09	0.05	0.068	0	4.09	0.01	0.05	0.825	0
24	Index norms neighbors	0.00	0.03	0.06	0.614	0	0.08	0.06	0.173	0	0.00	0.04	0.06	0.533	0	0.05	-0.07	0.06	0.245	0
25	Index norms joke	0.00	0.16	0.06	0.009	1	0.05	0.06	0.426	0	0.06	0.05	0.06	0.442	0	0.07	0.03	0.06	0.641	0
26	Fear ingroup sanctioning	2.55	-0.15	0.05	0.002	1	-0.12	0.05	0.017	1	2.47	0.02	0.05	0.629	0	2.47	0.01	0.05	0.820	0
27	Index previous media contact	0.00	0.07	0.06	0.232	0	-0.03	0.06	0.556	0	0.00	0.07	0.06	0.269	0	-0.02	0.11	0.06	0.072	0
28	Enumerator treatment	0.50	0.01	0.03	0.861	0	-0.01	0.03	0.834	0	0.00	.	.	.	.	0.50	0.00	0.03	0.999	0
29	Male enum	0.23	0.03	0.03	0.222	0	0.04	0.03	0.165	0	0.00	.	.	.	.	0.26	-0.03	0.03	0.343	0
30	Fem enum	0.25	-0.02	0.03	0.380	0	-0.04	0.03	0.105	0	0.00	.	.	.	.	0.22	0.02	0.03	0.450	0
31	Black Mathises (survey experiment)	0.33	-0.03	0.03	0.367	0	0.03	0.03	0.277	0	0.30	0.03	0.03	0.364	0	0.30	0.02	0.03	0.460	0
32	Black police officer (survey experiment)	0.34	0.00	0.03	0.944	0	-0.02	0.03	0.524	0	0.37	-0.05	0.03	0.101	0	0.36	-0.03	0.03	0.238	0
33	Black mayor (survey experiment)	0.50	0.02	0.03	0.605	0	0.02	0.03	0.616	0	0.50	0.01	0.03	0.730	0	0.47	0.06	0.03	0.051	0
34	Gleam public contest	0.50	0.00	0.03	0.909	0	0.01	0.03	0.788	0	0.50	0.00	0.03	0.999	0	.	.	.	.	.
						3				1					0					1
					%	0.09			%	0.03				%	0.00				%	0.03

Notes: Table presents balance tests for the parasocial and vicarious contact experiments (Panel A) as well as for the enumerator and Gleam contest experiments (Panel B). Control mean refers to the raw mean in the respective control groups. Coefficients, standard errors, and p-values are obtained from separate regressions of each covariates on the two treatment indicators in the full sample. The count column is 1 if the p-value is $\leq .05$.

G Attrition and Attention

Table G.1 reports tests of the parasocial contact and virtual enumerator experiments on a battery of attrition and attention questions.

Attrition: Attrition was low overall. In the contact experiment, 92 percent of control participants reached the contest (column 1) and 87 percent completed the study (column 2). There is no evidence that assignment to outgroup parasocial contact affected the likelihood of reaching the contest or the end of the study. Similar patterns hold for the enumerator experiment.

Attention: Attention was high across all groups. In the parasocial contact control group, participants answered 6.44 (out of 7) attention checks correctly (column 3). There is no evidence that assignment to the parasocial contact treatment group had any effect on treatment or survey attention checks (Panel I, columns 3-6). Equally important, assignment to the public condition in the enumerator experiment had no effect on attention to the survey (Panel II, columns 3-6).

Enumerator checks: Columns 7–9 report the effects of assignment to parasocial contact and to the public condition on correctly identifying the sex and race of the virtual enumerator. (In columns 7-8, a response is coded as correct if respondents assigned to the public condition correctly reported the enumerator's characteristics, and if respondents assigned to the private condition correctly reported that no enumerator was present). The most notable results are in column 9, which shows that only four percent of respondents in the private condition reported having a white enumerator. Assignment to the public condition increased this share by 93 percentage points, meaning that approximately 97 percent of respondents in the public condition reported having a white enumerator. These results indicate that participants noticed and registered the presence and race of the virtual enumerator.

Duration: In the parasocial contact experiment, the study duration for the control group (excluding outliers) was about 49 minutes. The contact treatment increased this by over five minutes, but this did not lead to higher attrition or reduced attention as shown in columns 1-9. The enumerator experiment also had no detectable effect on duration.

Survey purpose: Despite efforts to mask the study's intent, 61 percent of respondents in the contact control group correctly identified that it concerned racial attitudes. Assignment to parasocial contact treatment increased this by 19 percentage points (column 12). As further discussed in Appendix J.2, analysis of treatment effects by outgroup couple suggests that this awareness did not introduce experimenter demand effects. Assignment to the public condition in the virtual enumerator experiment had no detectable effect on guessing the study purpose.

Table G.1: Attrition and Attention Checks

	Attrition		Attention Treatment/Survey				Attention Enumerator Exp.			Duration		Reaction	
	Reached contest (1)	Reached end (2)	Total correct (3)	Survey correct (4)	Treat correct (5)	Final chk correct (6)	Sex correct (7)	Race correct (8)	Enum white (9)	Minutes w/outlier (10)	Minutes no outliers (11)	Survey purpose (12)	Reaction overall (13)
I. ATE: Parasocial Contact													
Contact	0.00	0.02	-0.03	0.00	-0.03	0.00	0.02	0.02	0.00	8.48**	5.35***	0.19***	0.01
se	(0.01)	(0.01)	(0.05)	(0.01)	(0.04)	(0.01)	(0.01)	(0.01)	(0.03)	(3.51)	(1.21)	(0.03)	(0.02)
pval	0.600	0.119	0.518	0.701	0.470	0.995	0.187	0.131	0.940	0.016	0.000	0.000	0.624
control mean	0.92	0.87	6.44	2.99	2.46	0.99	0.96	0.96	0.49	51.09	49.29	0.61	0.80
II. ATE: Public Condition, Enumerator Experiment													
Public	0.00	-0.01	-0.05	-0.01	-0.05	0.00	0.03***	0.00	0.93***	-4.05	-1.08	0.00	0.00
se	(0.00)	(0.01)	(0.04)	(0.01)	(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(3.72)	(0.95)	(0.02)	(0.02)
pval	0.844	0.282	0.177	0.212	0.151	0.314	0.005	0.944	0.000	0.277	0.254	0.864	0.870
control mean	0.94	0.89	6.46	2.99	2.47	0.99	0.94	0.96	0.04	53.64	50.04	0.64	0.79
N (Full)	1550	1550	1539	1545	1550	1539	1539	1538	1538	1550	1545	1538	1537
N (Para)	1118	1118	1034	1038	1048	1034	1034	1033	1033	1118	1113	1033	1032

Notes: Table reports regressions of attrition and attention check outcomes on indicators for assignment to the parasocial contact treatment and the public condition in the enumerator experiment. Binary outcomes include: *Reached contest*, *Reached end*, *Final check correct*, *Sex correct*, *Race correct*, *Enumerator white*, *Survey purpose*, and *Reaction overall*. *Total correct* (max = 7), *Survey correct* (max = 3), and *Treatment correct* (max = 3) are integer scores. *Duration* is measured in minutes.

H Social Costs

Concept. Testing heterogeneity in contact effects by visibility and social cost is central to both the theoretical framework and the empirical analysis. The pre-analysis plan specified that respondents would be classified as facing high social costs when three conditions were met: (1) they were embedded in an ingroup-homogeneous social network, (2) they perceived ingroup norms as condoning outgroup prejudice, and (3) they were personally sensitive to social sanctioning. For reasons discussed in Appendix O, the operational measure of social costs used in this paper incorporates only the first two components, each coded as a binary indicator. Respondents are classified as facing high social costs when both indicators are present.

Network homogeneity. Network homogeneity is coded as high if respondents report that most of the people they spend time with are white, a criterion met by 78 percent of the sample.

Perceived ingroup norms measures. To assess perceived ingroup norms, I use two sets of three questions. One set asked participants to estimate how many "white people in general/of your white family members/of your white friends" would like having more Black neighbors. The second set asked "how acceptable would it be, if at all, to make a racially insensitive comment or joke about Black people to [most White people in this country/your White family members/your White friends]." Figure H.1 shows the distribution for the six items.

Factor analysis. The main measure of prejudiced ingroup norms is constructed using factor analysis, which generates a binary indicator based on a latent trait derived from six theoretically related items. Factor analysis is well-suited to this task, as it captures the shared variance across items while down-weighting item-specific measurement error. This approach appeared to yield a more reliable and parsimonious measure than the inverse-covariance-weighted (ICW) index. Because this moderator is central to the heterogeneity analysis, minimizing measurement error was a key priority. Nevertheless, I assess the robustness of results using alternative high social cost indicators based on ICW indices. The main findings are largely robust to these alternative specifications.

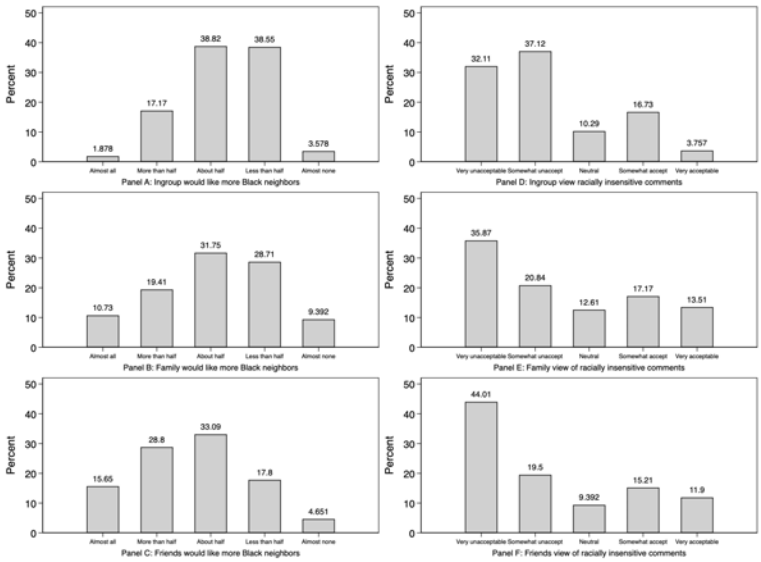
The factor analysis produced an eigenvalue of 1.92 for the first factor and 0.70 for the second, supporting a one-factor solution. Figure H.2 shows the distribution of the resulting factor scores, and exhibits a mostly normal distribution with a slight left skew. To maximize statistical power, I dichotomize the measure by coding individuals as having prejudiced ingroup norms if their factor score falls above the sample median.

Composite binary indicator. Combining the homogeneous network and perceived ingroup norms measures into a composite indicator of high social costs classifies approximately 40 percent of the sample as high-cost.

Correlates of high social costs. To validate these measures, Table H.1 presents regression results testing whether observed patterns align with theoretical expectations. For the main composite measure, personal affect toward Black individuals—as measured by "Index like outgroup (Black)"—is negatively associated with the high social cost measure: more tolerant individuals are less likely to perceive social norms as supporting prejudice. Reassuringly, the self-reported share of white individuals in one's network is positively correlated with living in a predominantly white census tract, based on 2020 U.S. census data linked to respondent location.

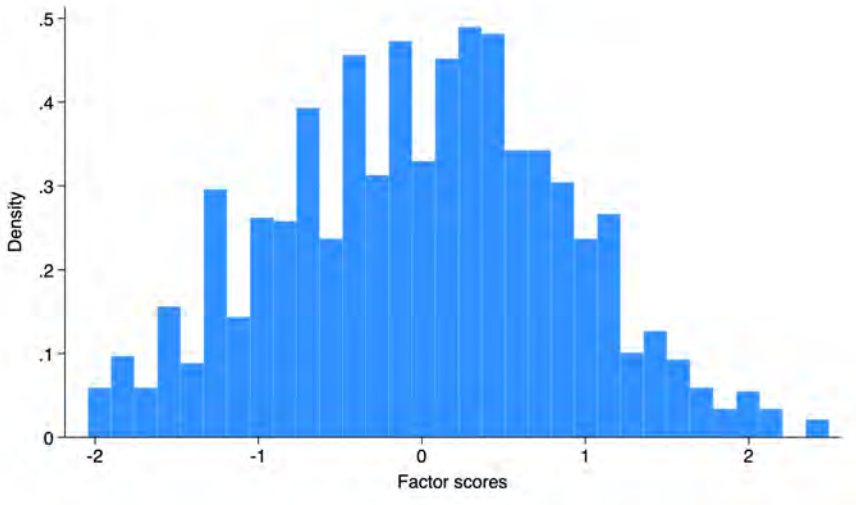
Convergence/divergence. Finally, the framework's emphasis on ex ante divergence between private and public outcomes finds empirical support in the relationship between private prejudice and perceived ingroup norms at baseline. Appendix Figure H.3 shows the joint distribution of baseline private prejudice and perceived ingroup norms. Although the two measures are positively correlated ($r = 0.29$), substantial misalignment exists: 17 percent of respondents hold tolerant private views while perceiving prejudiced ingroup norms (at risk of preference falsification), while 21 percent hold prejudiced private views while perceiving tolerant norms (at risk of misrepresentation). The remaining respondents exhibit convergence, with 44 percent showing low private prejudice and perceiving tolerant norms, and 28 percent showing high private prejudice

Figure H.1: Perceived Social Norms



Notes: Figure displays the response distributions for six measures of perceived ingroup norms. The first three items (left column) capture the perceived share of ingroup members—general public, family, and friends—who would like more Black neighbors. The remaining three items (right column) capture perceptions of how acceptable racially insensitive jokes are to the same ingroup categories.

Figure H.2: Perceived Social Norms



Notes: Histogram showing the distribution of factor scores from a one-factor solution based on six items measuring prejudiced ingroup norms.

and perceiving prejudiced norms. This heterogeneity underscores that both convergence and divergence between private beliefs and perceived norms—and, by extension, between private and public outcomes—are empirically plausible prior to treatment.

Table H.1: Correlates of Social Costs

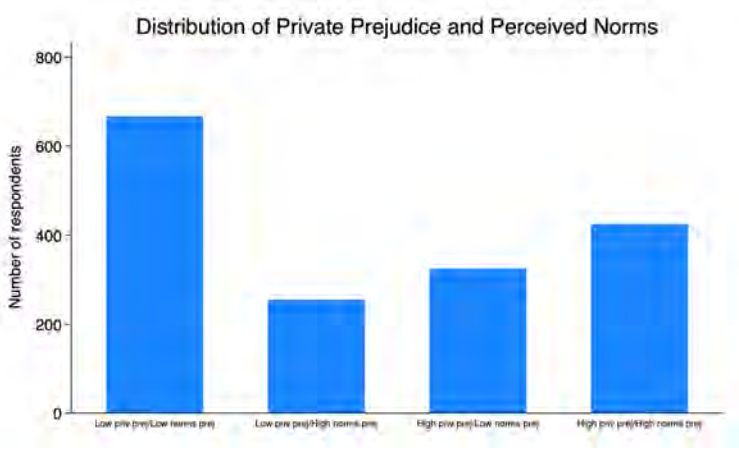
	Main Measure	Components		
	High Social Cost norms + homog network	Ingroup norms prejudiced	Network share white	
	b/se	b/se	b/se	b/se
Age	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Female (%)	-0.10***	-0.11***	0.06	0.07
	(0.03)	(0.03)	(0.04)	(0.05)
Education	0.01	0.01	-0.02	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)
Income	-0.01*	-0.01**	0.00	-0.01
	(0.00)	(0.00)	(0.00)	(0.01)
Urban	-0.01	0.04	-0.06	-0.06
	(0.03)	(0.03)	(0.05)	(0.06)
Republican (%)	0.04	0.02	0.00	0.02
	(0.03)	(0.03)	(0.05)	(0.06)
Share of network married	0.03**	0.00	0.13***	0.15***
	(0.01)	(0.01)	(0.02)	(0.02)
Like reality TV	0.03**	0.03**	-0.02	-0.02
	(0.01)	(0.01)	(0.02)	(0.02)
Region: Southeast	0.09	0.04	0.24**	0.28***
	(0.07)	(0.08)	(0.11)	(0.11)
Region: New England	-0.02	-0.07	0.45***	0.45***
	(0.07)	(0.07)	(0.11)	(0.14)
Region: Northeast	0.01	-0.05	0.21**	0.22**
	(0.05)	(0.05)	(0.09)	(0.09)
Region: Midwest East	0.05	0.01	0.26***	0.29***
	(0.05)	(0.05)	(0.09)	(0.09)
Region: Midwest West	0.13*	0.04	0.36***	0.41***
	(0.07)	(0.07)	(0.09)	(0.10)
Region: South Atlantic	0.11**	0.08	0.21**	0.22**
	(0.05)	(0.05)	(0.09)	(0.09)
Region: South Central	0.00	0.03	0.02	0.04
	(0.06)	(0.06)	(0.11)	(0.12)
Region: Mountain	-0.07	-0.02	-0.09	-0.05
	(0.07)	(0.07)	(0.11)	(0.12)
Index social media	0.00	0.01	-0.01	0.00
	(0.02)	(0.02)	(0.02)	(0.03)
Like white neighbors	0.01	-0.02	0.06**	0.05
	(0.02)	(0.02)	(0.03)	(0.03)
Index like outgroups (excluding Black)	-0.05**	-0.05**	-0.05	-0.04
	(0.02)	(0.02)	(0.04)	(0.04)
Index like outgroup (Black)	-0.13***	-0.15***	-0.01	-0.01
	(0.02)	(0.02)	(0.03)	(0.04)
Network: Share Black	-0.14***	-0.02	-0.61***	-0.62***
	(0.02)	(0.02)	(0.04)	(0.04)
Fear of ingroup sanctioning	-0.03*	-0.04**	0.01	0.01
	(0.02)	(0.02)	(0.03)	(0.03)
Like outgroup shows	0.02	0.00	0.06**	0.04
	(0.01)	(0.01)	(0.02)	(0.02)
Network: Share white		0.03		
		(0.02)		
Index norms neighbors			0.03	0.03
			(0.03)	(0.03)
Index norms racial jokes			0.01	0.01
			(0.02)	(0.02)
Census: Predominantly white				0.09*
				(0.05)
Constant	0.64	0.69	4.55	4.55
	(0.14)	(0.17)	(0.21)	(0.22)
N	1118	1118	1118	977

Notes: Table reports the correlates of the high social cost composite measure, as well as for its norms and network homogeneity components.

I Parasocial Contact Treatment Checks

Figure I.1 shows responses to a number of questions designed to capture whether the treatment created parasocial ties to the male and female characters. These questions were adapted from standard batteries of

Figure H.3: Perceived Social Norms



Notes: The figure shows the joint distribution of baseline private prejudice and perceived ingroup norms (both dichotomized).

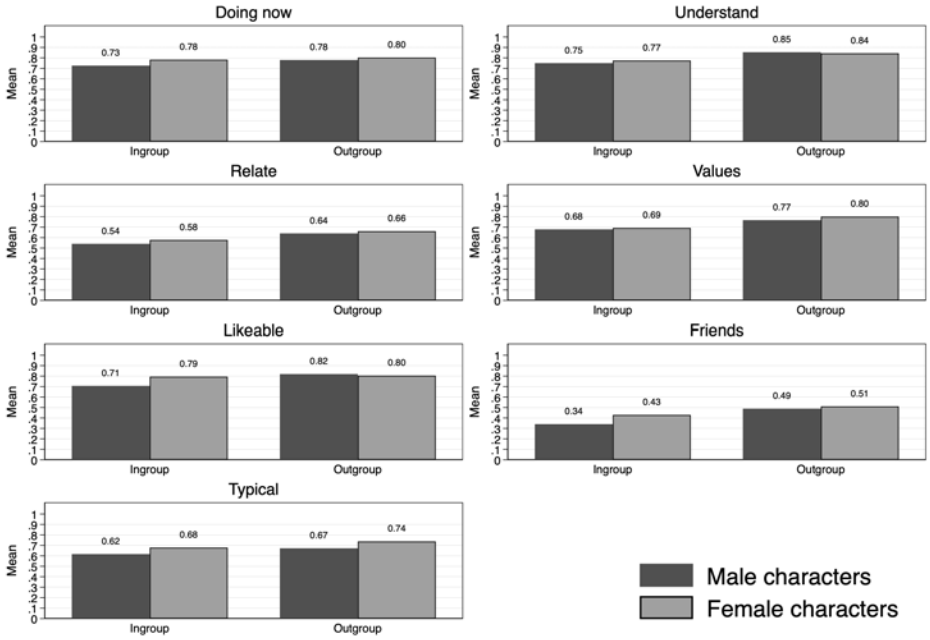
questions to measure parasocial relations and ties, as follows:

- To what extent, if at all, do you want to know how [NAME] is doing now that the show is over?
- To what extent, if at all, do you feel that you understand [NAME]?
- How much, if at all, did you relate to [NAME]?
- How much, if at all, would you say that [NAME] shares your values?
- Overall, do you think [NAME] is a likeable or unlikeable person?
- To what extent, if at all, could you see yourself being friends with [NAME]?
- How typical, if at all, is [MALE CHARACTER] of other [white/Black MEN] in this country?

Generally the responses were positive, were similar for ingroup and outgroup characters, and were similar for male and female characters.

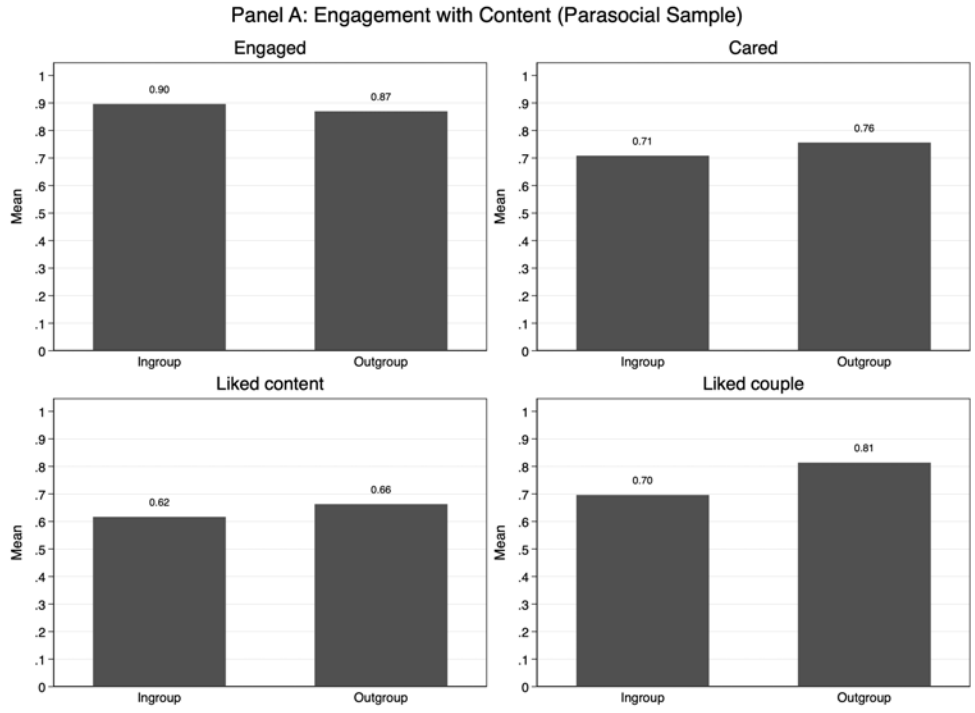
Figure I.2 shows reactions to four general measures of engagement with the content: engaged overall, cared about the couple, liked content, and liked couple. Overall, participants reacted with similar positivity to both ingroup and outgroup couples. If anything, they reacted more positively to the outgroup couples.

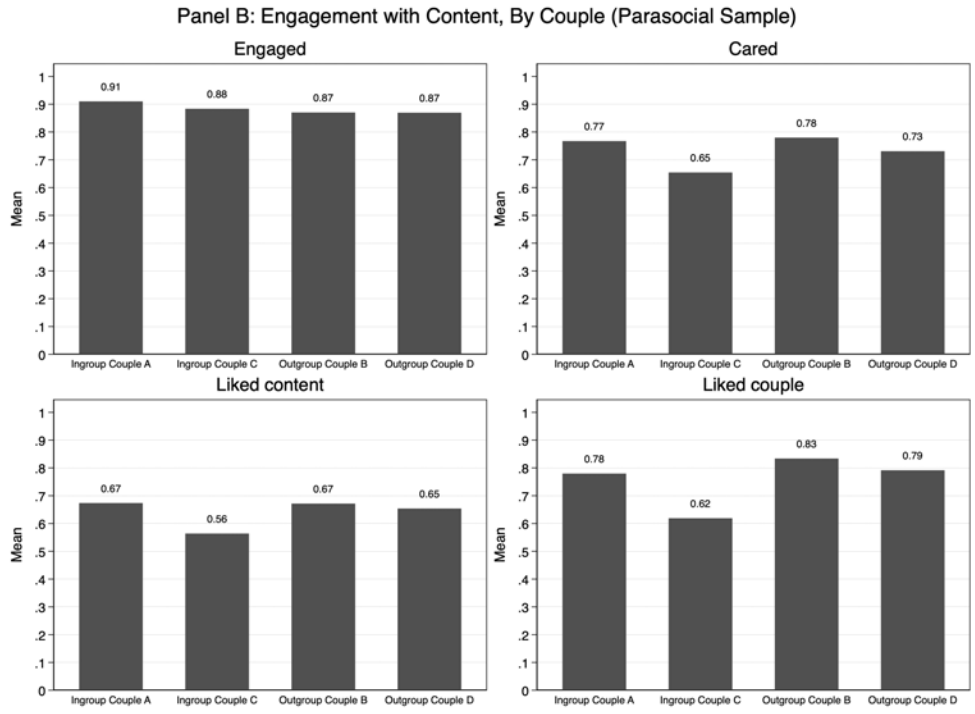
Figure I.1: Parasocial Relations for Ingroup/Outgroup Male and Female Characters



Notes: Figure shows responses to multiple measures of parasocial relations for the ingroup and outgroup male and female characters. Analysis is restricted to the parasocial contact sample

Figure I.2: Overall Engagement with Content





Notes: Figure shows participant engagement with the content for those assigned to outgroup and ingroup parasocial contact (Panel A) and by couple (Panel B). Individual figures display results for responses to survey questions measuring overall engagement, whether the respondent cared about the couple, whether they liked the content, and whether they liked the couple. Analysis is restricted to the parasocial contact sample (excludes those assigned to vicarious contact).

J Additional Analysis: Average Treatment Effects

J.1 Detailed Results

Table J.1 presents a more detailed version of the mechanism results shown in Table 1, showing the index components. The more detailed results support the exclusion of the "the not angry/violent" attribute from the main "positive attributes" indices. Parasocial contact appears to increase perceptions of Black Americans as not violent by about 0.20 units relative to white Americans ($p = .073$). However, the coefficients on the individual components indicate that this effect was driven by a perceived decline in favorability toward white Americans, which likely reflects a scene in which the ingroup couple argues. I therefore exclude this component from the main indices in the main text for both the "diff" and "outgroup" versions of the indices for consistency. Row 3c shows that the positive effect of parasocial contact on the positive attributes is robust when using the version of the "outgroup" index that incorporates all five attributes. Similarly, the main triple interaction results are robust regardless of whether using the positive attributes outgroup index that does and does not include this attribute (see columns 6 and 7 of Table K.3, respectively).

More generally, the results in Tables 1 and J.1 highlight the importance of focusing on the "outgroup" rather than the "difference" versions of the main measures. For more discussion of this reasoning, see Appendix O.

Table J.1: Average Treatment Effect on Mechanisms

	Control	Contact Effects			N	
	mean	coef	s.e.	p-val	Full	Para
Personal Beliefs Mechanisms						
3a. Positive attributes (outgroup) (excl. 'not angry')	0.00	0.11**	(0.05)	0.045	1540	1035
3b. Positive attributes (diff) (excl. 'not angry')	0.00	0.09*	(0.05)	0.082	1539	1034
Positive attributes (ingroup) (excl. 'not angry')	0.00	0.04	(0.06)	0.484	1540	1035
3c. Positive attributes (outgroup) (all)	0.00	0.09*	(0.05)	0.093	1540	1035
Not angry/violent						
Outgroup	6.09	0.05	(0.12)	0.661	1540	1035
Ingroup	5.92	-0.15	(0.13)	0.250	1540	1035
Diff	0.16	0.20*	(0.11)	0.073	1539	1034
4. Similarity index	0.00	0.03	(0.06)	0.579	1540	1035
Perceived similarity (general)	3.08	0.00	(0.05)	0.957	1540	1035
Perceived similarity (values)	3.50	0.06	(0.07)	0.375	1540	1035
5. Empathy index	0.00	0.02	(0.05)	0.645	1539	1034
Empathy (general)	3.30	-0.03	(0.06)	0.603	1539	1034
Empathy (policing)	5.24	0.12	(0.08)	0.159	1539	1034
6. Comfort around (diff)	-0.17	0.05	(0.04)	0.198	1542	1037
Outgroup strangers	3.78	0.06	(0.07)	0.383	1542	1037
Ingroup strangers	3.96	0.01	(0.07)	0.906	1542	1037
Perceived Ingroup Norms						
7. Perceived Ingroup Norms Index	0.00	0.07	(0.05)	0.205	1540	1035
General	2.62	0.04	(0.03)	0.251	1540	1035
Family	2.44	0.04	(0.04)	0.317	1540	1035
Friends	2.80	0.03	(0.03)	0.407	1540	1035

Notes: *p<.10, ** p<.05, *** p<.01. Table reports the effects of the contact treatment on pre-registered outcomes and mechanisms. Measures labeled "(difference)" represent the difference between ratings of Black and white individuals. All indices are inverse-covariance weighted. All measures are presented in their original scales, as summarized in Appendix Table E.1. Control group means are raw (unadjusted) averages; treatment effects are estimated from regressions that include covariates and fixed effects.

J.2 Results by Outgroup Couple

One potential concern is that estimated treatment effects reflect experimenter demand: exposure to parasocial contact may have increased respondents' ability to infer the study's purpose and adjust their responses accordingly. Table J.2 addresses this concern by disaggregating effects by the two outgroup couples. As shown in Panel A, assignment to either Couple B or Couple D increased the likelihood of correctly guessing the survey's purpose by 21 and 18 percentage points, respectively. If demand effects were driving the results, similar treatment effects would be expected across both couples. Instead, responses are consistently stronger for Couple D: respondents exposed to this couple were significantly more likely to report caring about the

couple, liking the content, and liking the study overall, and the positive treatment effects on tolerance, support for social change, and behavioral intentions are driven primarily by exposure to Couple D. This pattern is inconsistent with a pure demand effects explanation.

Table J.2: Effects of Contact by Couple

	Control	Couple B (Amani and Woody)			Couple D (Gregory and Deonna)		
	mean	coef	s.e.	p-val	coef	s.e.	p-val
Panel A: Engagement with Content							
Engagement with content (%)	0.90	-0.03	(0.03)	0.328	-0.03	(0.03)	0.250
Cared about couples (%)	0.71	0.02	(0.03)	0.613	0.09**	(0.04)	0.014
Liked content (%)	0.62	0.00	(0.04)	0.938	0.07**	(0.04)	0.042
Liked study overall (%)	0.70	0.05	(0.03)	0.146	0.19***	(0.03)	0.000
Guessed purpose (%)	0.61	0.21***	(0.04)	0.000	0.18***	(0.04)	0.000
Panel B: Main Attitudinal Outcomes							
1a. Tolerance index (outgroup)	0.00	0.06	(0.06)	0.323	0.17***	(0.06)	0.005
1b. Tolerance index (diff)	0.00	0.06	(0.06)	0.334	0.15**	(0.06)	0.014
Thermometer							
Outgroup	7.07	-0.10	(0.14)	0.498	0.17	(0.13)	0.201
Ingroup	7.07	-0.08	(0.14)	0.586	-0.12	(0.14)	0.403
Diff	0.00	-0.02	(0.14)	0.883	0.29**	(0.13)	0.031
Friend							
Outgroup	2.75	0.05	(0.05)	0.256	0.16***	(0.05)	0.001
Ingroup	2.68	0.01	(0.06)	0.864	0.01	(0.06)	0.898
Diff	0.07	0.04	(0.05)	0.425	0.15***	(0.05)	0.002
Support marriage (1-5)	4.53	0.07	(0.06)	0.256	0.08	(0.06)	0.212
2. Support for social change (1-3)	2.45	0.06	(0.05)	0.212	0.10**	(0.05)	0.046
Panel C: Personal Beliefs Mechanisms							
3a. Positive attributes index (outgroup) (excl. 'not angry')	0.00	0.04	(0.07)	0.542	0.17**	(0.07)	0.017
3b. Positive attributes index (diff) (excl. 'not angry')	0.00	0.10	(0.08)	0.201	0.09	(0.07)	0.171
3c. Positive attributes index (outgroup) (all)	0.00	0.03	(0.07)	0.636	0.15**	(0.07)	0.042

Notes: *p<.10, ** p<.05, *** p<.01. Table reports the effects of the contact treatment on pre-registered outcomes and mechanisms. Measures labeled "(difference)" represent the difference between ratings of Black and white individuals. All indices are inverse-covariance weighted. All measures are presented in their original scales, as summarized in Appendix Table E.1.

J.3 Multiple Hypothesis Testing Correction

I apply the False Discovery Rate (FDR) correction using sharpened q-values to account for multiple comparisons (Anderson, 2008). Panel A shows corrections using the results for the "outgroup" indices, Panel B shows corrections using the "diff" index results. As pre-registered, I correct for three primary outcomes: tolerance, support for social change, and the overall behavioral intentions action index. After correction, all three outcomes remain close to conventional significance thresholds at the 95 percent confidence level (column 2). When the correction is extended to include all five main outcomes (including the contest outcomes), the tolerance index and support for social change remain significant at the 90 percent level; the action index remains close to significant at the 90 percent confidence level (column 3).

Columns 4–6 apply corrections for multiple testing on the mechanism measures. These corrections are relatively stringent since the mechanisms were exploratory: I did not hypothesize positive effects on each one individually but rather expected to observe effects on at least one. Even under these adjustments, the effects on the main outcomes and on the main positive attributes mechanism remain close to significance at the 90 percent confidence level. For example, in column 4 the positive attributes index remains significant at the 90 percent level ($q = .099$), though this attenuates to $q = .136$ when using the version based on all components. Finally, as pre-registered, I also apply an FDR correction across all outcome and mechanism measures combined (column 6). Under this conservative correction, no individual test is significant. This is not surprising given that correcting across all outcomes, including exploratory ones, substantially increases stringency.

Table J.3: Multiple Hypothesis Testing Correction for Average Treatment Effects

		Orig	FDR Sharpened Q-Values				
		p-value	Main	Main All	Main+Mech	Main+Mech	Main +Mech
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Corrections using "outgroup" indices							
Main Outcomes							
(1)	Tolerance index	0.011	0.035	0.059	0.084	0.084	0.110
(2)	Support social change	0.034	0.036	0.073	0.099	0.114	0.137
(3)	Action index	0.095	0.054	0.105	0.118	0.136	0.167
(4)	Entered contest	0.196		0.173			0.247
(5)	Actions Difference	0.251		0.178			0.272
Personal Beliefs Mechanisms							
(6a)	Positive attributes (excl. 'not angry')	0.045			0.099		0.137
(6b)	Positive attributes (all components)	0.093				0.136	
(7)	Similarity index	0.579			0.383	0.383	0.423
(8)	Empathy	0.645			0.383	0.383	0.423
(9)	Comfort around (diff)	0.198			0.189	0.200	0.247
Panel B: Corrections using "diff" indices							
Main Outcomes							
(1)	Tolerance index	0.024	0.054	0.093	0.136		0.181
(2)	Support social change	0.034	0.054	0.093	0.136		0.181
(3)	Action index	0.095	0.054	0.105	0.136		0.200
(4)	Entered contest	0.196		0.173			0.272
(5)	Actions Difference	0.251		0.178			0.272
Personal Beliefs Mechanisms							
(6)	Positive attributes (excl. 'not angry')	0.082			0.136		0.200
(7)	Similarity index	0.579			0.383		0.423
(8)	Empathy	0.645			0.383		0.423
(9)	Comfort around (diff)	0.198			0.200		0.272

Notes: The table reports the raw p-values (column 1) along with FDR sharpened q-values for different configurations in columns 2-6.

K Additional Analysis: Private vs. Public Attitudes

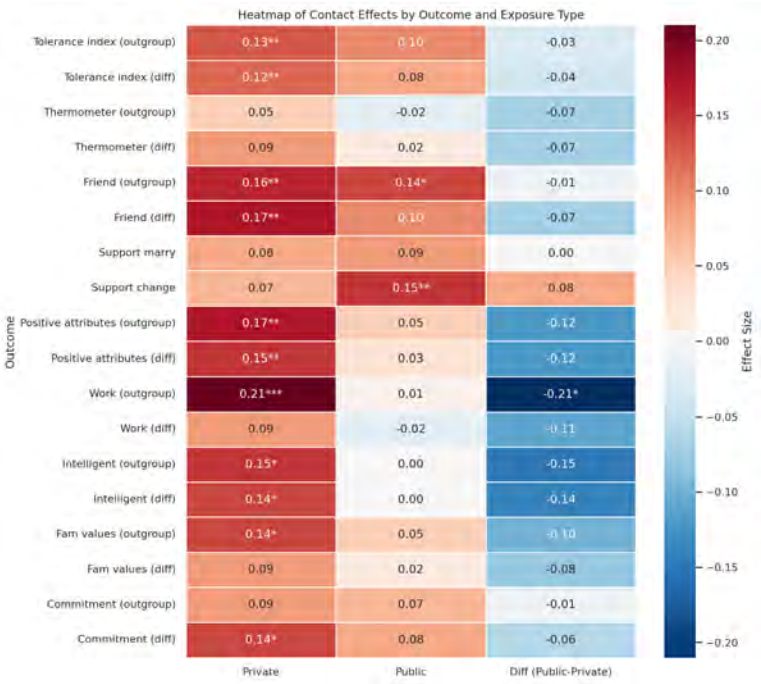
K.1 Contact × Public Interaction Results

Tables K.1 (outcomes) and K.2 (mechanisms) report the full regression results for the interaction between parasocial contact and assignment to the public condition in the enumerator experiment, corresponding to Figure 2 in the main text. Each table presents raw means for the 2×2 design (parasocial contact × private/public; Panel A), standardized regression coefficients (Panel B), and the marginal effects of contact in the public condition (Panel C). Figure K.1 summarizes the key patterns visually.

These results further clarify the patterns discussed in the main text and help rule out an alternative explanation based on reverse social desirability pressure (i.e., social benefits). If respondents in the parasocial contact control group faced pressure to overstate pro-outgroup attitudes in public, one would expect inflated public control responses and mechanically smaller treatment effects. Instead, the coefficients on β_2 are uniformly null or negative, indicating that weaker public effects reflect attenuation under social costs rather than elevated control-group responses.

The results also reinforce the decision to exclude the "not angry" item from the positive attributes index. As shown in columns 17–18 of Table K.2, although the difference score for this item aligns with the predicted pattern, the effect is driven primarily by shifts in perceptions of white Americans rather than the outgroup. Including this item would therefore risk overstating treatment effects on outgroup attitudes.

Figure K.1: Summary of 2x2 Interaction Results



Notes: Cells show estimated effects of parasocial contact in private and public conditions, as well as their difference. Values are standardized regression coefficients; red indicates positive effects and blue indicates negative effects. The expectation is positive effects in private (red), null or small positive effects in public (gray or light red), and a negative difference (blue). Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table K.1: Effects of Contact on Private vs. Public Attitudes—Main Outcomes

			Panel A		**Panel B**							**Panel C**
			Warmth indices		**Warmth Index Components**							**Support**
					Thermometer			Friends			Support	**Social**
		Hyp:	**Outgroup**	**Diff**	Outgroup	Ingroup	Diff	Outgroup	Ingroup	Diff	Marry	**Change**
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Raw Means												
(1)	Control-Private		0.05	0.05	7.12	7.10	0.02	2.76	2.33	0.09	4.59	2.47
(2)	Control-Public		-0.05	-0.05	7.02	7.04	-0.02	2.73	2.32	0.05	4.48	2.42
(3)	Contact-Private		0.05	0.04	6.99	7.01	-0.03	2.81	2.32	0.14	4.58	2.49
(4)	Contact-Public		0.04	0.00	6.94	6.98	-0.03	2.85	2.26	0.11	4.53	2.50
Panel B: Standardized Regression Coefficients												
(5)	B1: Contact (private)	*H1: +*	0.13**	0.12**	0.05	-0.05	0.09	0.16**	-0.02	0.17**	0.08	0.07
			(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)
			0.037	0.049	0.451	0.488	0.144	0.028	0.813	0.026	0.251	0.340
(6)	B2: Public		-0.05	-0.08	0.01	0.03	-0.01	0.01	0.04	-0.03	-0.11	-0.09
			(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)	(0.07)	(0.08)	(0.07)
			0.493	0.237	0.831	0.706	0.868	0.923	0.611	0.622	0.175	0.226
(7)	B3: Contact*Public	*H3: -*	-0.03	-0.04	-0.07	0.01	-0.07	-0.01	0.06	-0.07	0.00	0.08
			(0.09)	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)	(0.12)	(0.11)	(0.11)	(0.10)
			0.780	0.670	0.498	0.951	0.468	0.895	0.597	0.487	0.986	0.420
(8)	B0: Constant		0.03	1.58	-0.69	-2.92	1.94	-0.34	-1.98	1.57	0.81	0.46
			(0.27)	(0.26)	(0.27)	(0.26)	(0.25)	(0.28)	(0.33)	(0.30)	(0.30)	(0.28)
Panel C: Marginal Effects												
(9)	B1+B3: Contact (public)	*H2: 0/+*	0.10	0.08	-0.02	-0.04	0.02	0.14*	0.04	0.10	0.09	0.15**
			(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)	(0.07)	(0.08)	(0.07)
			0.128	0.213	0.817	0.543	0.716	0.050	0.611	0.185	0.268	0.039
	N		1543	1543	1543	1543	1543	1543	1543	1543	1543	1542
	Controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table reports results from separate regressions of each outcome on indicators for parasocial contact, assignment to the public (versus private) enumerator condition, and their interaction, as specified in Equation 2. All outcomes are standardized by the mean and standard deviation of the control group. Raw means are included to illustrate baseline levels across experimental conditions. Regression estimates are adjusted for controls and fixed effects.

Table K.2: Effects of Contact on Private vs. Public Attitudes—Main Mechanism

			Panel A: Pos. Attributes Indices			**Panel B: Positive Attribute Index Components**													
			(excluding 'not angry' component)			Work			Intelligent			Family Values			Want Commitment			Not angry/violent		
		Hyp:	Outgroup	Ingroup	Diff	Outgroup	Ingroup	Diff	Outgroup	Ingroup	Diff	Outgroup	Ingroup	Diff	Outgroup	Ingroup	Diff	Outgroup	Ingroup	Diff
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Panel A: Raw Means																				
(1)	Control-Private		0.04	0.07	-0.01	6.38	6.27	0.11	6.54	6.38	0.15	6.29	6.11	0.16	6.45	6.84	-0.39	6.25	6.18	0.07
(2)	Control-Public		-0.04	-0.07	0.01	6.31	6.30	-0.01	6.48	6.17	0.30	6.15	5.87	0.26	6.23	6.60	-0.39	5.92	5.66	0.25
(3)	Contact-Private		0.10	0.09	0.04	6.62	6.51	0.11	6.64	6.40	0.24	6.39	6.15	0.24	6.45	6.74	-0.29	6.02	5.70	0.32
(4)	Contact-Public		-0.01	-0.04	0.04	6.31	6.36	-0.05	6.49	6.20	0.29	6.17	5.92	0.26	6.37	6.64	-0.27	6.06	5.84	0.22
Panel B: Standardized Regression Coefficients																				
(5)	B1: Contact (private)	*H1: +*	0.17**	0.05	0.15**	0.21***	0.15*	0.09	0.15*	0.03	0.14*	0.14*	0.06	0.09	0.09	-0.04	0.14*	-0.02	-0.22**	0.20***
			(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
			0.028	0.519	0.039	0.006	0.077	0.228	0.075	0.686	0.064	0.083	0.514	0.276	0.301	0.640	0.078	0.806	0.011	0.010
(6)	B2: Public		-0.05	-0.10	0.04	0.00	0.04	-0.04	0.01	-0.10	0.12	-0.06	-0.10	0.03	-0.07	-0.10	0.01	-0.11	-0.22**	0.11
			(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.08)
			0.525	0.246	0.620	0.965	0.618	0.595	0.897	0.248	0.152	0.476	0.260	0.703	0.377	0.269	0.927	0.172	0.014	0.171
(7)	B3: Contact*Public	*H3: -*	-0.12	-0.02	-0.12	-0.21*	-0.11	-0.11	-0.15	-0.03	-0.14	-0.10	-0.02	-0.08	-0.01	0.05	-0.06	0.09	0.30**	-0.20*
			(0.11)	(0.12)	(0.11)	(0.11)	(0.12)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.12)
			0.258	0.844	0.280	0.060	0.338	0.322	0.196	0.784	0.231	0.408	0.861	0.541	0.907	0.672	0.593	0.424	0.019	0.080
(8)	B0: Constant		-0.28	-1.62	1.34	-0.10	-1.92	1.67	-0.89	-2.10	1.33	0.30	-0.39	0.65	-0.45	-1.24	0.80	0.87	-0.58	1.52
			(0.30)	(0.33)	(0.29)	(0.31)	(0.31)	(0.30)	(0.33)	(0.32)	(0.30)	(0.30)	(0.35)	(0.34)	(0.32)	(0.34)	(0.32)	(0.32)	(0.34)	(0.32)
Panel C: Marginal Effects																				
(9)	B1+B3: Contact (public)	*H2: 0/+*	0.05	0.03	0.03	0.01	0.04	-0.02	0.00	0.00	0.00	0.05	0.03	0.02	0.07	0.01	0.08	0.07	0.08	0.00
			(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)
			0.541	0.716	0.670	0.913	0.662	0.802	0.972	0.990	0.993	0.570	0.701	0.830	0.398	0.882	0.382	0.379	0.401	0.997
	N		1540	1540	1539	1540	1540	1539	1540	1540	1539	1540	1540	1539	1540	1540	1539	1540	1540	1539
	Controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table reports results from separate regressions of each outcome on indicators for parasocial contact, assignment to the public (versus private) enumerator condition, and their interaction, as specified in Equation 2. All outcomes are standardized by the mean and standard deviation of the control group. Raw means are included to illustrate baseline levels across experimental conditions. Regression estimates are adjusted for controls and fixed effects.

K.2 Parasocial × Public × Cost: Triple Interaction Results

To explore whether contact effects vary with perceived social costs, I estimate a triple interaction between parasocial contact, visibility (public vs. private), and an indicator for high social cost, defined by perceived ingroup norms (see Section 4). Table K.3 presents both regression coefficients and calculated marginal effects (see Figure K.2 for a visual summary). This table corresponds to Figure 3 in the main text.

To explore whether the effects of contact on private versus public attitudes depend on social costs, I estimate a triple interaction of the following form:

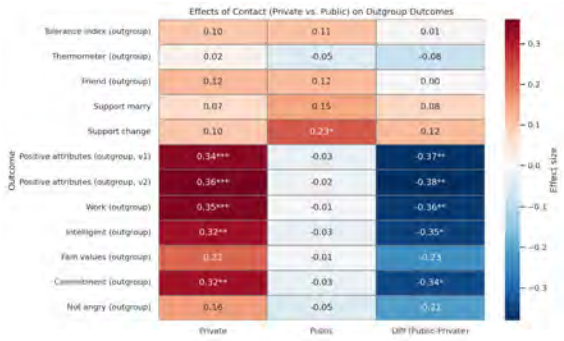
$$\begin{aligned} Y_i = \alpha &+ \beta_1 \,\text{Contact}_i + \beta_2 \,\text{Pub}_i + \beta_3 \,(\text{Contact}_i \times \text{Pub}_i) \\ &+ \beta_4 \,\text{HighCost}_i + \beta_5 \,(\text{Contact}_i \times \text{HighCost}_i) \\ &+ \beta_6 \,(\text{Pub}_i \times \text{HighCost}_i) + \beta_7 \,(\text{Contact}_i \times \text{Pub}_i \times \text{HighCost}_i) \\ &+ X_i'\psi + \mu_j + \epsilon_i \end{aligned}$$

where all terms are defined as in Equation 2, with the addition of an indicator for "high social cost," as defined in Section 4.

Table K.3 reports both the regression coefficients and the corresponding marginal effects plotted in Figure 3. Marginal effects were estimated using Stata's `margins` command, evaluated at $Vic = 0$. I exclude the "difference" measures here because they are difficult to interpret in a triple interaction: they embed treatment effects on ingroup as well as outgroup attitudes, with the former conditioned on perceived ingroup norms toward the outgroup. Since these ingroup effects are conceptually distinct from the main theoretical interest, the "difference" measures are not informative in this context. Figure K.2 summarizes the main results for those with high social costs.

The results provide additional insight into the patterns discussed in the main text and further validate the interpretation of social costs. Estimates for β_4 show that respondents facing high social costs express substantially more negative outgroup attitudes in the ingroup contact and private conditions. These differences are large, statistically significant across nearly all outcomes, and persist after controlling for baseline prejudice, supporting the social cost measure as theoretically and behaviorally meaningful. In addition, the negative or null effects of public exposure in the control condition (row 12) help rule out reverse social desirability pressure: if public visibility inflated baseline tolerance, outgroup ratings would be higher—not lower—in the public control condition. Instead, the pattern reinforces the interpretation that social costs suppress public expression of tolerant attitudes.

Figure K.2: Summary of Triple Interaction Results



Notes: Cells show estimated effects of parasocial contact in private and public conditions, as well as their difference. Values are regression coefficients; red indicates positive effects and blue indicates negative effects. The expectation is positive effects in private (red), null or small positive effects in public (gray or light red), and a negative difference (blue). Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table K.3: Triple Interaction Results

		Panel A: Outgroup Warmth				Panel B:	Panel C: Outgroup Positive Attributes						
		Warmth Index	Therm	Friends	Marry	**Support Change**	**Pos. Attributes Index (v1)**	**Index (v2)**	Work	Intelligent	Family	Commit	Not angry
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Standardized Regression Coefficients													
(1)	B1: Contact (when private, low cost)	0.11	0.03	0.15*	0.07	0.03	-0.06	0.02	0.10	0.01	0.06	-0.08	-0.17*
		(0.07)	(0.08)	(0.09)	(0.08)	(0.08)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)
		0.120	0.725	0.088	0.407	0.673	0.578	0.857	0.336	0.912	0.578	0.431	0.100
(2)	B2: Pub (when ingroup contact, low cost)	-0.04	0.02	0.03	-0.13	-0.10	-0.20**	-0.17*	-0.04	-0.05	-0.15	-0.23**	-0.20**
		(0.08)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
		0.594	0.779	0.744	0.169	0.244	0.039	0.080	0.708	0.615	0.150	0.025	0.048
(3)	B3: Contact*Pub (when low cost)	0.01	0.00	0.03	0.00	0.06	0.21	0.10	-0.05	0.03	0.04	0.23	0.35**
		(0.11)	(0.13)	(0.13)	(0.12)	(0.12)	(0.14)	(0.14)	(0.14)	(0.16)	(0.15)	(0.15)	(0.15)
		0.907	0.992	0.805	0.982	0.607	0.142	0.484	0.723	0.832	0.805	0.117	0.019
(4)	B4: Cost (when ingroup contact, private)	-0.23**	-0.17*	-0.15	-0.21	-0.15	-0.47***	-0.41***	-0.29**	-0.31**	-0.31**	-0.37***	-0.46***
		(0.11)	(0.10)	(0.11)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)	(0.12)
		0.030	0.073	0.176	0.115	0.190	0.000	0.001	0.016	0.010	0.014	0.004	0.000
(5)	B5: Contact*Cost (when private)	-0.01	-0.01	-0.03	0.00	0.07	0.39**	0.34**	0.25	0.31*	0.16	0.40**	0.33*
		(0.14)	(0.14)	(0.15)	(0.16)	(0.15)	(0.16)	(0.17)	(0.16)	(0.17)	(0.17)	(0.18)	(0.17)
		0.921	0.969	0.842	0.984	0.651	0.017	0.042	0.118	0.075	0.358	0.023	0.052
(6)	B6: Pub*Cost (when ingroup contact)	-0.08	-0.08	-0.11	0.00	0.03	0.27	0.25	0.06	0.11	0.18	0.38**	0.19
		(0.15)	(0.14)	(0.16)	(0.18)	(0.16)	(0.16)	(0.17)	(0.17)	(0.17)	(0.17)	(0.18)	(0.17)
		0.604	0.567	0.494	0.989	0.836	0.103	0.137	0.741	0.517	0.306	0.037	0.259
(7)	B7: Contact*Pub*Cost	0.00	-0.08	-0.03	0.08	0.06	-0.58**	-0.48**	-0.31	-0.38	-0.26	-0.58**	-0.56**
		(0.20)	(0.20)	(0.22)	(0.23)	(0.21)	(0.23)	(0.23)	(0.23)	(0.24)	(0.24)	(0.25)	(0.24)
		0.997	0.700	0.874	0.722	0.779	0.012	0.042	0.178	0.118	0.286	0.021	0.021
(8)	B0: Constant	0.36	-0.47	-0.05	1.05	0.44	0.22	0.01	0.14	-0.54	0.50	-0.27	0.93
		(0.24)	(0.23)	(0.25)	(0.26)	(0.24)	(0.26)	(0.26)	(0.26)	(0.27)	(0.26)	(0.27)	(0.28)
Marginal Effects													
(9)	B1+B5: Effect of outgroup contact (private, high cost)	0.10	0.02	0.12	0.07	0.10	0.34***	0.36***	0.35***	0.32**	0.22	0.32**	0.16
		(0.12)	(0.11)	(0.12)	(0.14)	(0.13)	(0.13)	(0.13)	(0.12)	(0.13)	(0.14)	(0.14)	(0.13)
		0.401	0.819	0.311	0.633	0.425	0.009	0.007	0.005	0.017	0.116	0.027	0.238
(10)	B1+B3+B5+B7: Effect of outgroup contact (public, high cost)	0.11	-0.05	0.12	0.15	0.23*	-0.03	-0.02	-0.01	-0.03	-0.01	-0.03	-0.05
		(0.12)	(0.11)	(0.12)	(0.13)	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)	(0.14)	(0.13)
		0.353	0.622	0.321	0.267	0.053	0.781	0.867	0.928	0.829	0.960	0.856	0.718
(11)	B3+B7: Effect of outgroup contact (diff: pub-private)	0.01	-0.08	0.00	0.08	0.12	-0.37**	-0.38**	-0.36**	-0.35*	-0.23	-0.34*	-0.21
		(0.17)	(0.15)	(0.17)	(0.20)	(0.18)	(0.18)	(0.19)	(0.18)	(0.19)	(0.20)	(0.20)	(0.19)
		0.934	0.609	0.991	0.685	0.482	0.040	0.042	0.045	0.061	0.253	0.087	0.274
(12)	B2+B6: Effect of public (ingroup contact, high cost)	-0.12	-0.05	-0.08	-0.13	-0.07	0.07	0.09	0.02	0.06	0.03	0.15	-0.01
		(0.13)	(0.11)	(0.13)	(0.15)	(0.13)	(0.13)	(0.14)	(0.14)	(0.13)	(0.14)	(0.15)	(0.14)
		0.348	0.611	0.541	0.391	0.578	0.587	0.538	0.887	0.663	0.819	0.319	0.965
	N	1543	1543	1543	1543	1542	1540	1540	1540	1540	1540	1540	1540
	Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table reports results from separate regressions estimating triple interactions between parasocial contact, enumerator condition (public vs. private), and social cost level (high vs. low). Reported estimates are the marginal effects of contact within each condition (public/private × high/low cost), derived from the triple interaction model. All variables are reported in their original scales. Positive attributes index v1 includes the 'not angry' attribute, v2—the main one reported in the text—excluded the 'not angry' attribute.

L Additional Analysis: Public vs. Private Behavior

L.1 Behavioral Intentions Factor Structure

To assess whether public and private political behaviors reflect distinct dimensions of participation, I conducted an exploratory factor analysis of eight behavioral intention outcomes. Table L.1 presents results using a two-factor structure with promax rotation, estimated separately for the control group (Panel A) and full sample (Panel B). The factor structure is highly stable across samples: voting items load on a "private" dimension, protest, confrontation, and sharing on a "public" dimension, and two behaviors—reading an article and signing a private petition—load on both, indicating intermediate visibility. The public petition is less clear-cut, loading ambiguously in the control group but more clearly on the public factor in the full sample.

Table L.1: Factor Analysis Results

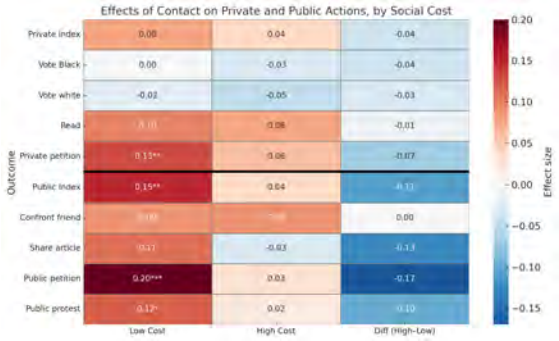
	Panel A: Control Only		Panel B: Para Sample	
	Factor 1	**Factor 2**	**Factor 1**	**Factor 2**
Variance				
Eigenvalues	4.416	4.353	4.500	4.250
Factor Loadings				
Vote black candidate	0.072	0.860	0.128	0.822
Vote white candidate	-0.013	0.890	0.017	0.878
Read article	0.552	0.306	0.544	0.301
Sign private petition	0.450	0.519	0.501	0.447
Share article	0.757	0.004	0.745	0.019
Confront a front	0.566	0.209	0.559	0.187
Sign a public petition	0.489	0.473	0.619	0.342
Public protest	0.785	0.003	0.757	0.010
Key				
Primarily loads on Factor 1				
Primarily loads on Factor 2				
Loads on both				

Notes: Promax-rotated factor loadings for behavioral intention items. Loadings are from a pca analysis with oblique (promax) rotation. Panel A is baseline factor structure using the control (ingroup parasocial contact) sample only. Panel B shows the factor structure using the full parasocial contact sample. Eigenvalues are post-rotation. Light gray shading indicates items that primarily load on Factor 2 (clearly private); dark gray shading indicates items that primarily load on Factor 1 (clearly public); medium gray shading indicates items that load moderately on both.

L.2 Additional Results

Table L.2 presents full results for regressions of intentions to take public and private actions on the contact treatment, high social costs, and their interaction, corresponding to Figure 4. Each table presents raw means for the 2×2 design (parasocial contact × high/low cost; Panel A), standardized regression coefficients (Panel B), and the marginal effects of contact in the high cost condition (Panel C). Figure L.1 summarizes the key patterns visually.

Figure L.1: Summary of Contact*Cost Interaction Results



Notes: Cells show estimated effects of parasocial contact in private and public conditions, as well as their difference. Values are regression coefficients; red indicates positive effects and blue indicates negative effects. The expectation is positive effects in private (red), null or small positive effects in public (gray or light red), and a negative difference (blue). Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table L.2: Detailed Contact*Cost Interaction Results

	Private Actions Index	Clearly Private Vote Black	Vote White	Intermediate Read Article	Private Petition	**Public Actions Index**	Confront Friend	Clearly Public Share Article	Public Petition	Public Protest
Panel A: Raw Means										
Control-Low Cost	0.05	3.82	3.80	3.99	3.76	0.00	3.74	2.92	3.65	2.79
Control-High Cost	-0.05	3.71	3.67	3.87	3.67	0.00	3.76	2.99	3.52	2.83
Contact-Low Cost	0.00	3.65	3.63	3.96	3.81	0.05	3.73	3.04	3.72	2.91
Contact-High Cost	0.04	3.74	3.65	4.03	3.83	0.09	3.86	3.02	3.72	2.91
Panel B: Standardized Regression Coefficients										
B1: Contact	0.08	0.00	-0.02	0.10	0.13**	0.15**	0.09	0.11	0.20***	0.12*
	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)
	0.175	0.949	0.777	0.139	0.038	0.012	0.130	0.125	0.002	0.079
B2: High Social Costs	-0.11	-0.11	-0.08	-0.15*	-0.05	-0.14*	-0.32***	0.00	-0.04	0.01
	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
	0.160	0.166	0.371	0.079	0.541	0.065	0.000	0.985	0.669	0.927
B3: Contact*Costs	-0.04	-0.04	-0.03	-0.01	-0.07	-0.11	0.00	-0.13	-0.17	-0.10
	(0.11)	(0.11)	(0.12)	(0.11)	(0.11)	(0.10)	(0.11)	(0.11)	(0.11)	(0.11)
	0.676	0.751	0.767	0.916	0.497	0.289	0.969	0.234	0.113	0.338
B0: Constant	0.18	0.29	-0.07	0.16	0.26	0.25	0.19	0.03	0.32	0.32
	(0.25)	(0.24)	(0.26)	(0.25)	(0.24)	(0.21)	(0.22)	(0.24)	(0.23)	(0.23)
Panel C: Marginal Effects										
B1+B3: Contact (high costs)	0.04	-0.03	-0.05	0.08	0.06	0.04	0.09	-0.03	0.03	0.02
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)
	0.649	0.726	0.554	0.373	0.497	0.600	0.305	0.771	0.763	0.831
N	1542	1542	1542	1542	1542	1542	1542	1542	1542	1542
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

L.3 Gleam Contest Results

Table L.3: Effects of Contact Treatment on Actual Private vs. Public Actions

	Control mean	coef	se	p-val	N (Full)	N (Para)
i. Private diff (race-nonrace private actions)	-0.03	0.04	(0.03)	0.189	770	520
Total race-related private actions (0-4)	0.97	-0.08	(0.14)	0.551	770	520
Netflix BLM (0/1)	0.27	-0.03	(0.04)	0.434	770	520
Documentary (0/1)	0.25	-0.01	(0.04)	0.734	770	520
Trailer (0/1)	0.28	-0.04	(0.04)	0.272	770	520
Petition (0/1)	0.17	0.00	(0.03)	0.897	770	520
Total nonrace related private actions (0-4)	1.00	-0.12	(0.14)	0.367	770	520
Netflix (0/1)	0.27	-0.04	(0.04)	0.304	770	520
Documentary (0/1)	0.27	-0.04	(0.04)	0.255	770	520
Trailer (0/1)	0.27	-0.03	(0.04)	0.386	770	520
Petition (0/1)	0.18	-0.01	(0.04)	0.865	770	520
ii. Public diff (race-nonrace public actions)	0.00	0.01	(0.01)	0.697	780	523
Total race-related public actions (0-4)	0.11	0.00	(0.05)	0.993	780	523
Netflix BLM (0/1)	0.03	0.00	(0.01)	0.827	780	523
Documentary (0/1)	0.03	0.00	(0.02)	0.788	780	523
Trailer (0/1)	0.03	0.01	(0.02)	0.769	780	523
Petition (0/1)	0.03	0.00	(0.01)	0.853	780	523
Total nonrace-related public actions (0-4)	0.12	0.00	(0.05)	0.931	780	523
Netflix (0/1)	0.03	0.00	(0.02)	0.804	780	523
Documentary (0/1)	0.03	0.00	(0.01)	0.844	780	523
Trailer (0/1)	0.03	0.00	(0.02)	0.890	780	523
Petition (0/1)	0.03	0.00	(0.02)	0.774	780	523

Notes: * p<.10, ** p<.05, *** p<.01. Table shows the effects of the contact treatment (estimated using Equation 1) where each row is a different outcome. Indices are z-scores produced using inverse covariance weighting.

M Regression Results with Covariates

This Appendix reports regression results along with the full set of covariates. Table M.1 presents results for the main average treatment effects reported in Table 1 (attitudes) and Table 2 (behavioral intentions) in the main text. Table M.2 reports main results from the interaction regressions reported in Figures 3 and 4.

N Research Ethics

This study involved human subjects and received approval from the Institutional Review Board (IRB) at [REDACTED]. Participants provided informed consent and were recruited through CloudResearch following their standard protocols. Participation was voluntary, and respondents could exit the survey at any point.

The study posed minimal risk. Treatments consisted of exposure to publicly available media content and survey questions concerning social and political attitudes, including attitudes toward racial groups. No deception was used beyond withholding specific hypotheses, and no personally identifying information was collected or retained. Data were analyzed in aggregate, and respondents were compensated according to standard platform practices.

Table M.1: Average Treatment Effects with Full Covariates

	Warmth index (outgroup)	Warmth index (diff)	Support social change	Positive attributes (outgroup)	Positive attributes (diff)	Private actions index	Public actions index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Parasocial Contact	**0.12****	**0.10****	**0.11****	**0.11****	**0.09***	**0.07**	**0.11****
	(0.05)	**(0.05)**	**(0.05)**	**(0.05)**	**(0.05)**	**(0.05)**	**(0.05)**
Vicarious Contact	0.03	-0.01	0.01	0.05	0.06	-0.02	0.01
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Age	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (%)	0.11***	0.13***	0.08*	0.10**	0.11**	0.18***	0.27***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
Education completed	-0.04**	-0.04**	0.03	-0.03	0.03	0.02	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Income	0.00	0.01*	-0.01**	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Urban (%)	0.01	0.02	0.00	0.05	-0.01	0.04	0.04
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Republican (%)	-0.15***	-0.17***	-0.85***	-0.27***	-0.21***	-0.70***	-0.54***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Network share: Married	0.02	0.01	0.02	0.04*	0.02	-0.01	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Like reality TV	0.05**	0.01	0.02	0.06**	0.01	0.07***	0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Region: Southeast (%)	0.19*	0.08	0.11	-0.14	-0.18*	-0.06	-0.16
	(0.10)	(0.09)	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)
Region: New England (%)	0.00	-0.05	0.23*	-0.17	-0.22*	-0.01	-0.04
	(0.10)	(0.11)	(0.12)	(0.15)	(0.13)	(0.12)	(0.11)
Region: Northeast (%)	0.08	0.03	0.09	-0.14	-0.19**	-0.03	0.00
	(0.07)	(0.07)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
Region: Midwest East (%)	0.13*	0.09	0.10	-0.04	-0.09	-0.03	-0.04
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Region: Midwest West (%)	0.04	0.00	0.25**	-0.31***	-0.17	-0.07	0.00
	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.10)	(0.09)
Region: South Atlantic (%)	0.07	-0.01	0.07	-0.18**	-0.22***	-0.09	-0.12
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)
Region: South Central (%)	0.09	0.04	0.12	-0.25**	-0.25**	-0.12	-0.12
	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)	(0.10)	(0.09)
Region: Mountain (%)	0.02	-0.05	0.15	-0.25**	-0.30***	-0.08	-0.06
	(0.10)	(0.10)	(0.10)	(0.11)	(0.12)	(0.11)	(0.10)
Index social media	-0.01	-0.07***	-0.01	0.04*	-0.01	0.02	0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Like white neighbors	-0.19***	-0.46***	-0.09***	-0.05	-0.32***	-0.07*	-0.05
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
Index like outgroups (excluding Black)	0.23***	0.27***	0.00	0.06	0.11**	0.04	0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Index like outgroup (Black)	0.35***	0.35***	0.25***	0.29***	0.31***	0.27***	0.29***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Index norms neighbors	-0.20***	-0.14***	-0.04*	-0.20***	-0.10***	-0.06**	-0.08***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
Index norms insensitive comments	-0.08***	-0.06***	-0.05**	-0.08***	-0.03	-0.06***	-0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Network share: Black	0.12***	0.07**	0.02	0.09**	0.04	0.05	0.18***
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Network share: White	0.06**	0.00	-0.01	0.05	-0.05	0.02	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Uncomfortable with ingroup disapproval	-0.01	-0.01	0.02	0.01	-0.01	0.01	-0.05*
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Index like outgroup shows	0.05**	0.03	0.09***	-0.01	0.02	0.10***	0.13***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Treatment assignment: police vignette (couple)	0.03	0.02	0.05	0.05	0.08	0.05	0.02
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Treatment assignment: police vignette (officer)	-0.02	0.01	0.02	0.04	0.04	-0.04	0.00
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Treatment assignment: mayor vignette	0.00	-0.04	-0.06	-0.06	-0.08*	-0.06	-0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	-0.01	1.51	0.41	-0.33	1.35	0.04	0.16
	(0.27)	(0.26)	(0.27)	(0.29)	(0.29)	(0.28)	(0.25)
N	1543	1543	1542	1540	1539	1542	1542

Notes: * p<.10, ** p<.05, *** p<.01. Table shows the effects of the contact treatment (estimated using Equation 1) on main outcomes and mechanisms with results for all covariates.

Table M.2: Contact × Public, by Social Cost with Full Covariates

	Warmth index (outgroup) (1)	Support social change (2)	Positive attributes (outgroup) (3)	Private actions index (4)	Public actions index (5)
Parasocial Contact	0.11	0.03	0.02	0.08	0.15**
	(0.07)	(0.08)	(0.10)	(0.06)	(0.06)
Public	-0.04	-0.10	-0.17*	-0.03	-0.01
	(0.08)	(0.09)	(0.10)	(0.04)	(0.04)
Parasocial Contact*Public	0.01	0.06	0.10		
	(0.11)	(0.12)	(0.14)		
HighCost	-0.23**	-0.15	-0.41***	-0.11	-0.14*
	(0.11)	(0.12)	(0.12)	(0.08)	(0.08)
Parasocial Contact*HighCost	-0.01	0.07	0.34**	-0.04	-0.11
	(0.14)	(0.15)	(0.17)	(0.11)	(0.10)
Public*HighCost	-0.08	0.03	0.25		
	(0.15)	(0.16)	(0.17)		
Parasocial Contact*Public*HighCost	0.00	0.06	-0.48**		
	(0.20)	(0.21)	(0.23)		
Vicarious Contact	0.06	-0.09	-0.01	-0.04	0.03
	(0.08)	(0.09)	(0.10)	(0.06)	(0.06)
Vicarious Contact*Public	-0.06	0.17	0.15		
	(0.11)	(0.13)	(0.14)		
Vicarious Contact*HighCost	-0.02	0.05	0.10	0.04	-0.08
	(0.15)	(0.16)	(0.17)	(0.11)	(0.10)
Vicarious Contact*Public*HighCost	-0.03	-0.01	-0.29		
	(0.22)	(0.22)	(0.24)		
Age	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female (%)	0.15***	0.10**	0.13***	0.20***	0.30***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Education completed	-0.05**	0.03	-0.04	0.02	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Income	0.00	-0.01**	0.00	0.00	-0.01
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Urban (%)	-0.01	0.00	0.03	0.04	0.03
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Republican (%)	-0.15***	-0.85***	-0.27***	-0.70***	-0.54***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Network share: Married	0.04*	0.02	0.06**	0.00	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Like reality TV	0.05**	0.02	0.06**	0.07***	0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Region: Southeast (%)	0.20**	0.11	-0.14	-0.05	-0.15
	(0.10)	(0.11)	(0.11)	(0.11)	(0.10)
Region: New England (%)	0.00	0.23*	-0.19	0.00	-0.03
	(0.11)	(0.12)	(0.15)	(0.12)	(0.11)
Region: Northeast (%)	0.09	0.09	-0.15*	-0.03	0.00
	(0.07)	(0.09)	(0.09)	(0.08)	(0.08)
Region: Midwest East (%)	0.14**	0.10	-0.05	-0.02	-0.03
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
Region: Midwest West (%)	0.09	0.24**	-0.28**	-0.05	0.03
	(0.10)	(0.10)	(0.11)	(0.10)	(0.09)
Region: South Atlantic (%)	0.04	0.06	-0.22***	-0.09	-0.14*
	(0.07)	(0.08)	(0.08)	(0.08)	(0.07)
Region: South Central (%)	0.05	0.11	-0.28***	-0.14	-0.14
	(0.09)	(0.10)	(0.10)	(0.10)	(0.09)
Region: Mountain (%)	0.00	0.14	-0.27**	-0.09	-0.07
	(0.10)	(0.10)	(0.11)	(0.11)	(0.10)
Index social media	-0.01	-0.02	0.04	0.02	0.03
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Like white neighbors	-0.17***	-0.10***	-0.03	-0.07*	-0.05
	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)
Index like outgroups (excluding Black)	0.23***	0.01	0.06	0.04	0.05
	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)
Index like outgroup (Black)	0.41***	0.26***	0.35***	0.29***	0.31***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Network: Share Black	0.06**	0.03	0.05	0.02	0.15***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Uncomfortable with ingroup disapproval	0.01	0.02	0.02	0.02	-0.04
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Index like outgroup shows	0.05**	0.09***	-0.01	0.10***	0.13***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Treatment assignment: police vignette (couple)	0.04	0.05	0.06	0.05	0.03
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Treatment assignment: police vignette (officer)	-0.02	0.02	0.04	-0.04	0.00
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Treatment assignment: mayor vignette	0.00	-0.05	-0.05	-0.06	-0.05
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Constant	0.36	0.44	0.01	0.18	0.25
	(0.24)	(0.24)	(0.26)	(0.25)	(0.21)
N	1543	1542	1540	1542	1542

Notes: * p<.10, ** p<.05, *** p<.01. Table shows the effects of the contact treatment (estimated using Equations 1 and 2) on main outcomes and mechanisms with results for all covariates. Regressions on attitudinal outcomes in columns 1-3 are from the triple interaction of Contact*Public*HighCost. Regressions on behavioral intentions outcomes in columns 4 and 5 are from an interaction regression of Contact*HighCost and include assignment to Public as a control.

O Deviations from Pre-Analysis Plan

This study was pre-registered on September 9, 2023. This appendix outlines both deviations from the PAP and analytic decisions that, while consistent with the PAP, warrant further explanation.

Exclusion of the vicarious contact experiment: As noted in the main text, the design included a vicarious contact treatment arm with pre-registered public–private hypotheses paralleling those for parasocial contact (see PAP, p. 10). This arm was intended to complement parasocial contact by primarily targeting perceived social norms. Results from the vicarious contact arm are not analyzed in this paper for two reasons: the treatment did not generate meaningful effects, and parasocial and vicarious contact were pre-registered as distinct interventions with separate analytic plans (PAP, pp. 41–42).

Heterogeneous effects by social cost: The PAP pre-registered heterogeneity analyses by social cost for the vicarious contact treatment but not for parasocial contact (p. 8). This reflected the initial focus on whether vicarious contact—designed to shift perceived ingroup norms—would be most effective among individuals facing higher social costs. While parallel heterogeneity analyses for parasocial contact were not explicitly pre-registered, the underlying logic is articulated throughout the PAP. For example, p. 2 states: "Parasocial contact, if it primarily operates on personal beliefs, might alter how a person thinks or feels about an outgroup in private but have no impact on what they say or do publicly if expressing tolerance or support for social change remains socially costly." See also p. 9: "Changes in personal beliefs will not necessarily be expressed publicly if doing so would be socially costly."

Conceptualizing high social costs: The pre-analysis plan defined high social costs as meeting three conditions: (1) having a homogeneous ingroup network, (2) perceiving that network to condone prejudice, and (3) being personally sensitive to ingroup sanctioning (see pp. 8, 22, 37). The analysis presented here excludes the third component both because the "discomfort with ingroup disapproval" measure was imbalanced (raising concerns about bias) and including it produced unstable estimates in the triple-interaction models. Accordingly, high social costs are operationalized as the intersection of homogeneous ingroup networks and perceived prejudiced ingroup norms. Robustness checks using alternative codings yield substantively similar results.

Focusing on "outgroup" rather than "difference" measures: The PAP specified outcomes based on difference measures (outgroup relative to ingroup ratings). During analysis, however, it became clear that ingroup parasocial contact in the control condition could shift ingroup evaluations, and that even small changes in ingroup ratings could materially affect difference scores (Appendix J.1). For this reason, outgroup-only measures provide the cleanest test of hypotheses about parasocial contact's effects on tolerance and support for social change toward the outgroup. Difference measures remain informative and are reported throughout the paper (e.g., Table 1, Figure 2), as they capture relative evaluations and guard against generalized positivity (e.g. watching happy couples increases warmth towards everyone).

Combining similarity items into an index: The pre-analysis plan specified perceived general and values similarity as separate mechanisms (PAP, p. 44). In practice, these items capture closely related perceptions of outgroup alignment and are empirically correlated, motivating their combination into a single index. Aggregation improves reliability and limits multiple-testing concerns. Disaggregated results are substantively similar and reported in Appendix J.1.

Renumbering/rewording of hypothesis phrasing: The hypotheses labeled H1–H3 in this paper correspond to $H3_{a,b}$, $H4_{a,b}$, and $H5_{a,b}$ in the pre-analysis plan. In the PAP, the interaction hypothesis was framed as a positive interaction effect, reflecting a larger effect of parasocial contact on private than public outcomes. In the paper, this is rephrased as a negative interaction, reflecting attenuation in the public condition. This change is purely semantic and does not affect the underlying estimand, estimation, or interpretation. Results for the PAP's main effect hypotheses (H1 and M1–M5) are reported in Table 1.