Can praise from peers promote empathy and political inclusion towards racial or ethnic outgroups?

Adeline Lo[1*], Jonathan Renshon[1], Lotem Bassan-Nygate[2]
[1] Department of Political Science, University of Wisconsin-Madison, Madison, WI, USA
[2] Kennedy School of Government, Harvard University, Cambridge, MA, USA
[*] Corresponding Author: Adeline Lo (aylo@wisc.edu).

Abstract

Outgroup bias is well-documented and pernicious, manifesting in negative attitudes and behavior towards outgroups. Addressing it is a first-order priority for Diversity, Equity and Inclusion programs, as well as society more generally. Empathy—taking the perspective and understanding the experiences of others—holds promise for attenuating outgroup bias, but existing methods are expensive. Through seven pilots, we develop a low-cost, easily scalable "peer praise" intervention that encourages empathy. In this report, we test whether our intervention promotes empathy and inclusive behavior/attitudes among white U.S. respondents towards Black and Latino/a Americans, a context where outgroup bias is particularly durable. We measure costly choices to engage in empathy, test whether peer praise promotes political and attitudinal inclusion, and if praise specifically from co-partisans can also promote inclusion. We find that peer praise for empathy does not motivate whites to engage in empathy with racial outgroups, nor does it change their attitudes or self-reported empathy towards outgroups. However, peer praise for empathy does encourage politically inclusive behavior towards racial outgroups in the form of writing letters on behalf of racial equality to the government. Other registered analyses show that peer praise for empathy can change attitudes both in the short term (Wave 1) and over time (in our longitudinal Wave 2) but only for certain subgroups. Overall, this study provides an unusually comprehensive examination of a treatment to promote outgroup empathy. That treatment is demonstrated to be effective for behavioral outcomes related to political inclusion across all respondents and can even change attitudes, though only for some demographics. Broadly, our study suggests the importance of targeting empathy-promoting interventions towards receptive groups as well as the difficulty of promoting outgroup empathy, particularly when group identity is highlighted. ~~Finally, we test if these effects endure over time, vary by partisanship and function through potential mechanisms such as happiness and norms.~~

Main Text

Outgroup bias—"among the most well documented and widely observed phenomenon in the social sciences"[1]—manifests in negative perceptions of, and behavior towards outgroups, who are viewed with suspicion and discriminated against in allocations of goods[2]. When outgroups are minorities, outgroup bias carries pernicious consequences across economic (job discrimination), health (e.g., odds of long-standing illnesses and maternal mortality), and judicial/policing (bail decisions, traffic stops) domains, and overall lower prosocial behaviors towards outgroup members[3;4;5;6;7;8;9]. In the U.S. context, racial and ethnic outgroup biases held by the dominant group have deep historical and institutional roots[10], leading to polarized attitudes as well as discriminatory behavior on many fronts. Bias against Black and Latino/a citizens is reflected in debates about virtually every important public policy issue—from immigration to policing and education[11]—and is similarly pervasive in online spheres[12]. The growth of Diversity, Equity and Inclusion programs designed to combat outgroup bias in both public, private (e.g., higher education[13]) and government spheres (see Biden Administration's series of Executive Orders around DEI[14;15;16]) stand as further testament to the issue's importance.

Empathy—the act of taking the perspective and understanding the experiences of others[17;18]—is one tool that holds promise for attenuating the worst effects of outgroup bias, even towards heavily stigamatized "others"[19]. Previous work has emphasized either affective aspects of empathy—e.g., when the focus is on the decision to engage in empathy[20]—or cognition, such as in the act of thinking through another's perspective (sometimes referred to as "perspective taking" or "mentalizing")[21;22].

In keeping with the growing tendency to see empathy as multi-faceted[23]—and neither purely cognitive nor purely affective—our conception of empathy focuses on imagining the experience of others (sometimes referred to as the "imagine him" or "imagine-other" perspective to to distinguish it from imagining one's self in a different situation[24;25]). This type of empathy—consistent with a host of other recent work[26;27;28]—involves more than simply inferring another's mental state, but instead is "a process of feeling into, in which Person A opens him or herself in a deeply responsive way to Person B's feelings and experiencing but without losing awareness that B is a distinct other self"[29]. Taking the perspective of others in this way is a powerful tool for shaping policy preferences and improving attitudes towards others more broadly[17;18;30]. Critically, the effects of empathy do not stop at attitudinal warmth, carrying over to affect partisan polarization[31;32;30] and behaviors towards outgroups as well[19;33].

Given the scale of the problem and the ameliorative properties of empathy, it's no surprise that much attention has focused on the effects of empathy on outgroup bias[34;35;36]. Research in this domain has included empathy-based exercises embedded within intergroup contact scenarios, including work by Broockman and Kalla[37;38], who successfully utilize resource-heavy face-to-face conversations and perspective-taking to reduce exclusionary attitudes towards outgroup members (see[39]and[40] for recent reviews). Other works utilize interactive exercises[41], online role-playing games[42] (sometimes requiring specialized virtual reality hardware[43;44]), long videos[45] and extended writing tasks[46], while another thread of research suggests that outgroup bias can be overcome through accumulated life experiences that engender outgroup empathy[47].

However, even these promising interventions encounter three related issues. First, the "modal treatments" in these studies are perspective-taking treatments which include second-hand or imagined contact with outgroups[48]. These designs beg the question: outside of scenarios in which experimenters assign people to engage in empathy-related tasks, how can decisions to engage in empathy be encouraged in the first place? That is, how can we overcome what one recent review referred to as the "challenge of motivation"[49]? Second, the interventions fielded thus far are costly and time-consuming, often requiring careful training of enumerators and almost always additional costs in equipment, time and footwork. Finally, empathy arises more easily for ingroup than outgroup members[50;47] and some perspective-taking interventions can actually amplify egocentric biases[51]. This can lead to a vicious cycle that one scholar has called "the power of otherness to block empathy"[52]. This means that successfully encouraging empathy towards outgroups, specifically, is an even taller order than just encouraging it more broadly. Moreover, promoting only "parochial" empathy (towards one's own group) can be "counterproductive"[20] and is unlikely to lead to the beneficial pro-social outcomes associated with general empathy. In addition to these larger substantive concerns, current work often suffers from a number of methodological issues, including small sample size, lack of pre-registration and outcomes that are nearly exclusively short-term and attitudinal (only 7% of outcomes in a recent survey of the literature were behavioral[48]).

The detrimental effects of outgroup prejudice and the salutary properties of empathy prompt our research question: How can we encourage greater empathy (and foster inclusive attitudes/behaviors) from members of the dominant (white) group towards racial and ethnic out-groups? Our argument is that—given its documented effects on pro-social behavior—praise from peers is a promising candidate for promoting empathy as well as changing behaviors and attitudes towards racial outgroups. Moreover, leveraging praise draws upon peoples' natural resources—the desire for positive esteem and admiration of peers—answering the call of a recent review to "align interventions with key psychological motivations"[49]. Our argument focuses additionally on the role of positive feelings and social norms as causal mechanisms linking peer praise to its effects on empathy and behavioral inclusion.

Praise is a promising candidate for an intervention to promote empathy given its documented effects on pro-social behavior (particularly in the large literature on child development[53]). Definitions of praise abound but generally agree that the concept centers on "positive evaluations…of another's products, performances or attributes"[54] (see also[55]). Praise can be about behavior, effort or personal qualities, and can occur either ex-ante or ex-post whatever is being encouraged; in fact, the most common use of praise is in trying to increase effort levels, motivation or encourage specific behaviors in the future. It can be distinguished from highly similar concepts such as "encouragement" (often associated with tasks with which a person is currently struggling or in which they performed negatively) and simple acknowledgment/feedback, which is inherently neutral and non-judgmental[56;57]. In line with a consensus that views behavior and effort-specific praise as more effective than "personal praise" (focused on peoples' attributes or qualities,[58]), we focus on praise for engaging in empathy randomly assigned to our respondents in advance of their choice to engage in empathy or not (ex-ante in order to cleanly identify the effects of the treatment on behavior). While our focus on the connection between praise and pro-social behavior is not new, previous work has often centered on child-parent relationships (with an emphasis on adolescent populations[59]).

While praise is a promising candidate for an empathy-encouraging intervention, extant literature suggests lessons for how to operationalize praise by leveraging the importance of peers in social networks[60]. A multidisciplinary literature on peer effect processes portrays this group as increasingly important upon broaching adulthood, with "respected" peers and ones who share values particularly influential[59]. Peer effects occur across contexts as diverse as uptake of education, future planning, emotional happiness and economic and welfare outcomes[61;62;63;64;65]. This work dovetails with a parallel literature on social learning, which has demonstrated the importance of "learning the value of stimuli, actions or knowledge from others" for outcomes specifically related to empathy[66;67].

Peer effects have two additional features that make them particularly relevant for our purposes. First, peer influence is particularly effective in encouraging pro-social behaviors and attitudes, including warmth towards outgroups and inclusive behavior[68;69;70]. Paluck and colleagues, for example, tap into peer networks to encourage anti-conflict norms and behavior in a middle school setting[71]. One explanation for the efficacy of peer influence is that "external social motivations" can reduce "parochial empathy"; the tendency to empathize more with one's own group and (relatively) less with outgroups[20]. In fact, among a plethora of lukewarm (and sometimes negative) results from diversity programs, one set of promising results comes from drawing on people's desire to look good in front of their peers[72]. Second, and critically for our purpose, peer effects seem to persist over time[73;74]. This is particular helpful in the context of interventions to promote empathy, where durable effects are rarely found[48].

Our argument highlights two mechanisms—positive emotions and norms—through which praise from peers might affect empathy-related outcomes, each related to a different component of our intervention. We focus on emotions given the natural connection between empathy and affect more generally—empathy is fundamentally "an affective response"[75]—as well as its more particular role in the first part of the empathetic process, when individuals choose whether to engage in or avoid empathy[76]. Indeed, the link between praise and positive emotions is a baseline expectation in much of the literature[77]. This is summed up by Delin & Baumeister, who write that the "obvious and immediate outcome" of praise is "simple, positive affect"[78]. And while there are strong links between praise and positive emotions, there are also connections between positive emotions and many other beneficial outcomes[79;80;81;82;83], with some recent work suggesting a link between positive mood and pro-social behavior[84] that might operate as a feedback loop or "virtuous cycle"[85]. Other recent studies have conclusively demonstrated that triggering positive emotions (towards outgroups) leads to behaviors such as higher outgroup donations and likelihood of approaching rather than avoiding[86]. Our theory emphasizes the role of happiness—the most common positive emotion[87]—but we note the possibility of other positive emotions at play in the causal effect chain.

Our second possible mechanism originates in the "peer" component of our intervention and focuses on norms: beliefs about what others do or how they expect us to behave[88]. Since praise is inherently non-neutral, one pathway through which it might operate is by affecting beliefs beliefs about what is valued by the sender of the praise (here, one's peers)[56]. As noted above, work on social learning suggests learning from others in this way might be particularly potent, even for "complex social phenomena" such as empathy[89] and there is a "robust" connection between social norms related to diversity and intergroup attitudes and behaviors"[60].

One reason to suspect this causal mechanism is the relative importance of social norms, the power and reach of which "can hardly be overestimated"[90]. In fact, one recent work finds exactly this in a set of lab and field experiments: a norms treatment shifted the attitudes of non-marginalized groups towards outgroups and their overall feelings towards pro-diversity initiatives[70] (see also[91] and[60]). In another set of studies, Nook et al. find that "people imitate others' prosocial behaviors"; instituting norms of generosity and prosocial behavior affected helping behavior, donation amounts and even respondents' own feelings (suggesting that there may even be a link between our two posited mechanisms)[92]. More broadly, the strength and content of norms has been found to affect behavioral patterns across societies as well[93]. The "normative pathway" we focus on is close to, but subtly distinct from a "priming" mechanism in which peer praise increases the "mental accessibility" of the concept of empathy or pushes people to evaluate the target group or individual "through the lens" of empathy (see discussion in[94]).

In this registered report, we develop and tested a noninvasive, low-cost intervention harnessing natural peer influence to encourage empathy and political/attitudinal inclusion towards racial and ethnic outgroups. Our intervention is praise from peers about engaging in empathy and comes in the form of a word-cloud of real praise elicited in pilot studies as well as text describing average thermometer ratings towards people who are empathetic. Since making group identity salient risks triggering parochial empathy only towards ingroup members, our treatment promotes only general empathy towards others (with no mention of outgroups, specifically)[20].

As highlighted in our "Scale-up proposal for Peer Praise" Figure in Appendix C, our intervention is relatively easy and low-cost to adapt to new online or in-person contexts. For example, to test the intervention in a field setting, one would field a small survey to generate real peer praise for empathy—among whatever group one is interested in treating—and then conduct a baseline survey of attitudes and behaviors before using the collected peer praise intervention to encourage empathy (either online or in person). The overall implementation costs will be particularly low for the many contexts in which data on attitudes towards others is already collected (e.g., college orientations, DEI training at private companies and public institutions). The cost of the initial survey to generate praise is also small, and especially so when compared to in-person interventions which require between ten and twenty minutes per person of contact with trained enumerators, multiplied by thousands of people[38;95]. Our supplementary materials include a downloadable packet of instructions for researchers to adapt our intervention to other contexts.

The real peer praise that fuels our treatment comes from one of two sources in our studies: either a neutral group of "online peers" on MTurk (in our main Peer Praise study) or a more identity-relevant group of online co-partisans (in our parallel Co-Partisan Peer Praise study). In both cases, we are leveraging the "external social motivations" that recent reviews have suggested as promising to reduce the empathy gap towards outgroups[20]. In keeping with the concern that researchers asking hypothetical questions will be rewarded with hypothetical answers, our design combines naturalistic real praise with a range of behavioral outcome measures, including donation behavior, letter-writing to the White House and a validated behavioral task in which respondents make real decisions about whether to engage in empathy or avoid it[96;97], answering calls to focus our studies on behavior as well as attitudes[49].

In addition to measuring (empathy and political) behavior alongside attitudes, our research design incorporates three additional and novel features. First, our design takes up the challenge set out by two recent reviews[48;49] to investigate whether effects from empathy-related interventions persist over time (analogous to recent work on long-term effects of correcting misperceptions[98]). To that end, we field a longitudinal survey experiment in which respondents are re-interviewed one week after the initial treatment in our main study. This provides a more stringent test of the intervention while also helping to shed greater light on mechanisms. Second, our main study features a placebo in Wave 1, addressing potential concerns about how the peer praise treatment works and ruling out alternative explanations. Finally, our parallel Co-Partisan Peer Praise study uses a treatment in which praise comes from co-partisans, allowing us to investigate whether the efficacy of praise depends on how "identity-relevant" the source of praise is[99].

Our light-touch intervention and behavioral task have been extensively tested across seven pilot studies and a total of 2,466 subjects (14,704 observations accounting for multiple trials). Our approach of "rigorous incrementalism"—embodied by our seven pilot studies—established several crucial steps in the causal process, honed our peer praise treatment, and generated clear baselines against which we will be able to judge the efficacy of our treatment in encouraging empathy towards racial and ethnic outgroups. Specifically, our pilots found:

1. a general aversion to empathy that required a 10% premium in wages to overcome.

2. respondents perceived engaging in empathetic description as more demanding, costly, difficult and anxiety-producing than engaging in pure description.

3. evidence of the effectiveness of peer praise in motivating empathy towards generalized "others" (i.e., without respondents knowing the identity of the target).

4. evidence on the role of positive emotions and norms as causal mechanisms, both of which were supported to varying degrees in our pilots.

In this registered report, we focused on the effect of peer praise on empathy, attitudes and measures of behavioral and political inclusion. In all cases, our interest was in attitudes and behavior among white respondents in the U.S. towards Black Americans and non-white Latino/a respondents, both contexts where outgroup bias is particularly durable and pernicious. Our first set of six hypotheses—in our "Peer Praise Study"—investigated the direct, immediate effect of praise from a neutral group of online peers on empathy and attitudinal and behavioral inclusion (See Peer Praise Design Table, Figure 1). The "peers" in this part of our study were other MTurk workers, based in part on evidence that the MTurk community functions similarly to more traditional workplaces[100]. Online communities such as MTurk are increasing in scope and depth[101] and function as a community through which workers can sustain and build identities[102;103], build reputations[104], and even garner better wages[105].

H1 focuses on the outcome of empathetic behavior—preempting concerns about social desirability bias or impression management on the part of our respondents—while H2 uses a more common measure of self-reported empathy[106;107]. In line with work referenced earlier that links praise with positive emotions and prosocial behaviors as well as attitudinal warmth, we also posit that respondents who receive peer praise for empathy are more willing to engage in behavior that is politically-inclusive towards racial outgroups as well as view those same groups more warmly. H3 focuses on the effects of the peer praise treatment on political inclusion, for which we elicit two semi-behavioral outcomes: willingness to (1) donate to civil rights and racial justice advocacy groups (UnidosUS, Black Lives Matter) and to (2) advocate on behalf of racial/ethnic equity to the White House. Those two measures are combined into an index of "semi-behavioral" political inclusion[108;41;109]. H4 assesses the extent to which peer praise can promote attitudinal inclusion, measured through an index of (1) respondent reported social distance and (2) thermometer measures towards Black and Latino/a outgroup members[110;111;112].

Our main Peer Praise study also assessed the durability of our peer praise intervention via a longitudinal design in which respondents are re-contacted one week following treatment. As a recent review noted, one of the biggest challenges of interventions to improve attitudes and behaviors towards outgroups is whether proposed methods have any lasting effect[48]. We measure two outcomes for the longitudinal component of the study: empathetic behavior (H5) and an attitudinal index (H6).

Our Peer Praise study also includes two additional sets of analyses—exploratory in nature—designed to probe mechanisms and ensure the validity of our treatment. To assess our proposed mechanisms, we both measured respondents' level of happiness and their perceptions of norms around empathy, as well as use the longitudinal study to calibrate the plausibility of different causal paths. To the latter point, any average treatment effect (ATE) found in the longitudinal follow-up is less likely to be driven by (fleeting) positive emotions induced by our intervention. To ensure that the average treatment effect of peer praise was not driven by extraneous elements of the praise intervention—e.g., the colors of the text—our main study included a placebo arm in which respondents also receive peer praise, but for "description" rather than empathy. If the elements of the treatment that matter for promoting empathy are only presentational, then "praise for description" should also cause a higher likelihood of choosing to engage in empathy in our choice task. Below we summarize the first six registered hypotheses:

H1: Peer praise increases likelihood of choosing to empathize (compared to no peer praise).

H2: Peer praise increases self-reported empathy (compared to no peer praise).

H3: Peer praise increases political inclusion (compared to no peer praise).

H4: Peer praise increases attitudinal inclusion (compared to no peer praise).

H5: Peer praise increases empathetic behavior (measured 1 week following treatment) (compared to no peer praise).

H6: Peer praise increases empathetic attitudes (measured 1 week following treatment) (compared to no peer praise).

Our second set of pre-registered hypotheses concerned the source of praise and are tested via a parallel study entitled Co-Partisan Peer Praise (depicted in the Co-Partisan Peer Praise Design Table 2). While H1-H4 assessed the efficacy of a peer praise treatment in which the praise came from a neutral group of online peers, it's possible that online peers might be too diffuse a group to matter to respondents, particularly if the efficacy of the treatment in promoting empathy depends on how much respondents value the opinion of the source of praise. And while some extant work on peer effects in online networks[103;100] suggests optimism that more generic peer praise might promote empathy, it's still possible that praise from a highly salient social identity grouping might be more effective.

To explore whether co-partisan peer praise can affect empathy, political inclusivity, and attitudinal warmth towards racial/ethnic outgroups, we preregistered and fielded our Co-Partisan Peer Praise study, identical to Wave 1 of the Peer Praise study except for the source of the praise, which comes from co-partisans rather than MTurk workers (control group receives nothing). Co-partisans are a particularly important group in the United States, where partisan identity has been described as one of Americans' "most salient social identities"[113] (see also[114]). The outcomes for H7-H10—empathy behavior, self-reported empathy, political inclusion and attitudinal warmth—exactly parallel H1-H4 and are summarized below:

H7: Co-Partisan Peer praise increases likelihood of choosing to empathize (compared to no co-partisan peer praise).

H8: Co-Partisan Peer praise increases self-reported empathy (compared to no co-partisan peer praise).

H9: Co-Partisan Peer praise increases political inclusion (compared to no co-partisan peer praise).

H10: Co-Partisan Peer praise increases attitudinal inclusion (compared to no co-partisan peer praise).

Additional exploratory analyses investigate the relative efficacy of co-partisan peer praise compared to the general MTurk peer praise. And because the tendency to empathize correlates with partisan identity[115;116], we anticipate finding heterogeneous treatment effects by party in our exploratory analysis. Given that Democrats have been described as the party of inclusivity[117], one might expect Democratic respondents to respond more to praise for empathy than Republican respondents, though conversely if the correlation between party ID and empathy towards racial outgroups is high enough, it's possible that "ceiling effects" among Democrats might suggest the opposite prediction.

Results

We fielded our survey experiment Study 1 Wave 1 on N=5,303 adult respondents registered in the U.S. on MTurk in July and August 2025. 90.31% of Study 1, Wave 1 respondents were successfully recontacted for Wave 2 (an attrition rate between waves of 9.69%). Given the attrition of roughly a tenth of the respondents across waves in Study 1, we conducted registered Manski bound calculations on our hypotheses tested with longitudinal data (H5 and H6). Study 2 was fielded in August 2025 on N=4,404 respondents.

In our results below, we distinguish between analyses that were (1) registered and powered upon; i.e., our main hypotheses, H1-10 (2) registered but not powered upon (i.e., listed in registration) and (3) purely exploratory analyses. As per our registration, analyses only include respondents passed both attention checks and completed the survey. Summary regression tables for H1-H6 and H7-H10 can be found in Appendices D and E respectively.

Peer Praise for Empathy from Increases Political Inclusion Towards Outgroups while not changing attitudes or prompting empathy behavior (pre-registered)

Our first set of hypotheses concern the effect of the "peer praise for empathy" treatment on four outcomes: the choice to empathize in a behavioral task (H1), self-reported empathy (H2), political inclusion (H3) and attitudinal inclusion (H4). Our outcomes thus represent a broad range of attitudinal (H2, H4) and behavioral (H1, H3) outcomes.

We find no support for the intervention affecting attitudinal outcomes: specifically, we find null results for both self-reported empathy (H2: $\beta = -0.011$; 95% CI[-0.038, 0.017]; $p = 0.442$) and our index of attitudinal inclusion, which combines social distance and warmth towards racial and ethnic outgroups (H4: $\beta = -0.0031$; 95% CI[-0.013,0.007]; $p = 0.563$). We also find null results for the behavioral outcome "choosing to engage in empathy": respondents randomized into the peer praise for empathy treatment were no more likely than those in the Control arm to choose to engage in empathy with a racial or ethnic outgroup (H1: $\beta = -0.012$; 95% CI[-0.044,0.019]; $p = 0.437$).

While the peer praise intervention evinced no effects on any attitudinal measure and no effect on the choice to engage in empathy with a member of a racial or ethnic outgroup, it did cause an increase in our index of *political inclusion* (H3: $\beta = 0.031$; 95% CI[0.012,0.049]; $p = 0.00156$), which combines two outcomes: a donation task in which respondents pay a portion of their bonus to BLM or UnidosUS and a letter-writing task in which respondents write an anonymous letter to the White House in support of prioritizing racial and ethnic equity policies. We conducted Benjamini-Hochberg multiple-hypothesis adjustments and our findings remain the same: H1, H2, and H4 tests cannot reject the null; H3 test does reject the null with a BH adjusted p-value of 0.009. We also conducted Bayesian analysis for null treatment effects for H1, H2, and H4 and find evidence for the null hypotheses (Bayes Factors all less than 1).

Study 2 was designed in a parallel fashion to Study 1, but the "praise for engaging in empathy" originated from co-partisans rather than the more neutral group of online peers. We found null effects for all outcomes in this study: self-reported empathy (H8: $\beta = -0.004$; 95% CI[-0.035, 0.028]; $p = 0.809$), attitudinal inclusion (H10: $\beta = -0.004$; 95% CI[-0.015, 0.007]; $p = 0.455$), choice to engage in empathy with outgroup members (H7: $\beta = 0.006$; 95% CI[-0.032, 0.044]; $p = 0.744$) and political inclusion (H9: $\beta = 0.014$; 95% CI[-0.011, 0.039]; $p = 0.273$). We conducted Bayesian analysis for null treatment effects of H8-H10 and find evidence for the null hypotheses (Bayes Factors all less than 1).

Attrition Leads to the Appearance of Backlash Effects (pre-registered)

Our second set of hypotheses concerned the effects of our treatment—peer praise for empathy—on empathetic behavior (H5) and empathetic attitudes (H6) measured in Wave 2, one week after the initial study. Consistent with our evidence for null effects on attitudes above (H2 & H4), we find no evidence that treatment affected empathetic attitudes 1 week post treatment (H6: $\beta = 0.001$; 95% CI[-0.010, 0.013]; $p = 0.814$). Bayesian analysis for null treatment effects find evidence for the null with Bayes Factors all less than 1.

We do find preliminary evidence that the peer praise treatment led to a decrease, or backlash, against empathy in Wave 2 (H5: $\beta = -0.045$; 95% CI[-0.079, -0.010]; $p = 0.0116$). However, our protocol dictated that we conduct analyses that account for attrition; while the overall re-contact rate was high ($\approx 90\%$), attrition still occurred in our sample 1 week after. We calculate Manski bounds for the ATE (-0.263, 0.195) and find they include 0, suggesting that under the weakest assumptions (only that the outcome lies in the specified bounds— here zero and one for the behavioral empathy task) the data are consistent with negative effects, zero effect, or positive effects; in sum, the treatment effect is not point identified and could be zero. Thus, we cannot rule out a zero treatment effect without further assumptions (e.g., missing-at-random, monotone selection, or other restrictions).

How Peer Praise works to promote empathy, and for whom (registered but not powered)

While our main results above focus our attention on the null effects of peer praise for empathy on attitudinal outcomes, analyses that were registered (and in our design table) but not powered on show that the intervention was successful for certain subgroups. For example, we find evidence of heterogenous treatment effects in both waves of Study 1: the more attitudinally inclusive respondents were *prior* to treatment, the greater the effect of the treatment in Wave 1 (coefficient on interaction between peer praise treatment $\times$ prior attitudinal index: $\beta = 0.058$; 95% CI[0.020, 0.095]; $p = 0.002$). That is, for peer-praised respondents, a one point increase in the prior attitudinal index (halfway across the index) is associated with a 0.339 standard deviation shift upwards in the post-attitudinal index. Similarly, the more positive respondents' attitudes were prior to treatment in Week 1, the greater the effect of the treatment 1 week later on attitudes towards racial outgroups (coefficient on interaction between peer praise treatment $\times$ prior attitudinal index on long-term post-attitudinal index: $\beta = 0.072$; 95% CI[0.012, 0.133]; $p = 0.019$). Our results from Study 2 that were registered (but not powered) also suggest individual-level dimensions that might shape the efficacy of the treatment. Here, we found a heterogeneous treatment effect by political party: the effect of co-partisan peer praise on choices to engage in empathy were positive and significant for Democrats compared to a baseline party of Republicans (coefficient on interaction between peer praise treatment $\times$ political party=Democrat: $\beta = 0.084$; 95% CI[0.007, 0.160]; $p = 0.031$; baseline is Republican party).

Additional analyses that were in our registration but not powered on also shed light on potential mechanisms. For example, the positive effect of peer praise for empathy on political inclusion (H3) were not found to have been mediated by changing beliefs about norms or by happiness (see Appendix D for more detail).

How and Why the intervention works, cont'd (exploratory)

Exploratory analyses shed further light on the questions of who the treatment worked best for and why. While registered analyses showed that the effect of peer praise for empathy on political inclusion were not mediated by positive emotions (H3), exploratory analyses suggested that the negative emotions of anxiety and anger might be responsible for part of the effect (Anxiety Average Causal Mediation Effect $\beta = 0.019$; 95% CI[0.010, 0.03]; $p =< 2e - 16$; Anger Average Causal Mediation Effect $\beta = 0.012$; 95% CI[0.004, 0.02]; $p = 0.012$). And while our intervention did not lead to a significant main effect on the choice to engage in empathy with racial outgroups, respondents in our treatment group did engage in empathy more often with White/Latino faces compared to the control group.

Exploratory analyses also help us better understand for whom the intervention worked. There were no heterogenous treatment effects by respondent gender or age for all H1-H10 hypotheses. However, in our Co-Partisan Peer Praise (Study 2), our analyses suggest moderation of the co-partisan peer praise effect on attitudinal inclusion by education levels; i.e., respondents who had up to and including high school education responded to the co-partisan treatment more positively than counterparts with bachelor's or professional degrees (coefficient on interaction between co-partisan peer praise treatment $\times$ baseline up-to H.S. education: $\beta = 0.161$; 95% CI[0.014, 0.309]; $p = 0.031$; for treatment X up-to bachelors: $\beta = -0.164$; 95% CI[-0.312, -0.017]; $p = 0.029$; treatment X professional and highest degrees: $\beta = -0.172$; 95% CI[-0.321, -0.024]; $p = 0.023$;]).

Discussion

In this report, we tested a low-cost, scalable intervention designed to encourage empathy towards racial and ethnic outgroups. This was done via two parallel studies: Study 1, in which the peer praise for empathy originated from a neutral group of online peers, and Study 2, in which the praise for empathy came from ideological co-partisans (Democrats or Republicans). Study 1 also included a longitudinal aspect, in which respondents were re-contacted 1 week post-treatment. Across all studies, we measured a wide array of outcomes under the umbrella of purely attitudinal DVs (self-reported empathy for outgroups, attitudinal inclusion), a semi-behavioral measure (political inclusion, measured as donation and letter-writing behavior) and the behavioral choice to engage in empathy towards outgroups. We sampled US respondents 18 years or older via the MTurk platform in order to ensure fair respondent wages and to retain control over respondent data and track attrition over the course of the study. Because our interest was in majority group empathy for minority outgroups, we sampled white US respondents, and our outcomes were related to empathy towards Black/hispanic Americans.

Our study differed in several important ways from previous work. Unlike previous interventions, ours focused on encouraging empathy towards racial and ethnic outgroups; a much more difficult task than broadly promoting empathy or doing so solely for ingroups. We also sought to develop a treatment that would promote greater "approach to empathy" rather than assigning an empathy-related behavior to investigate its downstream effects. Our intervention was light-tough, cheap and scalable, in contrast to many existing approaches which are incredibly costly in resources, manpower and time. Overall, and to distinguish it from much extant work, our studies were highly powered, extensively piloted, pre-registered, included a broad array of behavioral and attitudinal outcomes and a longitudinal element. Our studies thus represent an unusually comprehensive test of an empathy-related intervention.

Our highly-powered study found null effects across most outcomes: peer praise for empathy did not lead to an increase in self-reported empathy, attitudinal inclusion or the choice to engage in empathy with racial/ethnic outgroups. However, we did find that the treatment worked to increase political inclusion, measured as donation behavior, towards outgroup causes and letter-writing (to the White House in support of racial/ethnic equality). We also found tentative evidence that the peer praise intervention worked selectively: for some groups, e.g., those higher in pre-treatment empathy for outgroups were more likely to respond positively to the treatment in Study 1 in both the initial study and the follow-up 1 week later. Similarly, in Study 2, Democrats were more likely to respond with greater empathy for outgroups when praised by co-partisans.

We turn now to what broader lessons can be distilled from our studies. The first lesson comes from the clear disjuncture between our highly powered and pre-registered pilot studies and our main studies. In the former, we found that peer praise for empathy had strong and consistent effects in encouraging respondents to actually engage in empathy; in the latter, we found null results across Studies 1 and 2 for that outcome. What explains the difference? The most obvious cause is a critical change in how we structured our design: in the pilots, our peer praise treatment encouraged empathy towards generalized "others," but respondents made the choice to engage in empathy *before* seeing the faces of the potential targets. In the main studies—as a more challenging test of the intervention—they see photos of Black/hispanic faces and then made their choice, leading to null effects. This—along with null results along our attitudinal measures—reinforces the difficulty of encouraging empathy towards others when their identity as an outgroup is highlighted. In fact, in line with other work, we found that, while our peer praise for empathy intervention did not promote the choice to engage more in empathy with outgroups, it did prompt more empathy with ingroup (white) faces, a dynamic which has the potential to widen the "outgroup empathy gap."

The second lesson is a hopeful one. While promoting empathy towards racial outgroups is a difficult task, it is not impossible. Our modest intervention—quick, easy to scale and cheap—successfully promoted inclusionary political behavior towards racial outgroups on the part of white subjects, a result driven specifically by increases in letters written advocating racial and ethnic equality to the White House. If our normative commitments include both empathy on an individual level as well as policies designed to address inequality, then our treatment should be viewed as successfully encouraging the latter, a result that is particularly relevant to research and practice on generating political coalitions to address societal problems.

If the second lesson emphasizes that promoting empathy towards even racial/ethnic outgroups is possible, the final lesson concerns the question of who is most amenable to encouragement. Specifically, our work suggests the importance of proper targeting of "nudges" to encourage empathy: the intervention worked better for those that were higher in pre-treatment empathy towards outgroups and worked better when Democrats were encouraged to engage in empathy for racial outgroups by co-partisans. In fact, among those individuals that were higher in pre-treatment empathy, we found a long-term effect of the treatment—on attitudes towards racial minorities—after 1 week. All of these results point to the importance of targeting specific groups for interventions rather than relying on a "one size fits all" approach. Praise from some groups might work better than others, and some individuals may be more amenable to one class of encouragement than another group.

Methods

Ethics Information

All studies in this project received IRB approval and exemption from the University of Wisconsin Madison Educational and Social/Behavioral Science IRB (# 2020-0843-CP002). Informed consent was obtained for all pilot studies and will be obtained for the registered study. Participants in any of the pilot studies are prevented from re-enrolling in any other related study run by authors. Participants were paid according to the highest minimum wage across U.S. states (Washington) in 2020 on the online platform from which they were recruited (Amazon Mechanical Turk, or MTurk, for all pilots with the exception of Pilot 7, in which respondents were drawn from Harvard Digital

Lab for the Social Sciences (DLABSS)). Participants for the registered study are similarly recruited from MTurk. We aim to provide fair wages, especially given prior work indicating a median wage of MTurk workers at around \$2/hour[118]. For the current pre-registered study, we will follow this approach and pay subjects the highest minimum wage during survey administration.

Pilot Data

To demonstrate the feasibility of our approach, hone our intervention and collect data to refine our power analysis, we conducted seven pilot studies (summarized in Figure 1), with $n = 2,466$ respondents and a total of $N = 14,704$ respondent-task trials. The pilots all used a choice-based behavioral empathy task (the first of our major outcomes in this design and described in more detail below) and did the following:

- Pilot 1: allowed us to verify a general preference towards avoiding empathy
- Pilot 2: enhanced the realism of our peer praise intervention
- Pilot 3: proposed and test our light-touch intervention designed to encourage empathy through the use of peer praise
- Pilots 4 and 5: investigated the affective and norms-based mechanisms through which praise work
- Pilot 6: honed power calculations for testing the effect of peer praise on decisions to engage in empathy towards outgroups
- Pilot 7: preliminarily tested the effects of co-partisan peer praise towards outgroups

Pilot details are in Section A of SI.

Design

The substantive interest in our two parallel studies study is in assessing whether peer praise can encourage white Americans—the majority racial group in the United States—to empathize with and increase political and affective inclusion towards racial/ethnic outgroup members (here, Black or Hispanic individuals). Our design also allows us to estimate whether peer praise effects on empathy are durable over time, while our parallel Co-Partisan Peer Praise study focuses on whether the source of the peer praise matters for promoting empathy-related outcomes. The general procedure is illustrated in the consort diagram in Figure 4.

The core of our experimental design for our Peer Praise study was the randomization of peer praise in a between-subjects design. Respondents were randomized into Control and Peer Praise (Arms A and B) each with probability 0.4 (Arm C is a placebo arm utilized for a robustness check and assigned with probability 0.2). When respondents received peer praise, it was presented to respondents in the form depicted in Figure 2(a) and consisted of a word cloud of praise from peers—alongside numerical ratings—for people who engage in empathy. The word cloud uses word stems to group similar words, a commonly employed approach to avoid redundancy and to provide a concise summary of related words (see[119] and[120] among many others for examples of similar vocabulary processing and visualization approaches; see[121] for a summary). For example, the words “friendly” and “friendship” are often stemmed to the same root “friend”. Respondents in the control condition do not receive any additional information. Figure 2(b)-(c) present the peer praise interventions used for the Co-Partisan Peer Praise study.

Treatment was followed by a set of outcomes measuring empathy and political/attitudinal inclusion, randomized in order. One outcome was a (piloted and validated) behavioral empathy task in which respondents see images of either Black/Hispanic/white individuals (randomized and depicted as being drawn from a deck of cards) and then make the decision to either engage in empathy—by describing how the person depicted feels—or simply describe what the person looks like. The faces represent "racialized groups"[122] validated by the Chicago Faces Database and in our pilots[123]. Other outcomes in Wave 1 included self-reported empathy, donation willingness, letter writing, social distance, and thermometer ratings, detailed in the Outcomes section. Wave 2 followed up a week later with Arms A and B to assess long-term effects of peer praise (testing H5-H6).

All participants completed pre-treatment demographic questions upon study entry (as well as questions on race/ethnicity, political party, and prior measures of inclusion; see SI for details). All respondents were also asked several post-task questions near the end of the survey on religion, ideology, Trump approval and Biden approval, racial resentment (borrowed from the American National Election Studies[124]), and zipcode.

Placebo: To further investigate how our treatment works, and rule out explanations centered on extraneous features such as the color of the word cloud in the treatment, we added a third "placebo" Arm (C) in the Peer Praise Study (Figure 4). Here, respondents received an intervention resembling peer praise — information in the form of a word cloud and ratings by peers — but displaying a word cloud (and similar paired language on peer attitudes) for peer praise for objective behavior. If the elements of the treatment that matter are only presentational, then "praise for description" should cause a higher likelihood of choosing to engage in empathy in our choice task (compared to control).

Parallel Study on Source of Praise: To test hypotheses about co-partisan peer praise effects on empathy and inclusion, we fielded a parallel study (Co-Partisan Peer Praise Study; see Figure 5). Treated respondents received peer praise from co-partisans, while control respondents received no stimulus, each with probability 0.5. This second independent study ~~explicitly~~ explored the peer praise effects of co-partisan peers on empathy and inclusion and was explicitly designed and powered to test H7-10.

Outcomes: Our focus was on behaviors and attitudes towards racial and ethic outgroups. Outcomes consisted of decisions to engage in empathy, attitudinal warmth (adapted from Williamson, Enos and others[107;111]), and finally, decisions regarding political inclusion (letter-writing and donation towards civil advocacy groups). All outcomes were presented in randomized order.

We measured engagement in empathy in two ways:

1. Behavioral empathy task: a behavioral choice to empathize with an outgroup image.

   Similar to the pilot studies, images were drawn from the Chicago Faces Database (CFD)[123] and randomized among the following features, representing racialized groups as perceived and validated previously[123]: Race=Black/Hispanic/white and Valence=Angry/Fearful. We used images that rank in the highest quintile valence for "angry" and "fearful" for each group, as coded by independent raters within the CFD. We used faces that displayed emotions because we require that respondents have something to empathize with in the first place. We selected angry/fearful faces for two reasons. First, statistical power concerns dictated the need to specify some specific emotions rather than randomizing across the emotional spectrum. Second, empathizing with negative emotions is inherently a more "aversive"[125] experience than empathizing with positive emotions, so their selection helps avoid putting our thumbs on the scale by making empathy artificially easier. Because our main hypothesis was concerned with white respondents' empathy towards images of Black/Hispanic individuals, we randomize images (in the behavioral task) such that the probability of seeing an image of a Black or Hispanic person is $p = 0.9$, and the probability of seeing an image of a white person is $p = 0.1$ (see SI B.1 for examples).

All respondents completed one practice trial of the behavioral empathy task prior to making their choice. In the practice trial, they engaged in both empathy and description. The practice trial both familiarized respondents with the task as well as guarded against the possibility that their choices in the study might be the result of misperceptions about the difficulty of the two different options.

2. Self-reported empathy: We asked respondents how much they agreed with the statements (1) "I feel empathy for the grievances expressed by Black and Hispanic Americans about how they are treated by police", (2) "I feel empathy for the grievances expressed by Black and Hispanic Americans about how they are treated by fellow citizens", (3) "I feel empathy for the grievances expressed by Black and Hispanic Americans about the hardships they face due to discrimination by race and ethnicity". We combined the responses to the statements into a single "Self-reported empathy" outcome measure (normalizing each and adding to an index taking values from 0 to 3).

We measured political inclusion towards racial and ethnic outgroups in the following two ways, and combined the measures into a single "political inclusion" index.

3. Donate: we offered all respondents a bonus and ask them whether they would be willing to pay a portion of their bonus (on a scale of $0-0.5, in increments of 10 cents) to the BLM movement and UnidosUS. We randomized the order of the organizations. This question is a functional equivalent of a dictator game with the two organizations as the "other" and the bonus as the allocation; see[108] as an example, and[109] for a donation-only version. We focused on the Black Lives Matter movement and UnidosUS given their scope of coverage across the United States, their historical roots, and the firm connection to advocating for Black and Hispanic racial/ethnic justice[126;127;128;129]. We combined the donation outcomes for BLM and UnidosUS into a single "Donation" outcome measure (taking values from 0 to 1).

4. Letter: we asked respondents if they would be willing to write an anonymous letter to the White House in support of prioritizing racial and ethnic equity policies in the United States, and provided them open-ended space to write such a letter should they choose to (see Adida and colleagues[41] for a similar behavioral measure). We plan to submit written letters to the White House anonymously on behalf of the respondents.

We asked two types of attitudinal inclusion outcomes that measured social distance and warmth towards racial and ethnic outgroups. These were then combined into a single attitudinal index in our analyses, by normalizing each measure and then combining them into a single index taking values from 0 to 2.

5. Social distance: We borrowed our scale from Enos[111], which measures individuals' exclusionary preferences by capturing their willingness to share social spaces with a member of another group.

6. Thermometer: measure for how warm the respondent feels towards each racial and ethnic outgroup.

Following all the outcome variable questions, respondents were asked a battery of questions to determine their task preference, such as: how they selected between the two tasks (FEEL/DESCRIBE), whether they developed a preference for one task, and a version of the NASA Task Load Index[130], which measured "subjective mental workload."

Mechanisms: Respondents were also asked questions designed to shed light on the mechanisms through which peer praise affected our treatments. We asked two questions designed to capture whether a norm-based mechanism was plausible: respondents were asked whether they believed peers of theirs on the platform valued objectivity or empathy more, and how much they agreed with the following statement, "Most people agree that being open to understanding and considering other cultures' and identities' struggles and empathizing with them is important." Respondents were finally asked about their emotional states immediately after the treatment as a measure for our alternative mechanism: positive emotional valence.

Subsequently, respondents completed empathy questions[131] and the Interpersonal Reactivity Index[96], along with ethnocentrism inquiries[41;132], consistent with pilot studies. A manipulation check asks if respondents heard that platform peers favor individuals who engaged in empathy [Yes/No/Not sure]. Additionally, respondents indicated if they considered other MTurkers as peers, learn from them, and value their opinions. They verified the perceived race/ethnicity of images viewed, reflected on peers on the platform, and answered a debriefing question on treatment and outcomes.

Sampling Plan

We recruited US respondents who were 18 years or older with MTurk. Our survey instrument was programmed in Qualtrics, which was then fielded on the MTurk platform. From there, we sampled among MTurk workers registered as residing in the U.S. MTurk is platform that delivers a reasonably demographically diverse, high-quality set of respondents who are directly recruited[133]. It is likely to over-sample digitally-literate respondents and under-sample low-skill respondents[134], but as we discuss below, we particularly value the transparency of MTurk compared with other survey vendors.

While other providers (such as Lucid) have become more popular in recent years, we selected MTurk for three principal reasons: its reliability, its transparency and to maintain consistency with our extensive piloting of our treatment. In general, MTurk quality is high and, in fact, MTurk is still the main baseline comparison against which new survey firms are compared for quality: Coppock & McClellan, for example, use MTurk to assess the utility of alternative platforms[135]. MTurk's marketplace is also among the largest of online sample platforms, and its quality of respondents and their efforts is particularly high for respondents who have clear directions, are paid a fair wage for their work[136], and who are "vetted" (for example, for attention and work history on MTurk[137]). To comply with these best practices, we ~~have~~ piloted our instructions for each aspect of the tasks many times, provided above-average (on the platform) wages for tasks—based on actual highest minimum hourly wages in the U.S.—and vetted MTurk workers for attention (in our study) and success in previous MTurk tasks.

Fielding our study on MTurk also helps to assuage recent concerns over transparency on other platforms. The key issue here is that many more expensive survey vendors (e.g., YouGov or Lucid) do not actually maintain or control their own panels, and instead outsource them to subcontractors (and even sub-sub-contractors) without providing any information on how the panel is constructed (and in many cases, only providing "completes": responses from those who finished the full survey). In these cases, it can be difficult or impossible to know who is being sampled, how sampling is being conducted and which respondents attrite (as well as why they attrite)[138;139]. Finally, fielding on MTurk maintains consistency with our pilots in a critical way: our intervention specifically referenced "peers like you on this platform," so by using MTurk for the pre-registered study, we kept the source of praise constant in our experiment.

Our desired sample size for both peer praise and co-partisan peer praise experiments was suggested by power analyses conducted in R (see Figure 6 and SI for further details). In the power calculations, we made assumptions for our simulated experiments around expected (main) effect sizes for peer praise and their standard deviations (SD). Wherever possible, we took values for these from our extensive pilots; when not possible, we estimated these parameters from studies that were as similar as possible along key dimensions (e.g., similar setups and same outcome).

For H1, we took effect and SD information from Pilot Study 6, which also examined the effect of peer praise on our behavioral choice outcome (0.17, $SD = 0.47$). Because we did not have a piloted peer praise study to reference for the DV of "self-reported empathy" (H2), we borrowed effect size/SD (0.171, $SD = 1$) from Williamson et al.[107], which utilized a similar outcome to ours (self-reported empathy), though a different intervention (family history perspective).

H3 includes two semi-behavioral outcomes, which we combined into one political inclusion index. For the "donation" outcome, we again borrow from a Williamson et al.[107] study in which the outcome is "expressed support for more open policies" (0.1, $SD = 0.06$). For the "letter" outcome, we borrowed effect size/SD (0.165, $SD = 0.8$) information from a similar outcome in Adida et al.[41]—a letter to the White House. We combine "Donation" and "Letter" into a single "Political Inclusion" Index in which "Donation" and "Letter" take on equal weight; the Index ranges from $[0, 2]$.

For H4, we used two attitudinal measures—social distance and a thermometer—and indexed them into one attitudinal outcome index. Because we did not have an equivalent study to reference for Social Distance, we assumed a conservative increment of 0.1 rank move (in a $0 - 6$ rank scale) with a similar sized SD (0.25, $SD = 0.1$). For the Thermometer outcome, we borrowed effect/SD information (2.4, $SD = 0.08$) from a similar outcome in Williamson et al.[107]. We combined Social Distance and Thermometer into a single Attitudinal Outcome Index in which Social Distance and Thermometer took on equal weight; the Index ranges from $[0, 2]$. H5 and H6 focus on the Behavioral Empathy Task and Attitudinal Index in a longitudinal context. We assumed long-run effects would be reduced by $\frac{1}{4}$, and that the follow-up in Wave 2 would suffer from 50% attrition. In the peer praise study, our allocation ratio between groups was 40:40:20 for Control, Peer Praise and Placebo groups. Our power calculations account for Benjamini-Hochberg multiple-hypothesis adjustments for H1 through H6.

For the Co-Partisan Peer Praise parallel study, we did not have priors on whether peer praise from co-partisans would be stronger or weaker in its effects compared to peer praise from a generalized peer group. As a result, we made the same assumptions of peer praise (main) effect sizes as before for the four tests H7-H10 tests (main effects of co-partisan peer praise). Power calculations account for Benjamini-Hochberg multiple-hypothesis adjustments for H7-H10.

Because our registered hypotheses are for white subjects—and some proportion of respondents recruited will not be white—our simulations and power calculations considered both the overall number of subjects as well as the "usable" number of subjects. Because our study is focused on promoting empathy among the dominant group towards racial and ethnic outgroups, our subjects must self-identify as white. In our pilots, an average of 66% of respondents on the survey platform self-identify as white, which is factored into our power calculations (i.e., 33% of recruited subjects would not be "usable" for our purposes). We simulated datasets, and in each simulated data set, we estimated tests and evaluate MH adjusted p-values. From this, we discover how often, under each possible N, we woulld be able to recover treatment effects (for H1-H6 in the peer praise study, and H7-H10 for the Co-Partisan peer praise study) at a statistical significance level of $\alpha = 0.05$. We aimed for a power level above 0.95. Finally, the peer praise study required a follow-up panel design, for which we conservatively assume a 50% re-contact rate.

Our power analysis suggested targeting an N size of 5,300 for our peer praise study to achieve a power level of 0.95; similarly, we required N=4,400 for our co-partisan peer praise for H7-10. This resulted in two designs totaling N=9,700 respondent observations. These numbers represent the total subjects recruited and account for rates of attrition, and hypotheses which focus only on certain subgroups (such as white Americans or co-partisans).

Analysis Plan

The main hypotheses we tested for in Study 1 Peer Praise are depicted in Peer Praise Design Table (Figure 1). We were powered to test our first 6 hypotheses in this study, which pertain to the effects of peer praise on empathetic behavior towards racial/ethnic outgroups (H1 and H2), on political inclusion (H3), on attitudinal inclusion (H4), and whether these effects persist over time (H5 and H6). We hypothesized that (white) respondents who receive the randomized peer praise treatment are more likely to choose to empathize and self-report empathizing with racial/ethnic outgroup individuals (where outgroup always means "randomized to receive Black/Hispanic image") than white respondents who do not receive peer praise (H1 and H2). H3 explored peer praise effects on political inclusion, expecting treated respondents to score higher on a political inclusion index (based on donation and letter-writing outcomes). H4 assessed attitudinal inclusion, hypothesizing that treated respondents would rate higher on an attitudinal index (based on social distance and thermometer measures). H5 and H6 posited that peer praise would impact behavioral empathy and attitudes in the longer term, following the one week delay before re-contacting and eliciting follow-up measurements of outcomes, mediators, manipulation checks and post outcome variables.

H7-10 were tested in Study 2 (Co-Partisan Peer Praise design, Figure 5). These hypotheses posited that co-partisan peer praise influences behavioral empathy, self-reported empathy, political inclusion, and attitudinal inclusion. For H1-H10, we employed OLS regression to estimate the effect of peer or co-partisan peer praise on each outcome. We use OLS with adjustments for any chance imbalances in pre-treatment covariates (e.g., age, education, political party)[140]. Pre-measures of attitudinal inclusion were controlled in H5 and H6 tests for precision. Further details are provided in the Design Tables 1-2.

Analyses were conducted on all respondents who complete the survey and pass both attention checks. Participants were excluded from analyses if they (a) did not provide a response and are thus missing variables involved in the analysis or (b) failed either of the pre-treatment attention checks. We included two types of pre-treatment attention checks: multiple-choice and screening questions in a grid[141], as well as filler questions in the grid that were designed to elicit non-politically oriented opinions from respondents so as to minimize possible priming effects downstream.

Robustness checks were conducted on an alternative control condition — a placebo (Peer Praise Study Figure 4 Arm C) with a similar length messaging intervention offering praise for objectivity (rather than empathy). Checks were also done on how much attrition has occurred throughout the experiment (as well as in the wave 2 panel survey portion); should there be significant covariate imbalance in treatment arms due to differential attrition, we will further conduct Manski bound estimations for the treatment effect[142].

Contingency plans

Our extensive piloting should assuage numerous concerns about issues that might arise in our study. For example, one worry with a study like this is that respondents might interpret the treatment or measure of behavioral empathy in an unexpected manner, leading to potential compliance issues. In our pilot studies, however, we were able to verify that our treatment and behavioral empathy task worked in the manner they were designed to. We did this through piloting of the study more generally, but also through manipulation checks, debriefing questions and assessing potential variation in outcomes across studies that took place at different stages of the COVID pandemic (pandemic vs post-ish pandemic). In our experiments, we include debriefing questions to allow us to assess whether respondents are considering the treatment in different light across time and study.

However, there are still several concerns might arise in the implementation of our design that we have considered and prepared for in advance. The first and most obvious potential issue is participant attrition. We plan to examine differential attrition using an `R` attrition package `attritevis` created for the express purpose of assessing attrition in experiments[139]. The package allows for visualization of attrition over time in the experiment—allowing easy diagnosis of potentially problematic missingness—as well as the implementation of balance tests and the estimation of bounds.

We have also incorporated above-average and reasonable wages in the design in order to ward off attrition[143]. This is in line with, and no different from our usual procedure which has been followed in our pilot studies (see SI B), which experienced little to no attrition.

We used an α level of 0.05 to conduct statistical significance testing and construction of 95% confidence intervals. For tests in which we cannot reject the null hypothesis, we will conduct Bayesian analysis for null treatment effects using the "BayesFactor" package in R, using a noninformative Jeffreys prior on the variance of the normal population, and a Cauchy prior on the standardized effect size. We utilize standard ranges for Bayes Factors (BF) from[144] for interpretation: BFs$<$1 as evidence for the null hypothesis, $1-3$ as anecdotal evidence for the alternative hypothesis, $3-10$ as moderate evidence for the alternative, $10-30$ as strong evidence for the alternative, $30-100$ as very strong evidence for the alternative, and BF$>$100 as extreme evidence for the alternative.

Protocol Registration

The link to the approved Stage 1 protocol can be found here: `https://springernature.figshare.com/articles/journal_contribution/Can_praise_from_peers_promote_empathy_and_political_inclusion_towards_racial_or_ethnic_outgroups_Registered_Report_Stage_1_Protocol_/28715198?file=53375999`

Data Availability

All data (inclusive of collected anonymized experimental data and CFD faces data used) and materials are available on the Github website at this link: `https://github.com/adelinelo/Praise-and-Empathy`.

Code Availability

All analysis code are openly available on the Github website `https://github.com/adelinelo/Praise-and-Empathy`.

Acknowledgments

For comments and suggestions that aided the manuscript, the authors thank Claire Adida, Donghyun Danny Choi, Scott Clifford, Erin Hartman, Deborah Kanter, Rose McDermott, Yoav Nygate, Héctor Pifarré i Arolas, Dan Posner, Michelle Schwarze, Cigdem V. Sirin, and Jessica Weeks.

J.R. was supported by the Institute for Humane Studies as well as a grant from the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation and the University of Wisconsin-Madison Trice Faculty Award. A.L. was supported by funding provided by the Office of the Vice Chancellor for Research at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author Contributions Statement

A.L. and J.R. conceived and designed the experiments and contributed materials/analysis tools. A.L. and L.B.N. performed the experiments and analyzed the data. A.L., J.R. and L.B.N. wrote the paper.

Competing Interests Statement

The authors declare no competing interests.

Figure Legends/Captions

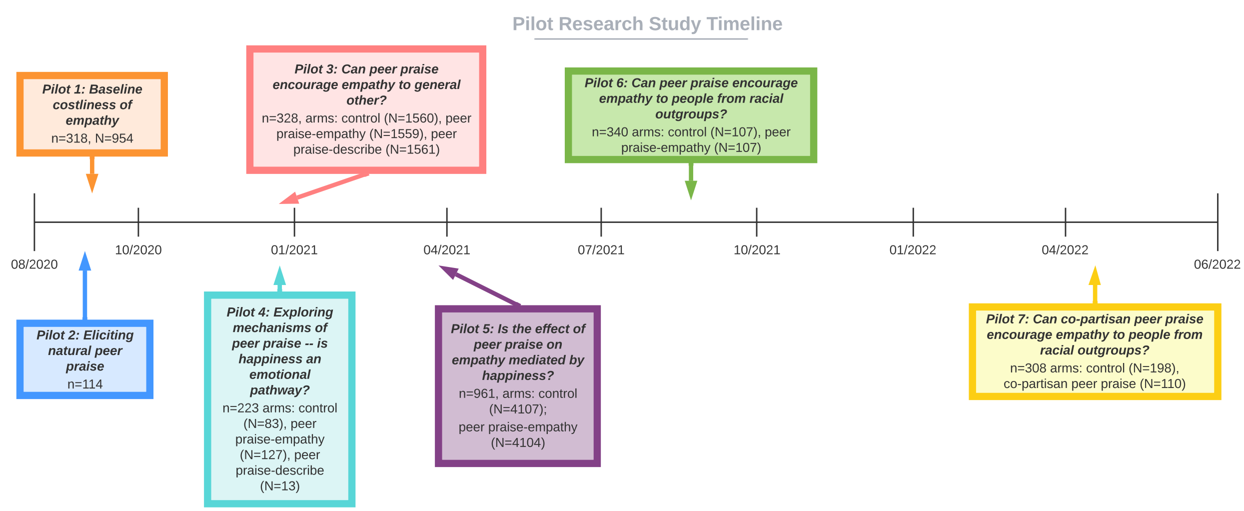

Figure 1: Pilot study timeline. Number of observations are presented for each pilot as N, while number of respondents as n, for studies with multiple trials of the main task per respondent.

(a) Peer praise (general).

(b) Peer praise (Democrat).

(c) Peer praise (Republican).

Figure 2: Real peer praise gathered for engaging in empathetic behavior in panel (a); Real peer praise by co-partisan group Democrats (b) and Republicans (c).

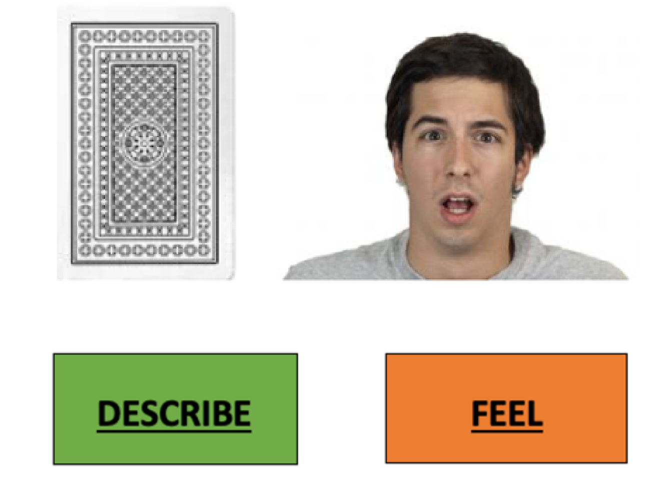

Figure 3: Top: Behavioral Empathy Task illustration. Example of an image (Race = white, Valence = Fearful) randomly drawn from a deck of cards, presented to respondents before choice task. Bottom: Respondents choose between DESCRIBE and FEEL buttons to select their task.

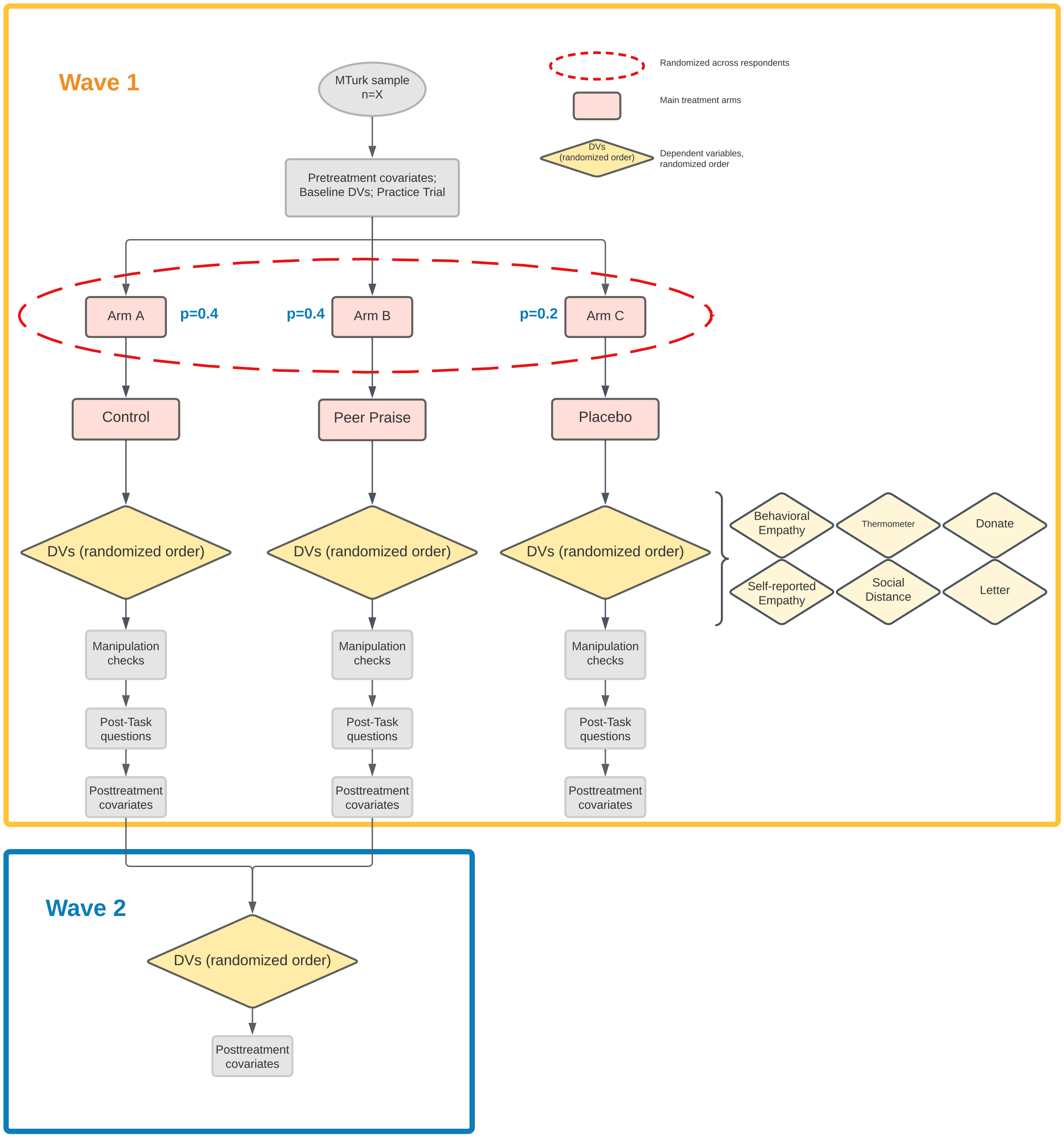

Figure 4: Peer Praise Study Consort Diagram. The design features three arms (A, B, C) where the first two arms are Control and Peer Praise, and the last is Placebo. All respondents are asked pretreatment covariate questions and complete a practice trial of the behavioral empathy task, after which they are randomized into one of the three A, B, C arms with probabilities (0.4, 0.4, 0.2). Respondents then see each of the outcome questions ("DVs") in random order, then finish with questions about task difficulty and preference and posttreatment covariates. This completes the Wave 1 survey. Wave 2, fielded a week later, follows up with respondents in Arms A and B to survey DVs and posttreatment covariates again.

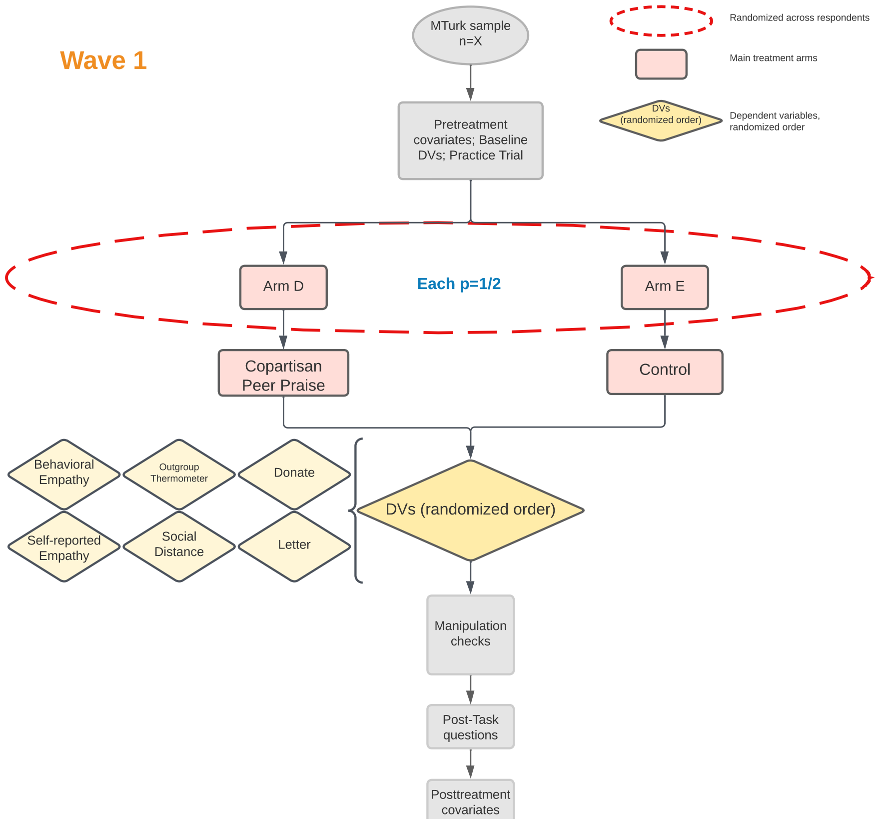

Figure 5: Co-Partisan Peer Praise Consort Diagram. The design features two main treatment arms, Co-Partisan Peer Praise and Control. All respondents are asked pretreatment covariate questions and complete a practice trial of the behavioral empathy task, after which they are randomized into one of the two arms with equal probability. Respondents then see each of the outcome questions ("DVs") in random order, then finish with questions about task difficulty and preference and posttreatment covariates.

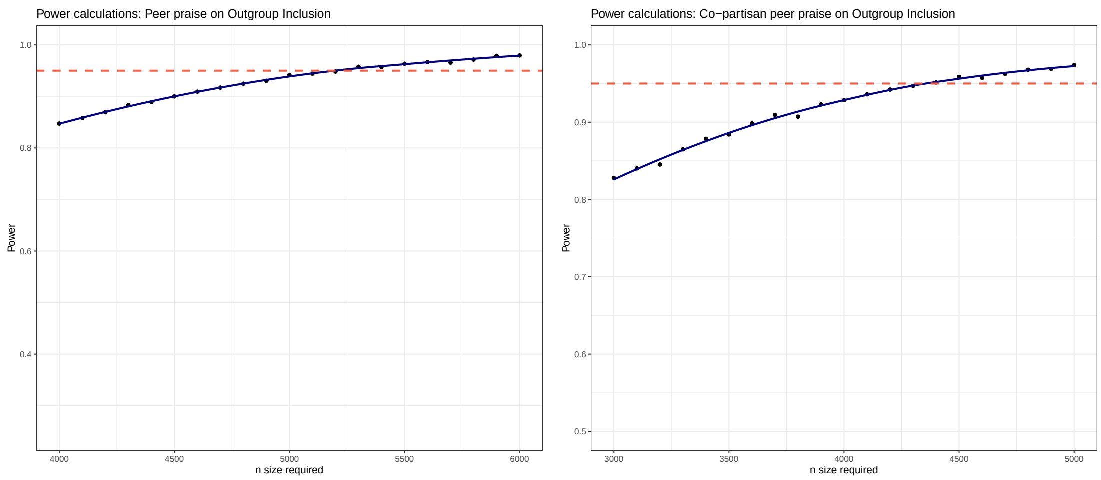

Figure 6: Power analyses. Left panel: Study 1 (Peer Praise Study) Power analysis. X-axis is N size, Y-axis is power level. Data points represent average power from 10,000 simulations at a given N level for multiple-hypothesis adjusted H1 through H6. To power at 0.95 for all 12 hypotheses, we require N=5,300. Right panel: Study 2 (Co-Partisan Peer Praise Study) Power analysis. X-axis is N size, Y-axis is power level. Data points represent average power from 10,000 simulations at a given N level for multiple-hypothesis adjusted H7 through H10. To power at 0.95 for all 4 hypotheses, we require N=4,400.

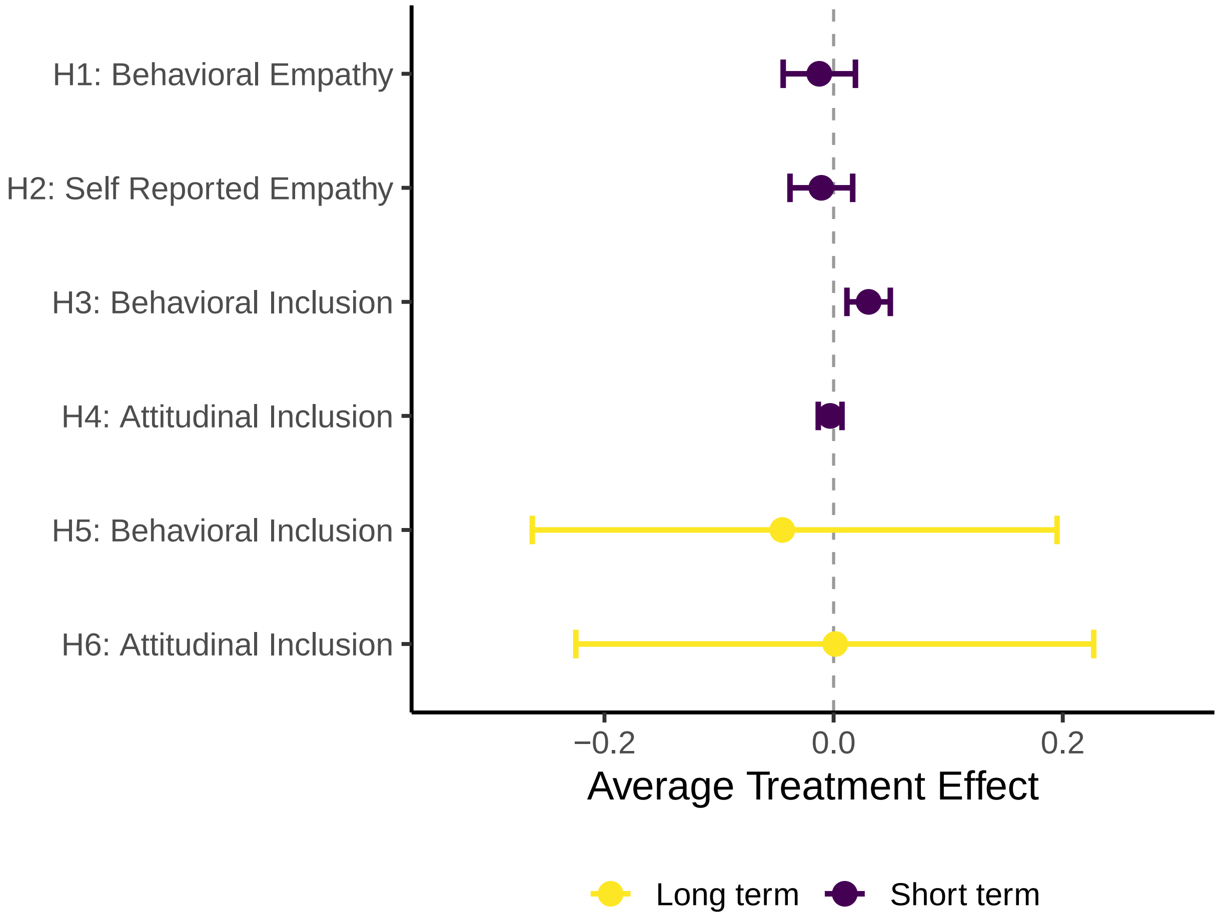

Figure 7: Study 1 Estimated Average Treatment Effects (ATE) H1-H6. X-axis is estimated ATE of peer praise, Y-axis are hypotheses tested. For each hypothesis except H5 and H6 the ATE is plotted with 95% CI; for H5 and H6 we plot the Manski bounds for ATE which take into account attrition experienced in the follow-up. Wave 1 hypotheses are in purple, while long-run Wave 2 hypotheses are in yellow.

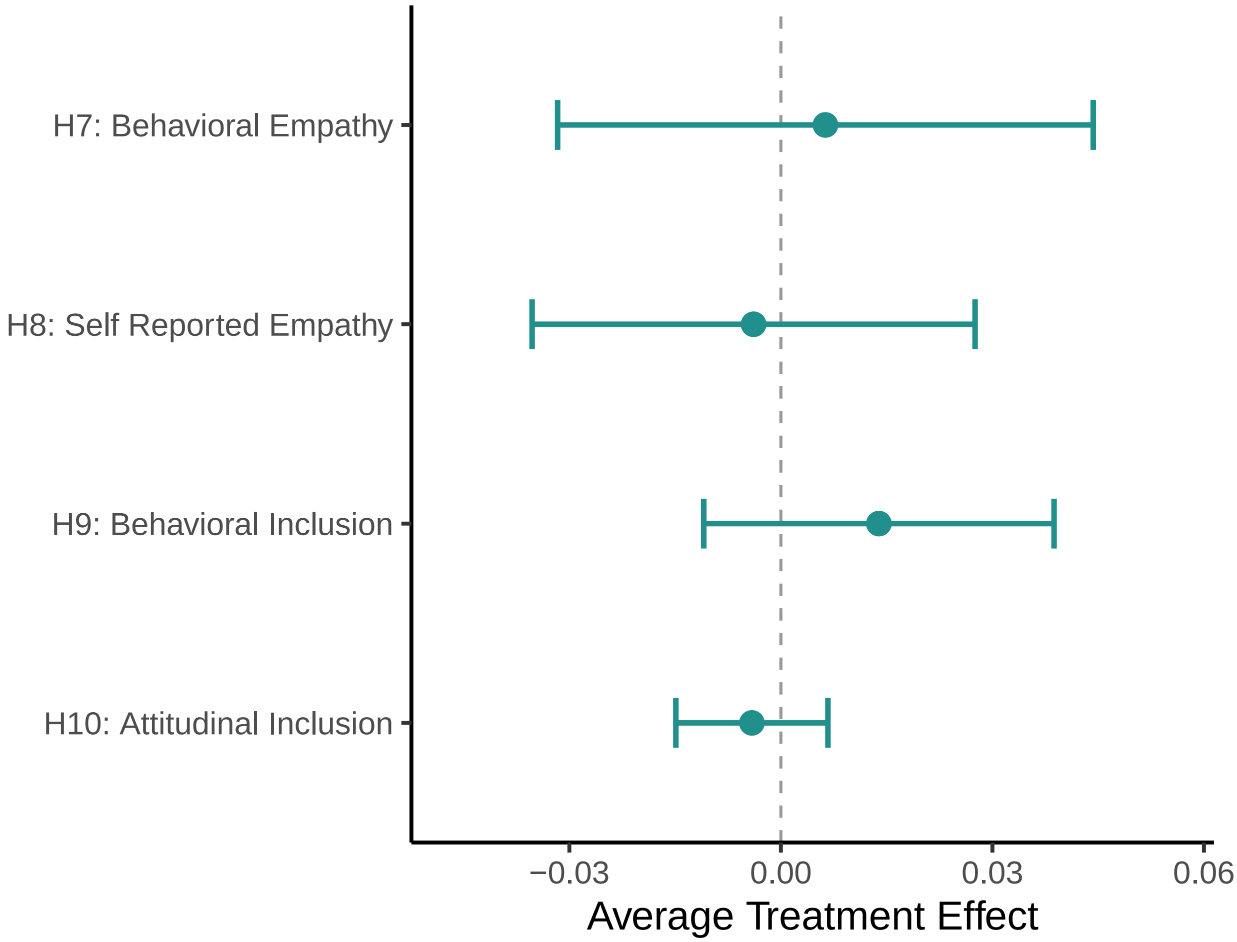

Figure 8: Study 2 Estimated Average Treatment Effects (ATE) H7-H10. X-axis is estimated ATE of peer praise, Y-axis are hypotheses tested. For each hypothesis the ATE is plotted with 95% CI.

Tables

Question	Hypothesis	Sampling Plan	Analysis Plan	Interpretation Given to Different Outcomes
Can peer-praise increase choosing to empathize with a racial/ethnic outgroup?	H1: White respondents who receive peer praise are more likely to empathize with racial/ethnic outgroup members than respondents who do not receive peer praise.	We sample N=5,300 respondents based on our R power analysis which suggests N=5,300 as sufficient to recover a main effect assuming peer praise effects of 0.17 and SD=0.47 (from Pilot Study 6) at a power level of 0.95 and significance level of 0.05, accounting for multiple hypothesis testing H1-H6. We recruit US respondents 18+ years, who have not participated in any of our pilot studies, and have above-average approval rating with MTurk. We exclude from our analysis respondents who do not pass both attention checks. We follow this sample inclusion rule for all H1-H10. We run this test on white respondents who see a racial/ethnic outgroup image.	OLS coef test of Peer Praise on behavioral empathy task. Robustness check: OLS regression with controls for pre-treatment covariates that are imbalanced. Similar robustness check used for all H1-H10.	If p-value ≤ 0.05, we interpret it as evidence supporting the alternative hypothesis of a peer praise effect on empathy. Otherwise, Bayesian analysis for null treatment effects is conducted using the "BayesFactor" package in R, using a noninformative Jeffreys prior on the variance of the normal population, and a Cauchy prior on the standardized effect size. We use Bayes Factors ranges from Kass & Raftery 1995: BF<1 supports null, 1-3 is anecdotal evidence for alternative, 3-10 moderate, 10-30 strong, 30-100 very strong, and BF>100 extreme evidence for alternative.
	H2: White respondents receiving peer praise are more likely to report empathizing with racial/ethnic outgroup members, compared to respondents who did not receive peer praise.	We sample N=5,300 respondents based on R power analysis, assuming hypothesized peer praise effects of 0.171 and SD=1 (from Williamson et al. 2021) at a power level of 0.95, significance level 0.05, and accounting for multiple hypothesis testing of H1-H6.	OLS coef test of Peer Praise on self-reported empathy, an index created from three agreement statements, normalized and summed.	If p-value ≤ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a peer praise effect on self-reported empathy. Otherwise, we conduct the same Bayesian analysis as above.
Can peer-praise increase political inclusion towards a racial/ethnic outgroup?	H3: White respondents who receive peer praise are more likely to act politically inclusively towards racial/ethnic outgroup members, compared to respondents who don't receive peer praise.	We sample N=5,300 based on R power analysis, assuming effect size for donation (0.1, SD=0.06 from Williamson et al. 2021) and letter (0.165, SD=0.8 from Adida et al. 2018) at a desired power level of 0.95, significance level 0.05 and accounting for multiple hypothesis testing H1-H6.	OLS coef test of Peer Praise on political inclusion index that includes the donation (summed responses to BLM and UnidosUS) and letter outcomes, normalized/weighted equally.	If p-value ≤ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a peer praise effect on political inclusion. Otherwise, we conduct the same Bayesian analysis as above.
Can peer-praise increase attitudinal inclusion towards a racial/ethnic outgroup?	H4: White respondents who receive peer praise are more likely to be attitudinally inclusive towards racial/ethnic outgroup members, compared to respondents who did not receive peer praise.	We sample N=5,300 based on R power analysis, assuming effect size for Social Distance (0.25, SD=0.1, no direct equivalent study, so we assume an increment of 0.1 rank move with a similar sized SD) and Thermometer (2.4, SD=0.08 from Williamson et al. 2021) at a desired power level of 0.95, significance level 0.05, and accounting for multiple hypothesis testing of H1-H6.	OLS coef test of Peer Praise on attitudinal index composed of social distance and thermometer outcomes, normalized and weighted equally. Robustness check adds pre-treatment attitudinal index.	If p-value ≤ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a peer praise effect on attitudes. Otherwise, we conduct the same Bayesian analysis as above.
Can peer-praise increase long-run choosing to empathize with a racial/ethnic outgroup?	H5: A week after intervention, white respondents who receive peer praise are more likely to choose to empathize with racial/ethnic outgroup members than respondents who do not receive peer praise.	We sample N=5,300 based on R power analysis, assuming a 25% decrease in effect size from H1 and 50% recontact rate in Wave 2 follow-up, a power level of 0.95, significance level 0.05, and accounting for multiple hypothesis testing of H1-H6.	OLS coef test of long-term Peer Praise on behavioral empathy task.	If p-value ≤ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a long-run peer praise effect on behavioral empathy. Otherwise, we conduct the same Bayesian analysis as above.
Can peer-praise increase long-run attitudinal inclusion towards a racial/ethnic outgroup?	H6: A week after intervention, white respondents who receive peer praise rate attitudinal index higher towards racial/ethnic outgroup members than respondents who do not receive peer praise.	We sample N=5,300 based on R power analysis, assuming a 25% decrease in effect size from H4, 50% recontact rate in Wave 2 follow-up, power level of 0.95, significance level 0.05, and multiple hypothesis testing of H1-H6.	OLS coef test of long-run Peer Praise on attitudinal index. Robustness check adds pre-treatment attitudinal index.	If p-value ≤ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a long-run peer praise effect on attitudes. Otherwise, we conduct the same Bayesian analysis as above.

Table 1: Peer Praise Design Table

Question	Hypothesis	Sampling Plan	Analysis Plan	Interpretation Given to Different Outcomes
Can co-partisan peer-praise increase choosing to empathize with a racial/ethnic outgroup?	H7: White respondents who receive co-partisan peer praise are more likely to empathize with racial/ethnic outgroup members than respondents who do not receive peer praise.	We sample N=4,400 respondents based on our R power analysis which suggests N=4,400 as sufficient to recover a main effect assuming co-partisan peer praise effects of 0.17 and SD=0.47 (from Pilot Study 6) at a power level of 0.95 and significance level of 0.05, accounting for multiple hypothesis testing H7-H10. We recruit US respondents 18+ years, who have not participated in any of our pilot studies, and have above-average approval rating with MTurk. We exclude from our analysis respondents who do not pass both attention checks. We follow this sample inclusion rule for all H1-H10. We run this test on white respondents who see a racial/ethnic outgroup image.	OLS coef test of Co-Partisan Peer Praise on behavioral empathy task. Robustness check: OLS regression with controls for pre-treatment covariates that are imbalanced. Similar robustness check used for all H1-H10.	If p-value $\leq$ 0.05, we interpret it as evidence supporting the alternative hypothesis of a co-partisan peer praise effect on behavioral empathy. Otherwise, Bayesian analysis for null treatment effects is conducted using the “BayesFactor” package in R, using a noninformative Jeffreys prior on the variance of the normal population, and a Cauchy prior on the standardized effect size. We use Bayes Factors ranges from Kass & Raftery 1995: BF$<$1 supports null, 1-3 is anecdotal evidence for alternative, 3-10 moderate, 10-30 strong, 30-100 very strong, and BF$>$100 extreme evidence for alternative.
	H8: White respondents receiving co-partisan peer praise are more likely to report empathizing with racial/ethnic outgroup members, compared to respondents who did not receive peer praise.	We sample N=4,400 respondents based on R power analysis, assuming hypothesized peer praise effects of 0.171 and SD=1 (from Williamson et al. 2021) at a power level of 0.95, significance level 0.05, and accounting for multiple hypothesis testing of H7-H10.	OLS coef test of Co-Partisan Peer Praise on self-reported empathy, an index created from three agreement statements, normalized and summed.	If p-value $\leq$ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a co-partisan peer praise effect on self-reported empathy. Otherwise, we conduct the same Bayesian analysis as above.
Can co-partisan peer-praise increase political inclusion towards a racial/ethnic outgroup?	H9: White respondents who receive peer praise are more likely to act politically inclusively towards racial/ethnic outgroup members, compared to respondents who don’t receive peer praise.	We sample N=4,400 based on R power analysis, assuming effect size for donation (0.1, SD=0.06 from Williamson et al. 2021) and letter (0.165, SD=0.8 from Adida et al. 2018) at a desired power level of 0.95, significance level 0.05 and accounting for multiple hypothesis testing H7-H10.	OLS coef test of Co-Partisan Peer Praise on political inclusion index that includes the donation (summed responses to BLM and UnidosUS) and letter outcomes, normalized/weighted equally.	If p-value $\leq$ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a co-partisan peer praise effect on political inclusion. Otherwise, we conduct the same Bayesian analysis as above.
Can co-partisan peer-praise increase attitudinal inclusion towards a racial/ethnic outgroup?	H10: White respondents who receive co-partisan peer praise are more likely to be attitudinally inclusive towards racial/ethnic outgroup members, compared to respondents who did not receive peer praise.	We sample N=4,400 based on R power analysis, assuming effect size for Social Distance (0.25, SD=0.1, no direct equivalent study, so we assume an increment of 0.1 rank move with a similar sized SD) and Thermometer (2.4, SD=0.08 from Williamson et al. 2021) at a desired power level of 0.95, significance level 0.05, and accounting for multiple hypothesis testing of H7-H10.	OLS coef test of Peer Praise on attitudinal index composed of social distance and thermometer outcomes, normalized and weighted equally. Robustness check adds pre-treatment attitudinal index.	If p-value $\leq$ 0.05, we interpret this as evidence in favor of the alternative hypothesis of a co-partisan peer praise effect on attitudes. Otherwise, we conduct the same Bayesian analysis as above.

Table 2: Co-Partisan Peer Praise Design Table

References

[1] Ruffle, B. J. & Sosis, R. (2006). Cooperation and the in-group-out-group bias: A field test on Israeli kibbutz members and city residents. *Journal of Economic Behavior & Organization*, 60(2), 147–163. https://doi.org/10.1016/j.jebo.2004.07.007

[2] Moy, J. & Ng, S. H. (1996). Expectation of outgroup behaviour: can you trust the outgroup? *European Journal of Social Psychology*, 26(2), 333–340. https://doi.org/10.1002/(SICI)1099-0992(199603)26:2¡333::AID-EJSP747¿3.0.CO;2-1

[3] Pager, D., Bonikowski, B. & Western, B. (2009). Discrimination in a low-wage labor market: A field experiment. *American Sociological Review*, 74(5), 777–799. https://doi.org/10.1177/000312240907400505

[4] Karlsen, S. & Nazroo, J. Y. (2002). Relation between racial discrimination, social class, and health among ethnic minority groups. *American Journal of Public Health*, 92(4), 624–631.

[5] Arnold, D., Dobbie, W. & Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, 133(4), 1885–1932. https://doi.org/10.1093/qje/qjy012

[6] Goncalves, F. & Mello, S. (2021). A few bad apples? Racial bias in policing. *American Economic Review*, 111(5), 1406–1441. https://doi.org/10.1257/aer.20181607

[7] Saluja, B. & Bryant, Z. (2021). How implicit bias contributes to racial disparities in maternal morbidity and mortality in the United States. *Journal of Women's Health*, 30(2), 270–273.

[8] Stepanikova, I., Triplett, J. & Simpson, B. (2011). Implicit racial bias and prosocial behavior. *Social Science Research*, 40(4), 1186–1195. https://doi.org/10.1016/j.ssresearch.2011.02.004

[9] Pierson, E., Camelia, S., Jan, O., Corbett-Davies, S., Jenson, D., Shoemaker, A., Vignesh, R., Barghouty, P., Phillips, C., & Shroff, R. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, 4(7), 736–745.

[10] Acharya, A., Blackwell, M. & Sen, M. (2016). The political legacy of American slavery. *The Journal of Politics*, 78(3), 621–641. https://doi.org/10.1086/686631

[11] Hartman, T. K., Newman, B. J. & Bell, C. S. (2014). Decoding prejudice toward Hispanics: Group cues and public reactions to threatening immigrant behavior. *Political Behavior*, 36(1), 143–163.

[12] Freelon, D., Marwick, A. & Kreiss, D. (2020). False equivalencies: Online activism from left to right. *Science*, 369(6508), 1197–1201. https://doi.org/10.1126/science.abb2428

[13] Adida, C. L., Lo, A., Rhee, I. & Williams, S. (2023). Do equity, diversity and inclusion (EDI) requirements change student political attitudes? *Politics, Groups, and Identities*, 1–10.

[14] House, The White. (2021). Executive Order on Diversity, Equity, Inclusion, and Accessibility in the Federal Workforce. *The White House*. Available at: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/06/25/executive-order-on-diversity-equity-inclusion-and-accessibility-in-the-federal-workforce/.

[15] House, The White. (2021). Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government. *The White House*. Available at: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/.

[16] House, The White. (2021). Executive Order on Advancing Equity, Justice, and Opportunity for Asian Americans, Native Hawaiians, and Pacific Islanders. *The White House*. Available at: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/28/executive-order-on-advancing-equity-justice-and-opportunity-for-asian-americans-native-hawaiians-and-pacific-islanders/.

[17] Decety, J. & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science*, 9(5), 525–537. https://doi.org/10.1177/1745691614545130

[18] de Waal, F. B. M. (2012). The antiquity of empathy. *Science*, 336(6083), 874–876. https://doi.org/10.1126/science.1220999

[19] Batson, C. D., Chang, J., Orr, R. & Rowland, J. (2002). Empathy, attitudes, and action: Can feeling for a member of a stigmatized group motivate one to help the group? *Personality and Social Psychology Bulletin*, 28(12), 1656–1666.

[20] Behler, A. M. C. & Berry, D. R. (2022). Closing the empathy gap: A narrative review of the measurement and reduction of parochial empathy. *Social and Personality Psychology Compass*, 16(9), e12701.

[21] Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, 140(6), 1608.

[22] Davis, M. H. (2006). Empathy. In J. E. Stets & J. H. Turner (Eds.), *Handbook of the sociology of emotions* (pp. 443–466). Springer.

[23] Muradova, L. & Arceneaux, K. (2022). Reflective political reasoning: Political disagreement and empathy. *European Journal of Political Research*, 61(3), 740–761.

[24] Stotland, E. (1969). Exploratory investigations of empathy. In *Advances in Experimental Social Psychology* (Vol. 4, pp. 271–314). Elsevier.

[25] Batson, C. D. (2014). *The altruism question: Toward a social-psychological answer*. Psychology Press.

[26] Batson, C. D., Early, S. & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin*, 23(7), 751–758.

[27] Batson, C. D. & Ahmad, N. Y. (2009). Using empathy to improve intergroup attitudes and relations. *Social Issues and Policy Review*, 3(1), 141–177.

[28] Gubler, J. R., Karpowitz, C. F., Monson, J. Q., Romney, D. A. & South, M. (2022). Changing hearts and minds? Why media messages designed to foster empathy often fail. *The Journal of Politics*, 84(4), 2156–2171.

[29] Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a nuclear concept. *Journal of Counseling Psychology*, 28(2), 91.

[30] Baron, H., Blair, R., Choi, D. D., Gamboa, L., Gottlieb, J., Robinson, A. L., Rosenzweig, S., Turnbull, M. & West, E. A. (2021). Can Americans depolarize? Assessing the effects of reciprocal group reflection on partisan polarization. *OSF Preprints*.

[31] Bayram, A. B. & Holmes, M. (2020). Feeling their pain: Affective empathy and public preferences for foreign development aid. *European Journal of International Relations*, 26(3), 820–850. https://doi.org/10.1177/1354066119890915

[32] DeSante, C. D. & Smith, C. W. (2020). Fear, institutionalized racism, and empathy: The underlying dimensions of Whites' racial attitudes. *PS: Political Science & Politics*, 53(4), 639–645. https://doi.org/10.1017/S1049096520000414

[33] Faulkner, N. (2018). "Put yourself in their shoes": Testing empathy's ability to motivate cosmopolitan behavior. *Political Psychology*, 39(1), 217–228.

[34] Galinsky, A. D. & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, 78(4), 708.

[35] Weyant, J. M. (2007). Perspective taking as a means of reducing negative stereotyping of individuals who speak English as a second language. *Journal of Applied Social Psychology*, 37(4), 703–716.

[36] Shih, M., Wang, E., Trahan Bucher, A. & Stotzer, R. (2009). Perspective taking: Reducing prejudice towards general outgroups and specific individuals. *Group Processes & Intergroup Relations*, 12(5), 565–577.

[37] Broockman, D. & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282), 220–224. https://doi.org/10.1126/science.aad9713

[38] Kalla, J. L. & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2), 410–425. https://doi.org/10.1017/S0003055419000923

[39] Tropp, L. R. & Barlow, F. K. (2018). Making advantaged racial groups care about inequality: Intergroup contact as a route to psychological investment. *Current Directions in Psychological Science*, 27(3), 194–199. https://doi.org/10.1177/0963721417743282

[40] Vorauer, J. D. & Petsnik, C. (2024). Empathy and perspective-taking interventions in intergroup contexts. In E. Halperin, B. Hameiri & R. Litman (Eds.), *Psychological Intergroup Interventions: Evidence-based Approaches to Improve Intergroup Relations* (pp. 31–44). Routledge.

[41] Adida, C. L., Lo, A. & Platas, M. R. (2018). Perspective taking can promote short-term inclusionary behavior toward Syrian refugees. *Proceedings of the National Academy of Sciences*, 115(38), 9521–9526.

[42] Simonovits, G., Kezdi, G. & Kardos, P. (2018). Seeing the world through the other's eye: An online intervention reducing ethnic prejudice. *The American Political Science Review*, 112(1), 186–193.

[43] Van Loon, A., Bailenson, J., Zaki, J., Bostick, J. & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PloS One*, 13(8), e0202442.

[44] Herrera, F., Bailenson, J., Weisz, E., Ogle, E. & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PloS One*, 13(10), e0204494.

[45] Pashak, T. J., Conley, M. A., Whitney, D. J., Oswald, S. R., Heckroth, S. G. & Schumacher, E. M. (2018). Empathy diminishes prejudice: Active perspective-taking, regardless of target and mortality salience, decreases implicit racial bias. *Psychology*, 9(06), 1340.

[46] Bientzle, M., Eggeling, M., Kanzleiter, M., Thieme, K. & Kimmerle, J. (2021). The impact of narrative writing on empathy, perspective-taking, and attitude: Two randomized controlled experiments on violations of Covid-19 protection regulations. *PloS One*, 16(7), e0254501.

[47] Cikara, M. & Fiske, S. T. (2011). Bounded empathy: Neural responses to out-group targets' (mis)fortunes. *Journal of Cognitive Neuroscience*, 23(12), 3791–3803. https://doi.org/10.1162/jocn_a_00069

[48] Paluck, E. L., Porat, R., Clark, C. S. & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, 72, 533–560.

[49] Ćehajić-Clancy, S. & Halperin, E. (2024). Advancing research and practice of psychological intergroup interventions. *Nature Reviews Psychology*, 3(9), 574–588.

[50] Sirin, C. V., Valentino, N. A. & Villalobos, J. D. (2021). *Seeing Us in Them: Social Divisions and the Politics of Group Empathy*. Cambridge University Press.

[51] Eyal, T., Steffel, M. & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology*, 114(4), 547.

[52] Segal, E. A., Gerdes, K. E. & Lietz, C. A. (2017). *Assessing Empathy*. Columbia University Press.

[53] Waugh, W., Brownell, C. & Pollock, B. (2015). Early socialization of prosocial behavior: Patterns in parents' encouragement of toddlers' helping in an everyday household task. *Infant Behavior and Development*, 39, 1–10.

[54] Kanouse, D. E., Gumpert, P. & Canavan-Gumpert, D. (1981). The semantics of praise. In *New Directions in Attribution Research* (Vol. 3, pp. 97–115).

[55] Kelly, F. D. & Daniels, J. G. (1997). The effects of praise versus encouragement on children's perceptions of teachers. *Individual Psychology*, 53(3), 331.

[56] Henderlong, J. & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin*, 128(5), 774–795.

[57] Pety, J., Kelly, F. D. & Kafafy, A. E. A. (1984). The Praise-Encouragement Preference Scale for Children. *Individual Psychology*, 40(1), 92–101.

[58] Mueller, C. M. & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, 75(1), 33.

[59] Brechwald, W. A. & Prinstein, M. J. (2011). Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on Adolescence*, 21(1), 166–179. https://doi.org/10.1111/j.1532-7795.2010.00721.x

[60] Murrar, S. & Brauer, M. (2024). Using social norms to promote positive relations between social groups. In E. Halperin, B. Hameiri & R. Littman (Eds.), *Psychological Intergroup Interventions* (pp. 45–57). Routledge.

[61] Duflo, E. & Saez, E. (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics*, 85(1), 121–148. https://doi.org/10.1016/S0047-2727(01)00098-6

[62] Duflo, E., Dupas, P. & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5), 1739–1774.

[63] Fowler, J. H. & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ*, 337, a2338. https://doi.org/10.1136/bmj.a2338

[64] Bertrand, M., Luttmer, E. F. P. & Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3), 1019–1055. https://doi.org/10.1162/003355300554971

[65] Carrell, S. E., Fullerton, R. L. & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3), 439–464. https://doi.org/10.1086/600143

[66] Pärnamets, P., Espinosa, L. & Olsson, A. (2020). Physiological synchrony predicts observational threat learning in humans. *Proceedings of the Royal Society B*, 287(1927), 20192779.

[67] Pan, Y., Novembre, G. & Olsson, A. (2022). The interpersonal neuroscience of social learning. *Perspectives on Psychological Science*, 17(3), 680–695.

[68] Barry, C. M. & Wentzel, K. R. (2006). Friend influence on prosocial behavior: The role of motivational factors and friendship characteristics. *Developmental Psychology*, 42(1), 153–163. https://doi.org/10.1037/0012-1649.42.1.15

[69] Balconi, M. & Canavesio, Y. (2013). Emotional contagion and trait empathy in prosocial behavior in young people: The contribution of autonomic (facial feedback) and Balanced Emotional Empathy Scale (BEES) measures. *Journal of Clinical and Experimental Neuropsychology*, 35(1), 41–48. https://doi.org/10.1080/13803395.2012.742492

[70] Murrar, S., Campbell, M. R. & Brauer, M. (2020). Exposure to peers' pro-diversity attitudes increases inclusion and reduces the achievement gap. *Nature Human Behaviour*, 4(9), 889–897.

[71] Paluck, E. L., Shepherd, H. & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3), 566–571. https://doi.org/10.1073/pnas.1514483113

[72] Dobbin, F. & Kalev, A. (2022). *Getting to diversity: What works and what doesn't*. Harvard University Press.

[73] Duflo, E., Dupas, P. & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5), 1739–1774.

[74] Lépine, A. & Estevan, F. (2021). Do ability peer effects matter for academic and labor market outcomes? *Labour Economics*, 71, 102022.

[75] Hoffman, M. L. (1984). Interaction of affect and cognition in empathy. In C. E. Izzard, J. Kagan & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 103–131). Cambridge University Press.

[76] Sirin, C. V., Valentino, N. A. & Villalobos, J. D. (2017). The social causes and political consequences of group empathy. *Political Psychology*, 38(3), 427–448.

[77] Morton, J. S., Mikolajczak, M. & Luminet, O. (2020). New perspectives on the praise literature: Towards a conceptual model of compliment. *Current Psychology*, 1–13.

[78] Delin, C. R. & Baumeister, R. F. (1994). Praise: More than just social reinforcement. *Journal for the Theory of Social Behaviour*, 24(3), 219–241.

[79] Erez, A. & Isen, A. M. (2002). The influence of positive affect on the components of expectancy motivation. *Journal of Applied Psychology*, 87(6), 1055–1067.

[80] Foo, M.-D., Uy, M. A. & Baron, R. A. (2009). How do feelings influence effort? An empirical study of entrepreneurs' affect and venture effort. *Journal of Applied Psychology*, 94(4), 1086–1094.

[81] Oswald, A. J., Proto, E. & Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics*, 33(4), 789–822. https://doi.org/10.1086/681096

[82] Storbeck, J., Dayboch, J. & Wylie, J. (2019). Fear and happiness, but not sadness, motivate attentional flexibility: A case for emotion influencing the ability to split foci of attention. *Emotion*, 19(4), 655–664.

[83] Subramaniam, K. & Vinogradov, S. (2013). Improving the neural mechanisms of cognition through the pursuit of happiness. *Frontiers in Human Neuroscience*, 7, Article 452. https://doi.org/10.3389/fnhum.2013.00452

[84] Aknin, L. B., Van de Vondervoort, J. W. & Hamlin, J. K. (2018). Positive feelings reward and promote prosocial behavior. *Current Opinion in Psychology*, 20, 55–59.

[85] Layous, K., Nelson, S. K., Kurtz, J. L. & Lyubomirsky, S. (2017). What triggers prosocial effort? A positive feedback loop between positive activities, kindness, and well-being. *The Journal of Positive Psychology*, 12(4), 385–398. https://doi.org/10.1080/17439760.2016.1198924

[86] Ćehajić-Clancy, S., Jamshed, N., Olsson, A. & Momčilović, A. (2024). From inspiration to restoration: Moral elevation as a catalyst for improving intergroup relations in contexts of conflict. *Journal of Personality and Social Psychology.*

[87] Davidson, R. J. (1993). The neuropsychology of emotion and affective style. In *Handbook of Emotion* (pp. 143–154). Guilford Press.

[88] House, B. R. (2018). How do social norms influence prosocial development? *Current Opinion in Psychology*, 20, 87–91.

[89] Olsson, A., Knapska, E. & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, 21(4), 197–212.

[90] McDonald, R. I. & Crandall, C. S. (2015). Social norms and social influence. *Current Opinion in Behavioral Sciences*, 3, 147–151.

[91] Stangor, C., Sechrist, G. B. & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin*, 27(4), 486–496.

[92] Nook, E. C., Ong, D. C., Morelli, S. A., Mitchell, J. P. & Zaki, J. (2016). Prosocial conformity: Prosocial norms generalize across behavior and empathy. *Personality and Social Psychology Bulletin*, 42(8), 1045–1062.

[93] House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebioglu, S., Crittenden, A. N., Erut, A., Lew-Levy, S., Sebastian-Enesco, C. & Smith, A. M. (2020). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour*, 4(1), 36–44.

[94] Chong, D. & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10, 103–126.

[95] Kalla, J. L. & Broockman, D. E. (2023). Which narrative strategies durably reduce prejudice? Evidence from field and survey experiments supporting the efficacy of perspective-getting. *American Journal of Political Science*, 67(1), 185–204.

[96] Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E. & Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, 148(6), 962–976. https://doi.org/10.1037/xge0000595

[97] Ferguson, A. M. & Inzlicht, M. (2022). Reliability of the empathy selection task, a novel behavioral measure of empathy avoidance. *Behavior Research Methods*, 54, 1–14.

[98] Carnahan, D., Bergan, D. E. & Lee, S. (2021). Do corrective effects last? Results from a longitudinal experiment on beliefs toward immigration in the US. *Political Behavior*, 43(3), 1227–1246.

[99] Simas, E. N., Clifford, S. & Kirkland, J. H. (2020). How empathic concern fuels political polarization. *American Political Science Review*, 114(1), 258–269.

[100] Martin, D., Hanrahan, B. V., O'Neill, J. & Gupta, N. (2014). Being a turker. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 224–235).

[101] Scholz, T. (2012). *Digital labor: The internet as playground and factory.* Routledge.

[102] Irani, L. (2015). Difference and dependence among digital workers: The case of Amazon Mechanical Turk. *South Atlantic Quarterly*, 114(1), 225–234.

[103] Wang, X., Zhu, H., Li, Y., Cui, Y. & Konstan, J. (2017). A community rather than a union: Understanding self-organization phenomenon on Mturk and how it impacts Turkers and requesters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2210–2216).

[104] Brawley, A. M. & Pury, C. L. S. (2016). Work experiences on MTurk: Job satisfaction, turnover, and information sharing. *Computers in Human Behavior*, 54, 531–546.

[105] Kaplan, T., Saito, S., Hara, K. & Bigham, J. (2018). Striving to earn more: A survey of work strategies and tool use among crowd workers. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (Vol. 6, pp. 70–78).

[106] Arceneaux, K. (2017). Anxiety reduces empathy toward outgroup members but not ingroup members. *Journal of Experimental Political Science*, 4(1), 68–80.

[107] Williamson, S., Adida, C. L., Lo, A., Platas, M. R. & Prather, L. (2021). Family matters: How immigrant histories can promote inclusion. *American Political Science Review*, 115(2), 686–693.

[108] Ben-Ner, A., McCall, B. P., Stephane, M. & Wang, H. (2009). Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior & Organization*, 72(1), 153–170.

[109] Linos, K., Jakli, L. & Carlson, M. (2021). Fundraising for stigmatized groups: A text message donation experiment. *American Political Science Review*, 115(1), 14–30.

[110] Winter, N. & Berinsky, A. J. (1999). What's your temperature? Thermometer ratings and political analysis. In *Annual Meeting of the American Political Science Association, Atlanta, GA*.

[111] Enos, R. D. & Gidron, N. (2018). Exclusion and cooperation in diverse societies: Experimental evidence from Israel. *American Political Science Review*, 112(4), 742–757.

[112] Weisz, E., Ong, D. C., Carlson, R. W. & Zaki, J. (2021). Building empathy through motivation-based interventions. *Emotion*, 21(5), 990.

[113] Iyengar, S. & Krupenkin, M. (2018). Partisanship as social identity: Implications for the study of party polarization. In *The Forum* (Vol. 16, No. 1, pp. 23–45). De Gruyter.

[114] West, E. A. & Iyengar, S. (2022). Partisanship as a social identity: Implications for polarization. *Political Behavior*, 44(2), 807–838.

[115] Allamong, M. B. & Peterson, D. A. M. (2021). Screw those guys: Polarization, empathy, and attitudes about out-partisans. *Political Psychology*, 42(3), 349–369.

[116] Clifford, S., Kirkland, J. H. & Simas, E. N. (2019). How dispositional empathy influences political ambition. *The Journal of Politics*, 81(3), 1043–1056.

[117] Levitsky, S. & Ziblatt, D. (2019). *How democracies die*. Crown.

[118] Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C. & Bigham, J. P. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, Montreal QC, Canada, 2018), 1–14. https://doi.org/10.1145/3173574.3174023.

[119] Bashri, M. F. A. & Kusumaningrum, R. Sentiment analysis using Latent Dirichlet Allocation and topic polarity wordcloud visualization. In *2017 5th International Conference on Information and Communication Technology (ICoIC7)* (IEEE, 2017), 1–5.

[120] Mustaqim, T., Umam, K. & Muslim, M. A. Twitter text mining for sentiment analysis on government's response to forest fires with VADER lexicon polarity detection and k-nearest neighbor algorithm. In *Journal of Physics: Conference Series* vol. 1567, no. 3 (IOP Publishing, 2020), 032024.

[121] Grimmer, J., Roberts, M. E. & Stewart, B. M. *Text as data: A new framework for machine learning and the social sciences* (Princeton University Press, 2022).

[122] Martinez, J. E. Facecraft: race reification in psychological research with faces. *Perspectives on Psychological Science* (SAGE Publications, Los Angeles, CA, 2022), 17456916231194953.

[123] Ma, D. S., Correll, J. & Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods* vol. 47, no. 4 (Springer, 2015), 1122–1135.

[124] ANES. The American National Election Studies (ANES) 2016 Time Series Study (Datafile). (University of Michigan and Stanford University, Ann Arbor, MI and Stanford, CA, 2017).

[125] Andreychik, M. R. Feeling your joy helps me to bear feeling your pain: Examining associations between empathy for others' positive versus negative emotions and burnout. *Personality and Individual Differences* vol. 137 (Elsevier, 2019), 147–156.

[126] Buchanan, L., Bui, Q. & Patel, J. K. Black Lives Matter may be the largest movement in US history. *The New York Times* vol. 3 (2020).

[127] Francis, M. M. & Wright-Rigueur, L. Black Lives Matter in historical perspective. *Annual Review of Law and Social Science* vol. 17 (Annual Reviews, 2021), 441–458.

[128] Minta, M. D. Diversity and Minority Interest Group Advocacy in Congress. *Political Research Quarterly* vol. 73, no. 1 (SAGE Publications, Los Angeles, CA, 2020), 208–220.

[129] UnidosUS. UnidosUS - GuideStar Profile. Available at: https://www.guidestar.org/profile/86-0212873 (Accessed: 2023-01-01).

[130] Hart, S. G. & Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology* vol. 52 (Elsevier, 1988), 139–183.

[131] Davis, M. H. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology* vol. 44, no. 1 (American Psychological Association, 1983), 113–126.

[132] Neuliep, J. W. & McCroskey, J. C. The development of a US and generalized ethnocentrism scale. *Communication Research Reports* vol. 14, no. 4 (Taylor & Francis, 1997), 385–398.

[133] Buhrmester, M., Kwang, T. & Gosling, S. D. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? (American Psychological Association, 2016).

[134] Guess, A. M. & Munger, K. Digital literacy and online political behavior. *Political Science Research and Methods* vol. 11, no. 1 (Cambridge University Press, 2023), 110–128.

[135] Coppock, A. & McClellan, O. A. Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics* vol. 6, no. 1 (2019), 205316801882217. https://doi.org/10.1177/2053168018822174.

[136] Lovett, M., Bajaba, S., Lovett, M. & Simmering, M. J. Data quality from crowdsourced surveys: A mixed method inquiry into perceptions of Amazon's Mechanical Turk Masters. *Applied Psychology* vol. 67, no. 2 (Wiley Online Library, 2018), 339–366.

[137] Berry, C., Kees, J. & Burton, S. Drivers of data quality in advertising research: differences across MTurk and professional panel samples. *Journal of Advertising* vol. 51, no. 4 (Taylor & Francis, 2022), 515–529.

[138] Enns, P. K. Do you know where your survey data come from? *3Streams* (2022). Available at: https://medium.com/3streams/surveys-3ec95995dde2 (Accessed: 2023-02-08).

[139] Lo, A., Renshon, J. & Bassan-Nygate, L. A Practical Guide to Dealing with Attrition in Political Science Experiments. *Journal of Experimental Political Science* (2023), 1–15.

[140] Lin, W. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. (2013).

[141] Berinsky, A. J., Margolis, M. F., Sances, M. W. & Warshaw, C. Using screeners to measure respondent attention on self-administered surveys: Which items and how many? *Political Science Research and Methods* vol. 9, no. 2 (Cambridge University Press, 2019), 1–8.

[142] Manski, C. Nonparametric bounds on treatment effects. *American Economic Review* vol. 80, no. 2 (1990), 319–323.

[143] Göritz, A. S. Determinants of the starting rate and the completion rate in online panel studies. In *Online Panel Research: A Data Quality Perspective* (Wiley Online Library, 2014), 154–170.

[144] Kass, R. E. & Raftery, A. E. Bayes factors. *Journal of the American Statistical Association* vol. 90, no. 430 (Taylor & Francis, 1995), 773–795. https://doi.org/10.1080/01621459.1995.10476572. (Refinement of Bayes factor interpretation scales.)