







People believe political opponents accept blatant moral wrongs, fueling partisan divides

Curtis Puryear ^{a,*}, Emily Kubin ^{b,c}, Chelsea Schein ^d, Yochanan E. Bigman ^e, Pierce Ekstrom ^f and Kurt Gray ^b

^aDepartment of Management and Organizations, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA

^bDepartment of Psychology & Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^cDepartment of Psychology, University of Kaiserslautern-Landau (RPTU), Landau 67663, Germany

^dDepartment of Legal Studies and Business Ethics, The Wharton School of Business, University of Pennsylvania, Philadelphia, PA 19104, USA

^eHebrew University Business School, The Hebrew University of Jerusalem, Jerusalem 9190500, Israel

^fDepartment of Political Science, University of Nebraska Lincoln, Lincoln, NE 68588, USA

*To whom correspondence should be addressed: Email: cwpuryear12@gmail.com

Edited By: James Druckman

Abstract

Efforts to bridge political divides often focus on navigating complex and divisive issues, but eight studies reveal that we should also focus on a more basic misperception: that political opponents are willing to accept basic moral wrongs. In the United States, Democrats, and Republicans overestimate the number of political outgroup members who approve of blatant immorality (e.g. child pornography, embezzlement). This “basic morality bias” is tied to political dehumanization and is revealed by multiple methods, including natural language analyses from a large social media corpus and a survey with a representative sample of Americans. Importantly, the basic morality bias can be corrected with a brief, scalable intervention. Providing information that just one political opponent condemns blatant wrongs increases willingness to work with political opponents and substantially decreases political dehumanization.

Keywords: polarization, politics, morality, perception, false polarization

Significance Statement

The United States is witnessing historic levels of political hostility and gridlock. This animosity is partly grounded in misperceptions of opponents’ political beliefs, but we find many Americans overestimate political opponents’ willingness to accept even the most basic moral wrongs. These findings suggest individuals and practitioners working to foster cross-partisan interaction might first correct this basic morality bias. Specifically, we show that learning a single opponent condemns basic moral wrongs increases behavioral engagement with political opponents and decreases dehumanization of the entire political outgroup.

People believe political opponents accept blatant moral wrongs, fueling partisan divides

Political animosity in the United States has grown steadily over the past 40 years. Americans report hating the opposing party more than they love their own party (1), and growing partisan animosity is associated with rising support for political violence (2). Research shows that political animosity is tied to misperceptions: Democrats and Republicans both believe their opponents are more extreme (3), harbor more prejudice (4), and conform more closely to demographic stereotypes than they actually do (5). For example, Democrats overestimate how much they disagree with Republicans on tax policy and vastly overestimate how many Republicans earn over \$250,000 annually (by 19 times) (5). We propose beliefs about political opponents encompass an even more fundamental misperception—the basic morality bias.

The nature of the basic morality bias

The basic morality bias is the exaggerated perception that outgroup members lack basic moral values—that they accept basic moral wrongs. We use the word “basic” because our focus is not on beliefs about political issues and nuanced moral dilemmas, but on widely agreed-upon moral wrongs in society, like theft and animal abuse. Existing work has assessed how misperceptions about specific political beliefs (e.g. believing opponents hold more extreme beliefs about immigration policy than they actually do) can fuel partisan conflict (6–8). We argue that partisans may also (mistakenly) doubt opponents’ basic moral sense. To be clear, we are not suggesting Democrats and Republicans think one another are completely deprived of all moral capacities. But we do think partisans overestimate how many outgroup members are willing to accept basic moral wrongs. This misperception may have important consequences for political interactions.

Competing Interest: The authors declare no competing interest.

Received: January 9, 2024. **Accepted:** May 8, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Whether we believe others are moral is essential to how we understand their humanity (9), carrying greater weight in our treatment of others than their competence and warmth (10). Research shows that perceiving greater distance between our moral values and the values of outgroups leads us to view them as less human (e.g. lacking rationality and self-restraint) (11, 12). If partisans exaggerate how much their opponents are willing to accept basic moral wrongs, then this may help explain high levels of political dehumanization (13, 14).

Believing that political adversaries are—compared with our in-group—relatively lacking in basic moral values may also undermine the function of political institutions. Institutions require open discussion and compromise (15), and people are unwilling to engage or associate with groups that seem immoral (16, 17). Underestimating the basic moral values of opponents could exacerbate politicians' reluctance to collaborate across party lines (18), increasing political gridlock. The basic morality bias may also drive partisans to isolate themselves from political outgroups in the workplace and in their communities (1). This ideological self-segregation may seem rational to those who think opponents believe acts like theft and abuse are acceptable (19, 20).

Correcting the basic morality bias

We argue partisans believe many of their opponents are willing to accept basic moral wrongs, as suggested by conservatives who call liberals “baby murderers” and liberals who call conservatives “deplorable bigots” (21, 22). At the same time, partisans may be open to reminders of their opponents' basic morality.

People's political perceptions can be influenced both positively and negatively. Constant reminders of our opponent's depravity on social media can reinforce the idea that the other side is immoral (3, 23), but corrective information (e.g. information about what opponents truly believe) also affects political perceptions and can bridge divides (7, 8). The basic morality bias may be similarly corrected: reminding partisans that people on the other side respect basic moral values may reduce dehumanization and increase their willingness to engage with opponents.

Highlighting similarities between two people or groups can generally improve interactions (24), and moral traits are especially influential in social evaluations (10). Highlighting that someone

on the other side—like you—also condemns heinous acts may help prevent the damage the basic morality bias would otherwise wreak upon political interactions. Many bridge-building initiatives may overlook this simple correction strategy because they may not appreciate the existence of basic morality bias among partisans.

Research overview

We test for the existence of a basic morality bias, measure its impact upon dehumanization and willingness to engage with political opponents, and test a solution for correcting it. We first explore the basic morality bias in cross-partisan language on social media (Study 1). We then assess the basic morality bias with self-report measures in samples representative of gender, ethnicity, and age in the United States (Studies 2–3) and test whether it predicts dehumanization and willingness to engage with opponents. Next, we test whether a corrective intervention—learning that a single political opponent condemns basic moral wrongs—reduces dehumanization, increases willingness to collaborate with opponents, and encourages engagement with a bipartisan organization (Studies 4–6). We compare our intervention to broader reminders of similarity (i.e. shared activities; Study 7) and test whether corrective information about just one political opponent also reduces dehumanization for the entire political outgroup (Study 8). Table 1 contains an overview of study samples, designs, and key findings.

Documenting the basic morality bias

We first explored real political conversations on Twitter (currently named X) for evidence of the basic morality bias. Our aim was to assess the prevalence of explicit denials of basic morality in public political interactions. To accomplish this, we compiled a list of “basic morality bias words,” which referenced blatantly immoral acts or traits. We then examined how frequently these words appeared in political discussions.

We collected every tweet posted by a random sample of 5,806 active partisans (who passed thresholds for activity and ideology, see Method) on Twitter from January 2013 to November 2021 (Study 1; $N = 5,854,868$ tweets). We first constructed a list of

Table 1. Summary of study methods and results.

Study	Demographics	Design	Key Finding
Study 1	$N = 5.8$ M tweets from 5.8 K partisans	Archival Correlational	Basic Morality Bias words are especially prevalent in political tweets; this has not always been the case
Study 2	$N = 346$; $M_{\text{age}} = 37.23$, 175 women, 167 men, 4 other gender	Correlational Online Survey	Partisans overestimate opponents' approval of basic morality wrongs, which predicts more <i>dehumanization</i> and less <i>willingness to engage</i>
Study 3	$N = 700$; $M_{\text{age}} = 46.21$, 353 women, 327 men, 4 other gender, 16 no response	Correlational Online Survey	Replication of Study 2 in a sample representative of age, gender, and ethnicity in the United States; included an accuracy incentive.
Study 4	$N = 202$; $M_{\text{age}} = 25.96$, 93 women, 107 men, 2 no response	In-person Survey Experiment	Correcting the basic morality bias decreased <i>dehumanization</i> and increased <i>willingness to engage</i>
Study 5	$N = 174$; $M_{\text{age}} = 33.86$, 65 women, 109 men)	Online Survey Experiment	Correcting the basic morality bias reduced <i>opting out of collaborating with a political opponent</i>
Study 6	$N = 810$; $M_{\text{age}} = 46.70$, 389 women, 416 men, 5 other gender	Online Survey Experiment	Correcting the basic morality bias erased the tendency to <i>dehumanize</i> an opponent more than an ally
Study 7	$N = 600$; $M_{\text{age}} = 35.79$, 288 women, 311 men, 1 no response	Online Survey Experiment	Correcting the basic morality bias decreased <i>dehumanization</i> and increased <i>willingness to engage</i> more than information about shared hobbies
Study 8	$N = 326$; $M_{\text{age}} = 39.28$, 149 women, 173 men, 4 other gender	Online Survey Experiment	Correcting the basic morality for one member of the opposing party also decreased <i>dehumanization for the entire political outgroup</i> on average

Study 6 also recruited a sample via Qualtrics Panels that was representative of gender, income, education, age, region, and ethnicity in the United States. All participants were required to be from the United States and participants recruited via Mechanical Turk were required to have a 95% approval rate.

words, which we believed reflected a denial of basic moral values (e.g. “rapist,” “pedophile,” “felon,” “thief,” “sociopath,” and “murderer”). We then expanded this list with closely related words as discovered by pretrained word embeddings (e.g. “molester,” “homicidal,” and “psychopath”). To identify tweets about political targets versus nonpolitical targets, we constructed dictionaries of partisan identity words (e.g. “liberal” and “Republican”), the names and Twitter handles of political elites during this time period (e.g. presidents and Congress members), and celebrities (i.e. posts by the 1,000 most followed artists, athletes, socialites, and actors—excluding all politicians and pundits; see [Supplemental Materials](#) for full details).

We next looked at the proportion of tweets containing basic morality bias words. Consistent with the idea that the basic morality bias is especially prevalent in politics, basic morality bias words were more prevalent in discussions about political targets than nonpolitical targets (see Figure 1a).^a However, this has not always been the case, as basic morality bias words were no more prevalent in discussions of political targets than nonpolitical discussions in 2013. Looking at the raw volume of tweets containing basic morality bias words (see Figure 1b), the number of times partisans used basic morality bias words when discussing political targets increased from a few dozen in 2013 to more than 500 every year from 2016 onward. This suggests that basic morality bias words are not inherently more prevalent in discussions of political targets, but they have become relatively more prevalent in recent years. This is consistent with the idea that denials of basic morality may be fueled by the current political climate.^b

Across a few thousand accounts, we found hundreds of yearly tweets indicative of the basic morality bias. Although the proportion of political tweets containing basic morality bias words was modest, these tweets reflect extreme attributions (e.g. behaviors like pedophilia and committing felonies), often with little or no basis in fact. Even a handful of such tweets in a year may suggest that partisans are willing to deny the basic morality of others. Importantly, we focused only on words that obviously reflected the denial of basic morality. The word count approach used in this study could not capture the many nuances in cross-partisan language. Moreover, social media analyses cannot tell us what users privately believe, or the context that gives rise to those beliefs—a gap we address in the following studies.

Study 2 sampled 240 Democrats and 106 Republicans from Mechanical Turk (all MTurk samples were collected via CloudResearch). Participants were asked about the acceptability of 7 morally unambiguous issues (e.g. child pornography and homicide) and then indicated how they thought the opposing party would evaluate each issue (“acceptable” or “immoral”). We also assessed explicit dehumanization for the opposing party using the ascent of man scale (25), trait dehumanization (e.g. “the average Democrat/Republican is rational and logical” and “...Lacking in self-restraint, like an animal”),^c and willingness to engage with the other side (e.g. “how willing would you be to have a discussion about politics with an average Republican?”).

Across all issues, liberals and conservatives overestimated opponents’ approval of unambiguously immoral behavior. For example, while 8.75% of Democrats, and 11.32% of Republicans in

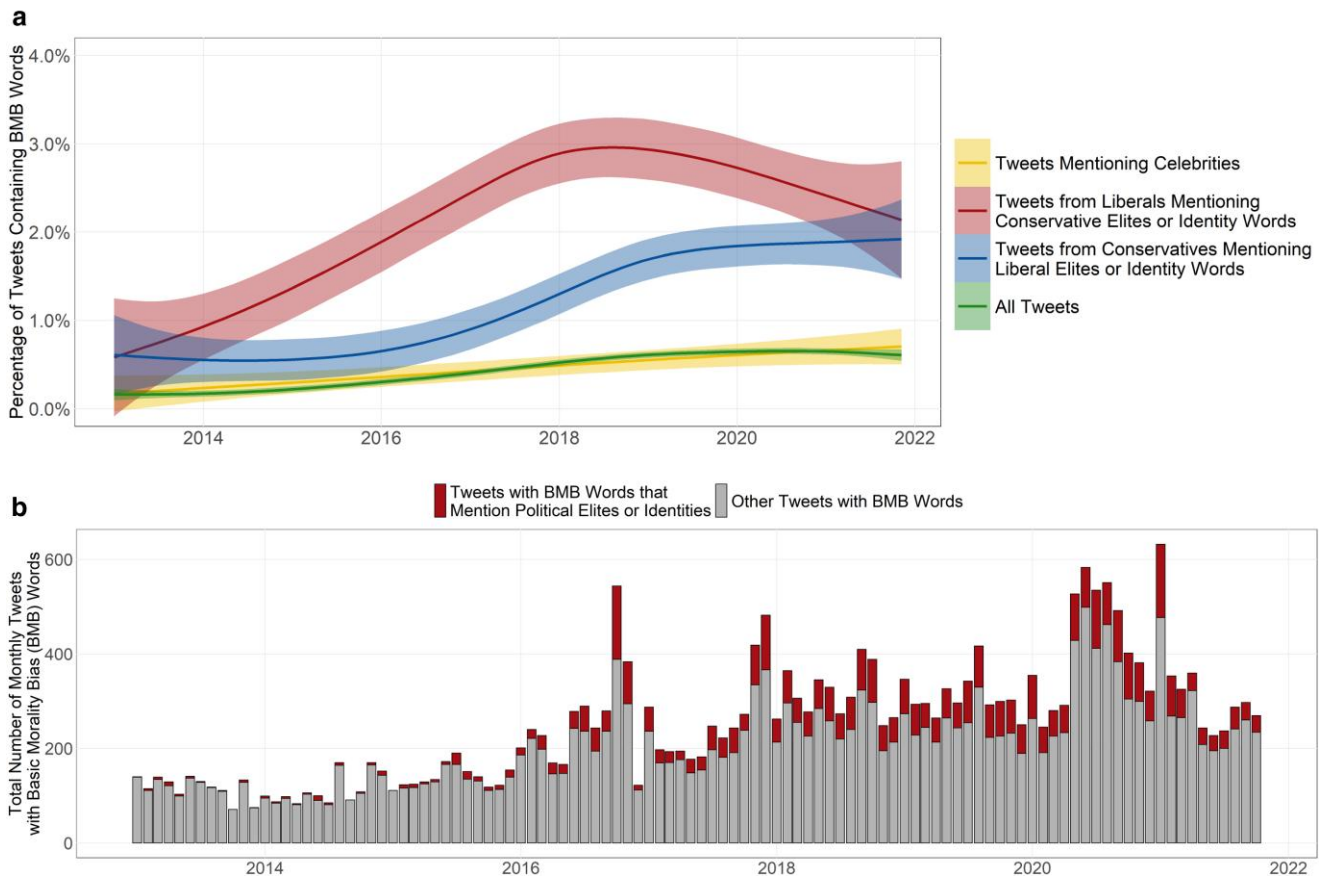


Fig. 1. Increases in basic morality words within political tweets. Trend lines in Figure 1a were drawn using GAM smoothing with 95% CIs. Data presented in Figure 1a and b are aggregated by week and month, respectively.

our sample labeled cheating on a spouse as acceptable, 27.36% of Republicans (and 37.91% of Democrats) reported that the average opponent viewed this behavior as acceptable. Participants expected political ingroup members to label 6.62 out of 7 issues, $SD = 0.96$, as immoral and the opposing party to only label 5.28, $SD = 2.03$, issues as immoral, $t(345) = 13.16$, $P < 0.001$, $d = 0.79$. The number of issues participants personally rated as immoral, $M = 6.60$, $SD = 1.01$, was nearly identical to the number they predicted for the average ingroup member.

We then examined the relationship between the basic morality bias, willingness to engage, and dehumanization. We operationalized the basic morality bias as the number of basic moral wrongs participants expected their outgroup to label “acceptable.” To specifically measure beliefs about outgroup members, we controlled for how much participants’ believed ingroup members rated basic moral wrongs as acceptable. Participants who expected the outgroup to label fewer transgressions as immoral expressed more trait dehumanization, $b = -0.27$, $t(343) = -8.98$, $P < 0.001$, more explicit dehumanization, $b = -4.24$, $t(343) = -5.74$, $P < 0.001$, and less willingness to engage, $b = 0.12$, $t(343) = 4.22$, $P < 0.001$. This study revealed that there is a robust difference between the perceived acceptance of basic moral wrongs among the ingroup and the outgroup, providing evidence for the basic morality bias. As anticipated, partisans did not believe most outgroup members approve of basic moral wrongs. But the extent to which partisans exhibited the basic morality bias predicted dehumanization and willingness to engage.

We next tested whether accuracy cues corrected the basic morality bias. Though partisans may often think opponents accept basic moral wrongs, research suggests they do not think their opponents are completely devoid of moral capacities (26, 27). One may also wonder how earnestly partisans believe their reported misperceptions, as partisans sometimes report beliefs to express their political identities (28). Thus, perhaps even a small accuracy incentive can reduce the basic morality bias.

In Study 3, we collected a representative sample of age, gender, and ethnicity in the United States ($N = 708$) and paid half the participants a \$0.50 bonus if their estimates were within 5% of the true average values. To ensure this sample was adequately powered to detect interaction effects, we conducted a sensitivity analyses, which showed that Study 3 had 80% power to detect effects of $R^2 = 0.003$. We asked participants to estimate the exact percentage of the political outgroup and ingroup that approved of each behavior (rather than indicating whether the average Democrat or Republican approved of each on a binary scale as in Study 2; see Figure 2).^d A 2(ingroup vs. outgroup rating) \times 2(incentive or no incentive) mixed ANOVA revealed a significant main effect of group; participants estimated a higher percentage of opponents labeled unambiguous moral issues as acceptable, $M = 22.77\%$, $SD = 21.74$, compared to their estimates of allies, $M = 13.18\%$, $SD = 17.83$, $F(1, 696) = 81.11$, $P < 0.001$, $\eta_p^2 = 0.10$. Participants paid for accuracy, $M = 16.97\%$, $SD = 19.40$, did not give significantly lower estimates than those not paid for accuracy, $M = 19.02\%$, $SD = 21.43$; $F(1, 696) = 2.23$, $P = 0.14$, $\eta_p^2 = 0.003$, nor were their estimates any more charitable toward opponents, test of interaction: $F(1, 696) = 2.69$, $P = 0.10$, $\eta_p^2 = 0.004$. The simple effect of the accuracy incentive within opponent estimates revealed a small, but nonsignificant effect, $b = -3.11$, $t(696) = -1.89$, $P = 0.06$. Once again, the basic morality bias predicted trait dehumanization, $b = 0.03$, $t(691) = 12.96$, $P < 0.001$, explicit dehumanization $b = 0.33$, $t(691) = 6.20$, $P < 0.001$, and willingness to engage $b = -0.02$, $t(690) = -7.28$, $P < 0.001$.

The results of Studies 1–3 document the basic morality bias and its potential impacts. Social media analyses suggested partisans on Twitter are increasingly willing to use words that deny basic morality. Surveys revealed that partisans overestimate approval of basic moral wrongs among opponents more than allies.^e This intergroup bias was resistant to accuracy incentives and predicted dehumanization and willingness to engage. At the same time, partisans reported that most of their opponents do condemn basic

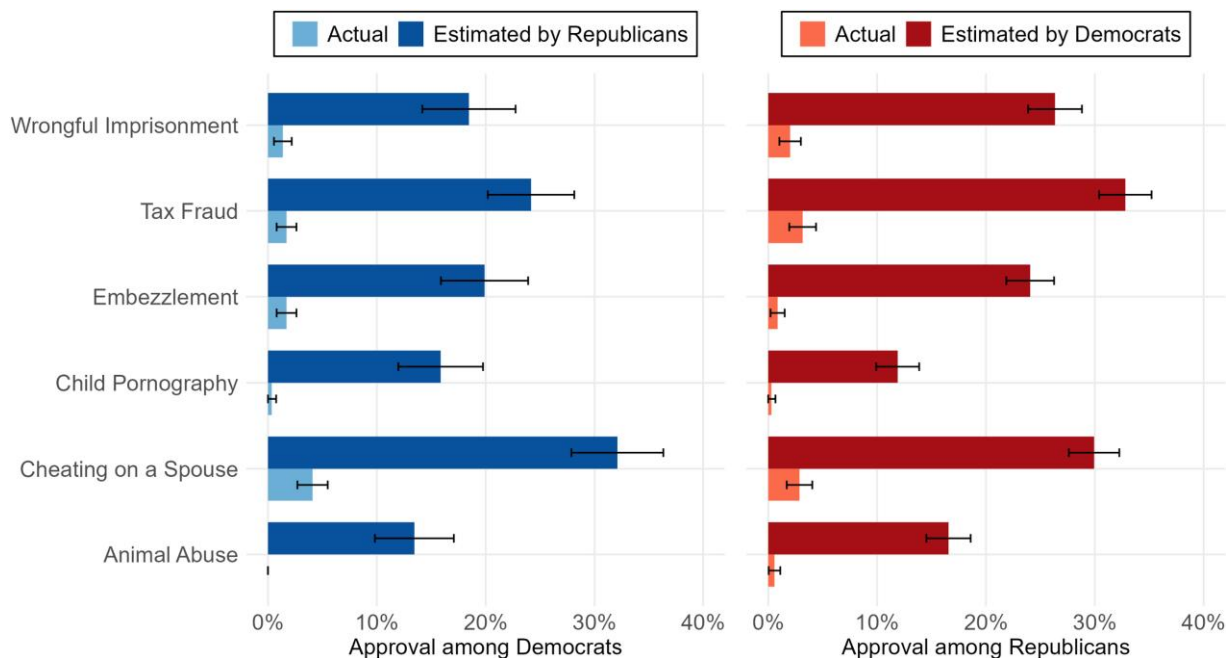


Fig. 2. Actual and estimated percent approval of basic moral wrongs. Estimated approval is data from Study 3. Actual approval is from a pilot Study with a sample of 641 Democrats and Republicans representative of age, gender, and ethnicity in the United States. Error bars represent 95% CIs.

moral wrongs, which provides hope that reminders of shared morality may correct the basic morality bias.

Correcting the basic morality bias

Republicans and Democrats exhibit the basic morality bias in their perceptions of outgroup partisans, but can simple, direct corrections to the basic morality bias undo its effects on dehumanization and willingness to engage? People generally think others are good “deep down” (26); so, perhaps simple reminders of shared, basic morality will counter the effects of the basic morality bias. On the other hand, if people generally think others are good “deep down,” then learning someone condemns obvious moral wrongs may seem uninformative.

A pilot study (Pilot 1; $N = 99$) revealed 80% of people believe explicit corrections of the basic morality bias would not change social evaluations. Further, 70.7% of participants indicated learning someone condemned six basic moral wrongs would not alter their views of them. Informational corrections of the basic morality bias seemed nondiagnostic to the participants in our pilot study, suggesting that reminding others of our basic morality is not obviously helpful. However, in cross-partisan interactions, people’s assumptions about their opponents may be so pessimistic that this information is useful.

In Study 4, we tested whether telling participants an opponent opposes obvious moral wrongs can reduce dehumanization, increase willingness to engage, and increase actual cross-partisan engagement behavior (i.e. choosing to learn about a cross-partisan organization).

Participants ($N = 202$) in Study 4 comprised students and community members who were recruited on and near a university in the southeastern United States. Participants evaluated an opponent who held strong, opposing political views (e.g. “Jeff is strongly liberal”). After learning that Jeff was an ideological opponent, half of participants read that Jeff opposed the same six morally unambiguous issues used in Studies 2–3 (e.g. he thinks child pornography is immoral). The control condition did not learn this information. Next, participants rated Jeff on trait dehumanization and reported their willingness to learn more about a cross-partisan organization (*Bridge the Divide*). We also included a behavioral measure of cross-partisan engagement: whether participants signed up for information from *Bridge the Divide*. Learning Jeff opposed basic wrongs reduced trait dehumanization, bias correction: $M = 3.06$, $SD = 1.10$; control: $M = 3.83$, $SD = 0.81$, $t(200) = -5.65$, $P < 0.001$, $d = -0.80$, increased willingness to engage with a bipartisan organization on the self-report measure, bias correction: $M = 5.00$, $SD = 0.82$; control: $M = 4.66$, $SD = 0.91$, $t(200) = 2.77$, $P = 0.006$, $d = 0.39$, and marginally so on the behavioral measure (57% signed in the bias correction condition; 45% in control, $\chi^2 = 2.86$, $P = 0.09$).

Study 4 suggested correcting the basic morality bias increases humanization and willingness to engage with opponents. Results for our behavioral measure of engagement—which tested the potential benefits of our intervention beyond its effects upon attitudes towards a single outgroup member—was in the predicted direction but was nonsignificant. The effect upon this behavioral measure was also nonsignificant in a similar study we conducted (see [Supplementary Material Study 1](#)), which used the same design as Study 4, except participants in the control condition did not also rate the acceptability of the six basic moral wrongs. Notably, our behavioral measure of engagement used a binary outcome in both studies, which often suffer from reduced statistical power relative to the other measurement scales on which we based our power analyses (29). An exploratory analysis

combining data from Study 4 and [Supplementary Material Study 1](#) found a significant effect on our behavioral measures (54% signed in the bias correction condition; 43% signed in control, $\chi^2 = 4.73$, $P = 0.03$). Still, the size of this effect was modest and should be interpreted with caution given this analysis was exploratory.

Addressing the basic morality bias helps teams work together

After the 2016 election, nearly a quarter of participants surveyed by the American Psychological Association (30) reported avoiding a colleague at work due to their political orientation. We examined whether corrections to the basic morality bias could overcome reluctance to work in teams with political opponents.

In Study 5, participants ($N = 174$) provided their own rating of the same six unambiguous moral issues from previous studies and then were told they would participate in a puzzle-solving competition with a teammate. They saw a profile of one potential teammate—a political opponent—and half of participants also read their teammate opposed each of the six moral wrongs. Participants learned that in previous competitions, people completed 7 puzzles on average. We then asked how many puzzles participants thought their teammate could complete (a measure of perceived competence) and whether they would like to switch partners (a behavioral measure of engagement). Participants who learned their political opponent opposed the unambiguous wrongs predicted they would complete more puzzles, bias correction: $M = 7.14$, $SD = 1.03$; control: $M = 6.76$, $SD = 1.34$, $t(172) = 2.12$, $P = 0.04$, $d = 0.32$, and were half as likely to switch teammates, bias correction: 11.9% switched partners; control: 25.6%, $\chi^2 = 5.27$, $P = 0.02$.

Study 6 ($N = 810$) sought to replicate the results of Study 5 in a large sample representative of gender, income, education, age, region, and ethnicity in the United States. We also included a third condition where participants’ partners agreed with them about politics. This condition enabled us to measure how much more participants dehumanized and avoided political opponents (versus allies) and to test whether our bias correction strategy fully erased preferences for political allies. Study 6 also measured cognitive trust (“I could rely on this person to follow through on commitments”) (31), an important predictor of workplace cooperation (32).

When the partner was a political opponent, learning they opposed unambiguous wrongs reduced dehumanization, bias correction: $M = 2.60$, $SD = 1.15$; control: $M = 3.75$, $SD = 1.18$, $t(807) = -11.66$, $P < 0.001$, $d = -0.98$. In fact, comparing the bias correction condition to the political ally condition, $M = 2.66$, $SD = 0.98$, revealed that the bias correction eliminated the dehumanization gap between political opponents and allies, $t(807) = -0.51$, $P = 0.61$, $d = -0.05$ (see [Figure 3](#)).

Corrections to the basic morality bias also increased cognitive trust, $M = 5.09$, $SD = 1.29$, and perceived competency, $M = 5.74$, $SD = 1.99$, compared to control, trust: $M = 4.47$, $SD = 1.27$, $t(807) = 5.74$, $P < 0.001$, $d = 0.49$; competency: $M = 5.10$, $SD = 1.77$, $t(807) = 3.82$, $P < 0.001$, $d = 0.34$. Again, the bias correction eliminated the gap between ratings of opponents and allies, as the bias correction condition did not differ from the political ally condition, trust: $M = 5.16$, $SD = 1.12$, $t(807) = 0.65$, $P = 0.52$, $d = -0.06$; puzzles solved: $M = 5.62$, $SD = 1.84$, $t(807) = 0.78$, $P = 0.44$, $d = 0.07$. Lastly, fewer participants switched partners from a political opponent in the bias-correction condition (9.8%) versus control (16.7%), $\chi^2 = 5.06$, $P = 0.02$. In the political ally condition, 5.8% of participants chose to switch partners (contrast with bias-

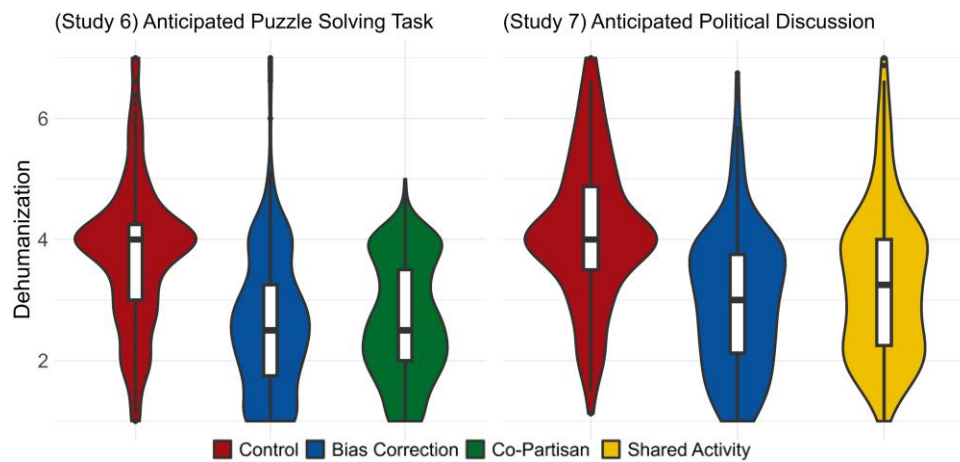


Fig. 3. Basic morality bias corrections reduce dehumanization of individual political opponents.

correction condition: $\chi^2 = 3.09$, $P = 0.08$). These results suggest that our bias correction strategy may also improve partisans' willingness to work together in teams with political opponents.

Correcting the basic morality bias versus other reminders of similarity

Telling partisans that an opponent condemns basic moral wrongs increases humanization and willingness to engage, but any information emphasizing similarity between political opponents may have some effect (24). Study 7 compared correcting the basic morality bias (i.e. highlighting moral similarity) to learning that an opponent enjoyed similar activities (i.e. highlighting everyday similarities)—testing whether the former was especially effective at increasing willingness to chat with a political opponent (i.e. at the start of the study, participants were told they would have a live discussion with another MTurk worker). In addition to rating the immorality of six basic moral wrongs, participants also rated whether they enjoyed six popular activities (e.g. “talking with friends” and “listening to music”). They then either learned that a political opponent condemned the basic moral wrongs, learned that they enjoyed the same activities, or received no information about the opponent. We then measured dehumanization and how positive participants expected the chat to be with eight items (e.g. “I expect the chat to be...,” “pleasant,” and “frustrating”).

Overall, corrections to the basic morality bias significantly reduced dehumanization, $M = 2.99$, $SD = 1.15$, relative to both the control condition, $M = 4.13$, $SD = 1.19$, $t(597) = -9.36$, $P < 0.001$, $d = -0.98$, and the shared activity condition, $M = 3.30$, $SD = 1.24$, $t(597) = -2.70$, $P = 0.007$, $d = -0.26$. Correcting the basic morality bias also significantly increased positive expectations about the upcoming chat, $M = 4.89$, $SD = 1.51$, compared to the shared activity, $M = 4.54$, $SD = 1.43$, $t(597) = 2.53$, $P = 0.02$, $d = 0.24$, and control conditions, $M = 4.09$, $SD = 1.40$, $t(597) = 5.53$, $P < 0.001$, $d = 0.57$. The shared activity condition also increased positive expectations, $t(597) = 3.13$, $P = 0.002$, $d = 0.32$, and decreased dehumanization, $t(597) = -0.677$, $P < 0.001$, $d = -0.68$, compared to the control condition.^f

Any reminders of similarity can help partisans see each other's humanity, but explicit corrections to the basic morality bias are more effective. Correcting the basic morality bias probably does increase perceived similarity. Since almost everyone rejects basic moral wrongs, learning someone rejects basic moral wrongs will almost always signal similarity. However, correcting the basic

morality bias should specifically increase perceived moral similarity, which is especially impactful in social evaluations (10). This explains why correcting the basic morality bias may be especially effective at combating political dehumanization. Moreover, it may not always be easy for liberals and conservatives to find similarities, given their often different personalities and interests (33). But since nearly everyone agrees on basic moral values, reminders of basic moral similarity should be especially feasible.

Is it possible to correct the basic morality bias for the entire outgroup?

Decreasing dehumanization for one opponent is a worthy goal, but can reminders of one persons' moral similarity also change how partisans view the opposing party overall? Study 8 ($N = 326$) tested whether correcting the basic morality bias for one political opponent also decreased the basic morality bias and dehumanization for the entire political outgroup. We also examined whether correcting the basic morality bias for just one moral issue—instead of six moral issues—sufficed to decrease dehumanization.

We again told participants they would have a live chat with a political opponent (named “TeresaJ”) either after correcting the basic morality bias or not. As in Study 7, we measured both trait dehumanization of TeresaJ and whether participants wanted to switch conversation partners (a behavioral measure of engagement). To test whether the effects of our basic morality bias correction generalized to the entire political outgroup, we measured how much participants dehumanized and exhibited the basic morality bias towards Democrats and Republicans as a whole. Lastly, to test whether learning that an opponent condemns just one moral wrong is enough to correct the basic morality bias, we included a third condition in which participants saw information about one basic moral wrong rather than six.

Again, correcting the basic morality bias decreased dehumanization towards a single opponent. Compared to control, $M = 4.29$, $SD = 1.38$, participants who learned that an opponent condemned all six, $M = 3.13$, $SD = 1.07$; $t(322) = -6.88$, $P < 0.001$, $d = -0.94$, or just one basic moral wrong, $M = 3.38$, $SD = 1.22$; $t(322) = -5.46$, $P < 0.001$, $d = -0.70$, dehumanized their partner significantly less. The two bias correction conditions did not differ from one another, $t(322) = 1.44$, $P = 0.151$, $d = 0.10$. These results suggested that even just one correction to the basic morality bias makes individual opponents seem more human.

We next tested whether these effects generalized to attitudes towards political outgroups in general and directly examined reductions in the basic morality bias (i.e. measuring the estimated percentage of outgroup members who approve of basic moral wrongs). The six-item bias correction reduced the basic morality bias, $M = 8.52\%$, $SD = 12.12$, $t(319) = -2.94$, $P = 0.003$, $d = -0.45$, but the single item correction did not, $M = 16.29\%$, $SD = 20.36$, $t(317) = 0.360$, $P = 0.718$, $d = 0.04$, compared to control, $M = 15.44\%$, $SD = 18.24$. This suggested that correcting the basic morality bias for the entire outgroup may require multiple pieces of converging information. Examining our behavioral measure of engagement, we again found the six-item correction significantly decreased participants' likelihood of switching conversation partners (Odds ratio = 0.43, 95%CI = [0.24, 0.76], $P = 0.004$) but the single item correction did not (Odds ratio = 0.68, 95%CI = [0.39, 1.17], $P = 0.17$).

Examining how much participants dehumanized their political outgroup, both the six-item bias correction, $M = 3.81$, $SD = 1.19$; $t(319) = -4.09$, $P < 0.001$, $d = -0.57$, and the single item correction, $M = 3.93$, $SD = 1.31$; $t(319) = -3.40$, $P < 0.001$, $d = -0.45$, significantly decreased group-level dehumanization compared to control, $M = 4.52$, $SD = 1.32$. These results suggest that—while a single piece of information may be insufficient to correct the basic morality bias—it may still have some effect upon political dehumanization. Importantly, the condition that successfully reduced the basic morality bias (i.e. the 6-item correction) also increased behavioral engagement with an outgroup member.

General discussion

Partisans in the United States believe some opponents are willing to accept obvious moral wrongs, which may help explain why the United States is burdened by political gridlock, partisan bias (18), and political dehumanization (13). Republicans believe that 15% of Democrats accept child pornography, and Democrats believe over a quarter of Republicans accept wrongful imprisonment. These estimates are significantly higher than partisans' estimates for their own party. Social media analyses also suggest partisans are increasingly likely to express these beliefs in online political interactions. Two correlational studies showed misperceptions of basic moral values predict negative outcomes like dehumanization and opposition to cross-partisan engagement.

Fortunately, the basic morality bias is correctable—at least within individual political interactions. Partisans often hear their opponents are morally depraved (3), but the present work reveals that reminders of moral similarity can counteract these negative messages. Learning that a political opponent condemns obvious moral wrongs increases willingness to engage with bipartisan groups, decreases dehumanization of a political opponent (erasing the gap between how much partisans dehumanize opponents versus allies), and decreases dehumanization of the political outgroup on average.

The present work suggests correcting the basic morality bias is an especially effective approach to combating political dehumanization. Research from contact theory has long demonstrated the potential for intergroup contact to reduce prejudice (34), but many cross-partisan interactions are highly negative (35, 36)—a form of contact which often backfires and increases prejudice (37). Correcting the basic morality bias may provide a promising strategy for fostering more positive intergroup contact and for correcting misperceptions outside of social interactions; however, the power of this correction may not be obvious. Eighty percent of participants in a pilot study indicated corrections to the basic morality bias would not affect their opinion of political opponents, but our

results suggest that these reminders do indeed humanize outgroup members, and even moreso than reminders of general similarity. Since people rarely disagree on basic morality, correcting the basic morality bias should provide a reliable humanizing strategy.

The present findings may raise the question, “Do partisans really believe that their opponents are ok with abuse, theft, and murder?” A large literature on expressive responding (28, 38) suggests partisans sometimes express beliefs in surveys simply to derogate their political opponents—raising concerns that some responses are insincere or artifacts of low-stakes surveys. However, other work suggests that rates of expressive responding are modest (39) and not necessarily insincere, since people often find ways to rationalize the beliefs they express (40, 41). We also employed an accuracy incentive in Study 3—a common strategy for reducing insincere responses—and observed only a small, nonsignificant reduction in the basic morality bias. Of course, this monetary incentive was small (\$0.50), but previous work has also found minimal differences across smaller and larger accuracy incentives (42). Lastly, to the degree that participants' responses were expressive, this does not mean they are unimportant. The same motivations that give rise to expressive responding in surveys (e.g. motivations to praise the ingroup and derogate the outgroup) are often even stronger in real-world interactions (43). Even if partisans' self-reports are expressive to some degree, it is still striking—and relevant to real-world interactions—that they are willing to claim that ~20% of their opponents are accepting of embezzlement and wrongful imprisonment.

We used multiple methods to document the basic morality bias, each of which carries strengths and limitations. Surveys allowed us to collect precise estimates of opponents' approval of basic wrongs, but people often provide inflated estimates of population base rates (44). That said, our social media data showed that a substantial portion of partisans are willing to openly express the basic morality bias, suggesting the phenomenon is more than a quirk of how laypeople guess statistics. Experiments then revealed that correcting the basic morality bias improves cross-partisan attitudes and behavioral engagement. These methods provide converging evidence that the basic morality bias affects cross-partisan behavior and attitudes in the United States.

The present studies also relied upon online samples, which tend to be more politically engaged than the average American (45). People who are more politically active, have stronger partisan identities, or more extreme political attitudes may be more likely to doubt opponents' basic morality. But in a political environment where people are increasingly sorted by party lines (46) and where the most politically active and extreme voices are elevated above others (23), understanding why more extreme, active, or strongly identified partisans dehumanize one another is as important as ever.

Future work should examine how the present approach to correcting misperceptions compares to approaches that provide statistical information about opponents' beliefs. For example, showing partisans the true percentage of political opponents who support political obstruction has been found to reduce false perceptions (7, 8). By contrast, our approach was to provide information about the basic moral values of a single exemplar. This approach also has advantages, as previous work has found that vivid information about a single political opponent can—in certain contexts—foster respect more effectively than statistical information (47). Results from Study 8 suggested that our approach could also reduce misperceptions of the entire outgroup on average. Future work might directly compare these approaches and discover whether certain correction strategies are more effective in some contexts over others.

Research should also examine whether these corrections to the basic morality bias improve cross-partisan attitudes outside of the United States. Since the United States has witnessed larger increases in partisan animosity than other countries (1, 46), the basic morality bias may be less pervasive elsewhere. Though corrections to political misperceptions have proven to be successful across countries (7), we theorize that direct corrections to the basic morality bias will only be effective in environments where this bias is prevalent.

Future work might also explore the longevity of corrections to perceptions of basic morality. While we argue that increasing engagement and humanization in a single interaction is worthwhile, long-term effects upon cross-partisan attitudes could transform political engagement even further. Unfortunately, some recent work has found that the effects of corrections to political misperceptions do not persist over time (8). If single corrections to the basic morality bias subside over time, then generating prolonged willingness to engage with the other side may require repeated reminders that political outgroups oppose basic moral wrongs.

Conclusion

In the United States today, people exaggerate the immorality of the other side not just when it comes to hot-button issues but also when it comes to basic moral values, believing that many on the other side see blatant wrongs as acceptable. Fortunately, correcting these misperceptions is possible and bridges divides more than people expect.

Method

Informed consent was obtained in all studies that were not archival. This research was approved by the institutional review board at the University of North Carolina at Chapel Hill (#20-2521).

Study 1

Data collection

We used the Twitter Academic API to collect every tweet (excluding retweets) from a random sample of 5,806 active partisans January 2013 to November 2021. Users were randomly sampled from a data base of 54 million users whose ideology was estimated in 2018 using a validated method (48). To ensure that users were partisan, we only included users with ideology estimates that were less than -0.5 (for liberals) and greater than 0.5 (for conservatives). Users beyond these thresholds are highly likely to be registered with the expected political party (e.g. liberals as Democrats) (48). We also followed recommendations from Barbera and colleagues (48) for minimizing risks posed by bot accounts. We excluded accounts with fewer than 100 tweets and fewer than 25 followers; we only included Twitter accounts who listed a US state in their Twitter profile; and we excluded accounts with more than 10,000 tweets. The number of unique users posting tweets from 2013 to 2021 ranged from a minimum of 3,897 in 2021 to a maximum of 4,950 in 2016.

Basic morality bias dictionary

We identified keywords for extreme, immoral behavior we theorized to represent the denial of basic moral values. Our list of initial words included “rapist,” “pedophile,” “felon,” “thief,” “sociopath,” and “murderer.” We then used GloVe embeddings that were pre-trained on 2 billion tweets to identify additional words for each

category. Two researchers examined the 50 most semantically similar words to each of our six initial words and recorded closely related words. For example, in the case of pedophile, this helped us identify alternative spellings and variations used on Twitter like “pedo” and “paedophile.” Both researchers then met to resolve discrepancies to create the final dictionary (available on our OSF page).

Political elites and identity dictionary

Our list of political target words consisted of the names and Twitter handles of top political elites, the Twitter handles of all Congress and Cabinet members during the period of data collection, and words representing liberal and conservative identity (e.g. “liberal” and “Republican”). Top political elites comprised all presidents and vice presidents (starting with Bush and Cheney), house and senate leaders, and all presidential primary frontrunners (starting with the 2012 primaries). Liberal and conservative identity words were generated by the researchers, using the same word embedding approach to make sure all common identity labels were included.

Study 2

Participants

To achieve our target sample size of 300 participants, we recruited 497 participants from Mechanical Turk, of which 240 were Democrat, 106 Republican, 133 Independent, and 18 something else. No participants failed either of two attention checks. Because we were specifically interested in perceptions of political opponents, we only included Democrats and Republicans in the data analysis, leaving a total of 346 ($M_{\text{age}} = 37.23$, $SD_{\text{age}} = 11.43$; 175 women, 167 men, 3 nonbinary, 1 other gender; 240 Democrats, 106 Republicans).

Procedure

Basic moral wrongs

Participants saw a list of seven moral issues in random order: tax fraud, watching child pornography, embezzlement, homicide, wrongful imprisonment, animal abuse, and cheating on a spouse. They were then asked to rate the immorality of each item on a binary scale from acceptable to immoral. They then saw the same list of actions twice more, but were asked to select what the average Democrat, and the average Republican believes, in random order. We then summed the number of issues each participant personally labeled immoral and expected ingroups and outgroups to label as immoral.

Dehumanization

All studies measured trait dehumanization using eight items from (9), presented on a seven point scale from 1-Not at All to 7-Extremely. Items were: “Refined and cultured (rev),” “Responsive and warm (rev),” “Rational and logical (rev),” “Mechanical and cold, like a robot,” “Lacking in self-restraint, like an animal,” “Unsophisticated,” “Open-minded, like they can think clearly about things (rev),” and “Superficial, like they have no depth.” The explicit dehumanization scale (25) contains a series of increasingly evolved humans, and asks people to place targets on this ascent of man scale.

Willingness to engage with opponents

Participants then indicated how willing they would be to engage with political opponents on four items: (i) Have a general

discussion with an AVERAGE REPUBLICAN[DEMOCRAT]; (ii) Have a discussion about politics with an AVERAGE REPUBLICAN[DEMOCRAT]; (iii) Attempt to understand an AVERAGE REPUBLICAN'S[DEMOCRAT'S] political viewpoint; (iv) Listen to an AVERAGE REPUBLICAN[DEMOCRAT] describe their political views.

Study 3

Participants

A US sample of 1,038 participants, representative of age, gender, ethnicity, was collected from Prolific. We excluded 34 participants who missed one of our two attention checks. Political Independents were also removed prior to analysis, leaving a final sample of 515 Democrats and 185 Republicans ($M_{\text{age}} = 46.21$, $SD_{\text{age}} = 16.04$; 353 women, 327 men, 4 other gender, 16 no response).

Procedure

Participants were told that we had previously collected a representative sample of 800 Democrats and Republicans and that they would receive a \$0.50 bonus if their estimates for each party were within 5% of the true average. Participants then estimated the percentage of political opponents and allies who consider the same issues used in Study 2 (except homicide) acceptable. They then completed measures for dehumanization and willingness to engage.

Pilot 1

Participants and procedure

Ninety-nine participants ($M_{\text{age}} = 36.2$, $SD_{\text{age}} = 11.35$, 55 women, 43 men, 1 no response) on Mechanical Turk were told that they were participating in a study examining how receiving new information might lead people to update their beliefs about others. They saw six statements saying that a target believes one of the six morally unambiguous wrongs used in Studies 2 and 3 is immoral. They also saw six statements saying a target believes one of the six acts is acceptable (e.g. X believes that tax fraud is immoral/acceptable). They were then asked to rate whether this information would lead them to change their opinion about the other person, by selecting "no, does NOT change my view" or "Yes, changes my view."

Study 4

Participants

Two hundred and two participants ($M_{\text{age}} = 25.96$, $SD_{\text{age}} = 10.56$; 93 women, 107 men, 2 no response, 133 slightly liberal/liberal/very liberal, 44 slightly conservative/conservative/very conservative, 25 neutral) completed the study around a public university in the southeastern United States in exchange for an ice-pop. Study 4 did not include attention checks.

Procedure

Participants were handed an iPad to complete the survey. Participants in the bias correction condition then saw a series of six statements about Jeff's beliefs. Participants read, "Jeff believes that [tax fraud/child pornography/embezzlement/animal abuse/wrongful imprisonment/cheating on a spouse] is immoral. What do you believe?" For each of the six issues, participants then selected either "immoral" or "acceptable." If they chose immoral (as occurred in 88% of the cases), they then saw a follow-up screen

stating "you both AGREE about this moral topic." If they chose acceptable, they proceeded to the next question without the screen saying they and Jeff agreed about the topic. After rating all six issues, participants saw a summary screen, with a report of the number of issues they and Jeff agreed upon. Participants in the control condition also rated the immorality of the same six actions before learning that they would be making judgments about a political opponent. Thus, the only difference between the bias correction and control condition was the information about Jeff's beliefs.

Participants then completed the trait and explicit dehumanization measures and a revised version of the willingness to engage scale: How willing would you be to (i) Attempt to understand Jeff's viewpoints, (ii) Empathize with Jeff, (iii) Have a discussion about politics with Jeff, and (iv) Have a general discussion with Jeff, measured on a six-point scale from Very Unwilling to Very Willing. Participants then saw the logo of Bridge the Divide, and read that, "Bridge the divide is an initiative to bridge the divide between right and left, Republican and Democrat, to engage in conversations that matter." They then selected on a five-point scale how interested they were in getting involved with an organization like Bridge the Divide. After the survey, the research assistant, blind to condition, then took out a clipboard, with the logo of "Bridge the Divide" on its back, and asked (yes or no) if they would like to write their names and emails down to receive more information. All participants saw a petition sheet with four names already signed.

Study 5

Participants

One hundred ninety-nine participants completed the study on Mechanical Turk. We included one attention check, which 25 participants failed. This left a final sample of 174 participants ($M_{\text{age}} = 33.86$, $SD_{\text{age}} = 9.82$; 65 women, 109 men; 89 slightly liberal/liberal/very liberal, 61 slightly conservative/conservative/very conservative, 24 neutral).

Procedure

Participants were told that the researchers were examining how people perform in teams. They were then told that they would be paired with another Mechanical Turk worker to complete puzzles for a 10 cent bonus. As they waited for a match, they entered a screenname for the study, answered a question about their political orientation, (Strongly Liberal/Liberal/Conservative/Strongly Conservative), and rated the immorality of the six basic moral wrongs from previous studies. Following these questions, participants saw a screen that said "matching" for four seconds, followed by a message saying that there were 8 MTurkers waiting for matches online. They then read that after seeing the profile of a potential teammate, they would have an option of switching teammates. After a two second waiting screen, all participants saw a profile of "TylerK." They read "You and TylerK DISAGREED about politics. You said you were ____. TylerK is: Strongly ____." The other teammate was always a political opponent. Participants in the bias correction condition then saw an addition screen stating that "You and TylerK AGREED on ____ out of 6 moral issues." Participants saw their responses, followed by TylerK's responses. TylerK labeled all acts as immoral.

To measure perceived competence, we asked "in previous competitions, the average participant completed 7 puzzles by themselves. How many puzzles do you think TylerK could complete on their own?" Options ranged from 4 to 10 puzzles. We then

asked whether they would like to switch partners (“yes, switch partner” or “no, stay with current partner”). Since our main focus was not on the actual puzzle performance, participants learned that there was an error in matching and that they would automatically earn a bonus after answering a few demographic questions.

Study 6

Participants

Participants were recruited through Qualtrics Panels. We planned to recruit 800 participants that reflect the US adult population on gender, income, education, age, region, and ethnicity. Our recruitment roughly reflected our targeted sample distribution, but participants with less than a high school education were more likely to stop the study once they saw the anagram task, and therefore were harder to recruit. Thus, we combined our recruiting goals for some high school or less, and only high school. We included one attention check, which all participants passed. Our final sample included 810 participants ($M_{\text{age}} = 46.70$, $SD_{\text{age}} = 17.21$; 389 women, 416 men, 5 other gender; 381 liberal/very liberal, 429 conservative/very conservative). Although the conditions were distributed randomly, more participants in the political disagreement conditions (disagreement alone $N = 258$; disagreement and bias correction, $N = 244$) dropped out of the study before completion than participants in the political agreement condition ($N = 308$). This likely led to a more conservative test of our predictions, as participants who likely wanted to avoid engaging with political opponents, or who held more negative views potentially dropped out of the study early.

Procedure

Participants were told that they were going to be paired with another online worker to complete anagrams and the pairs that completed the most puzzles would earn a bonus. As they were presumably being matched with another partner, participants completed a demographics screening page to help meet the quotas for the survey, and then the six unambiguous moral acts items. To increase the realism of this study, participants then completed a practice round of anagrams. All participants saw three anagrams (HAICR, TRUFI, DIWNWO). Participants then briefly saw an unsolvable anagram (UNAGAT) before a stop sign appeared on the screen. They then saw a screen with the number of puzzles solved.

After the practice round, participants were paired with TylerK. They then learned one of three sets of information: In the political agreement condition, they learned that “You and TylerK AGREED about politics. You said you were ____. TylerK is: Strongly ____.” In the political disagreement condition, agreed was replaced with the word “DISAGREED” and participants learned that TylerK was strongly the opposing political party. Finally, in the political disagreement, plus bias correction condition, participants learned that TylerK condemned the six unambiguous wrongs in the same manner as Study 5.

Participants completed the predicted number of puzzles and partner switching questions from Study 5, as well as the trait dehumanization scale from previous studies. To measure cognitive trust, participants were instructed to imagine they were working with TylerK on a team for work. They were then asked to rate their agreement (1-Strongly Disagree to 7-Strongly Agree) with the following items: I could rely on this person to follow through on commitments; I could assume this person’s work was done properly if I needed to use it; I would be comfortable having this person in a critical role on my team; and I would feel uneasy if I needed to depend on this person’s abilities (reversed)—adopted from (31).

Study 7

Participants

We recruited an initial sample of 662 participants from Mechanical Turk, targeting a sample of 600 participants after pre-registered exclusions. We included one attention check and an open response question, which we examined for nonsensical answers. Sixty-two participants either failed the attention check or wrote nonsense in their free responses, leaving a final sample of 600 participants ($M_{\text{age}} = 35.79$, $SD_{\text{age}} = 11.04$; 288 women, 311 men, 1 no response; 339 liberal/very liberal, 261 conservative/very conservative).

Procedure

Participants learned that they would have a chat with another MTurker. As they waited for a match, they entered a screenname for the study, answered a question about their political orientation, (Strongly Liberal/Liberal/Conservative/Strongly Conservative), and then rated the immorality of the six unambiguously harmful acts used in previous studies. They also rated a six-item activity survey (do not enjoy/enjoy: talking with friends, listening to music, reading, watching YouTube, playing games (board games or video), taking walks). A small pilot study suggested that MTurkers generally enjoy these activities. Participants saw a screen that said “matching” for four seconds, followed by a screen indicating that a match has been found. Participants then saw a screen stating that “You and TylerK DISAGREED about politics. You said you were ____. TylerK is: Strongly ____.” The other teammate was always a political opponent.

In the bias correction condition, participants learned that TylerK condemned all six of the moral wrongs. In the activity similarity conditions, participants learned that TylerK found all six of the activities enjoyable. Ninety-one percent of participants found at least five of the activities enjoyable, indicating our manipulation successfully communicated activity similarity. In the control condition, participants did not receive any additional information about TylerK after learning he disagreed with them on politics.

Participants rated their partner on trait dehumanization and then reported their expectations about the chat (I expect the chat to be... pleasant, enjoyable, informative, fun, interesting, frustrating annoying, useless, angering, upsetting) on a seven-point scale from Not at All to Extremely So. Negative items were reversed scored.

Study 7 also included a manipulation of whether the upcoming chat would be about politics or activities. Chat topic did not moderate the degree our basic morality bias correction improved expectations for cross-partisan interactions (see [Supplemental Materials](#) for results). Lastly, we asked participants to write an opening line to send to their conversation partner and to evaluate a line ostensibly written by their conversation partner that read “this should be fun.” (see [Supplemental Materials](#) for results).

Study 8

Participants

We recruited an initial sample of 360 participants from Mechanical Turk, targeting a sample of 300 participants after pre-registered exclusions. Thirty-four participants were excluded for missing multiple attention checks (out of three), leaving a final sample of 326 participants ($M_{\text{age}} = 39.28$, $SD_{\text{age}} = 11.99$; 149 women, 173 men, 4 other gender; 208 liberal, 108 conservative).

Procedure

Participants again learned that they would have a chat with another MTurker. After rating whether they approved of six basic moral wrongs, they saw a loading animation for 20 seconds with text indicating they were being matched with a partner. They then learned they had been matched with “TeresaJ” who disagreed with them on politics. Participants either learned that TeresaJ rated all six of the basic wrongs as immoral or received no additional information about TeresaJ’s answers. In the single item correction condition, participants chose one of the six basic wrongs that they wanted to learn whether or not TeresaJ condemned. Participants then rated TeresaJ on trait dehumanization and indicated whether or not they would like to switch conversation partners (yes or no). Finally, they rated Democrats and Republicans on dehumanization and estimated the percentage of each party that approved of each basic moral wrong.

Notes

^a There was also a trend of increasing negativity on Twitter (see [S4 in Supplemental Materials](#)), but this was not limited to political tweets. By contrast, the increases in words about basic moral wrongs occurred mostly among political posts.

^b Increases in basic morality bias words were not limited to a handful of users, nor were the effects driven by comments about a few political elites or a single type of accusation (see [Figures S1–S3](#)).

^c Dividing the trait dehumanization scale into items that deny human nature (e.g., “mechanical and cold, like a robot”) and items that deny human uniqueness (“lacking self-restraint, like an animal”) produced the same results. Therefore, we combined both subscales in all studies.

^d To ensure these issues truly are morally unambiguous for both parties, we conducted a nationally representative survey of 641 Democrats and Republicans, asking about the acceptability of the 6 basic wrongs used in Study 3. Each issue was labeled as acceptable by fewer than 5% of both conservatives and liberals.

^e Analyses also revealed the basic morality bias was not confined to ideologically extreme partisans in Studies 2–3. Though ideological extremity was positively related to misperceptions, even partisans with the lowest scores on ideological identification perceived opponents (versus ingroup) as significantly more accepting of basic moral wrongs.

^f The effect of the basic morality correction did not differ among partisans with stronger or weaker ideological identification, but the effect of the shared activity condition was weaker among partisans who at least moderately identified as liberal or conservative (see [Supplemental Materials](#) for details).

Acknowledgments

We thank Sergio Barbosa De La Torre and Corry Cook for their helpful feedback. This paper was posted as a preprint at <https://osf.io/preprints/psyarxiv/fk8g6>.

Supplementary Material

[Supplementary material](#) is available at *PNAS Nexus* online.

Funding

This work was supported by the Charles Koch Foundation (Center for the Science of Moral Understanding).

Author Contributions

C.P.: Conceptualization, methodology, formal analysis, investigation, data curation, writing (original draft, review, and editing). E.K.: Conceptualization, methodology, formal analysis, investigation, data curation, writing (review and editing). C.S.: Conceptualization, methodology, formal analysis, investigation, data curation, writing (original draft). Y.E.B.: Conceptualization, methodology, and writing (review and editing). P.E.: Methodology and writing (review and editing). K.G.: Conceptualization, methodology, writing (review and editing), supervision, funding acquisition.

Data Availability

Survey materials, data, analysis scripts are available our OSF page. Preregistrations documents (Studies 3, 5, 6, 7, and 8) are also available <https://osf.io/sgm25/>.

References

- Finkel EJ, et al. 2020. Political sectarianism in America. *Science*. 370:533–536.
- Bartels LM. 2020. Ethnic antagonism erodes Republicans’ commitment to democracy. *Proc Natl Acad Sci U S A*. 117:22752–22759.
- Levendusky M, Malhotra N. 2016. Does media coverage of partisan polarization affect political attitudes? *Polit Commun*. 33:283–301.
- Yudkin D, Hawkins S, Dixon T. 2019. The perception gap: how false impressions are pulling Americans apart. *PsyArXiv*. <https://doi.org/10.31234/osf.io/r3h5q>, preprint: not peer reviewed.
- Ahler DJ, Sood G. 2018. The parties in our heads: misperceptions about party composition and their consequences. *J Polit*. 80:964–981.
- Moore-Berg SL, Hameiri B. 2024. Improving intergroup relations with meta-perception correction interventions. *Trends Cogn Sci*. 28:190–192.
- Ruggeri K, et al. 2021. The general fault in our fault lines. *Nat Hum Behav*. 5:1369–1380. <https://doi.org/10.1038/s41562-021-01092-x>
- Lees J, Cikara M. 2020. Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nat Hum Behav*. 4:279–286.
- Bastian B, Haslam N. 2010. Excluded from humanity: the dehumanizing effects of social ostracism. *J Exp Soc Psychol*. 46:107–113.
- Goodwin GP. 2015. Moral character in person perception. *Curr Dir Psychol Sci*. 24:38–44.
- Pacilli MG, Rocco M, Pagliaro S, Russo S. 2016. From political opponents to enemies? The role of perceived moral distance in the animalistic dehumanization of the political outgroup. *Group Process Intergroup Relat*. 19:360–373.
- Kubin E, Kachanoff FJ, Gray K. 2023. Threat rejection fuels political dehumanization. *Soc Psychol Personal Sci*. 14:487–500. <https://doi.org/10.1177/19485506211068922>
- Cassese EC. 2021. Partisan dehumanization in American politics. *Polit Behav*. 43:29–50.
- Martherus JL, Martinez AG, Piff PK, Theodoridis AG. 2021. Party animals? Extreme partisan polarization and dehumanization. *Polit Behav*. 43:517–540.
- McCoy M, Scully P. 2002. Deliberative dialogue to expand civic engagement: what kind of talk does democracy need? *Natl Civ Rev*. 91:117–135.
- Delgado MR, Frank RH, Phelps EA. 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci*. 8:1611–1618.

- 17 Sperber D, Baumard N. 2012. Moral reputation: an evolutionary and cognitive perspective. *Mind Lang.* 27:495–518.
- 18 McCoy J, Rahman T, Somer M. 2018. Polarization and the global crisis of democracy: common patterns, dynamics, and pernicious consequences for democratic polities. *Am Behav Sci.* 62:16–42.
- 19 Boyd R, Richerson PJ. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol Sociobiol.* 13:171–195.
- 20 de Waal F. 2006. *Primates and philosophers: how morality evolved.* Princeton (NJ): Princeton University Press.
- 21 Smith J. 2019. As Trump fans the flames of anti-abortion rhetoric, Kansas offers a cautionary tale. *The Intercept.* [accessed 2023 Dec 15]. <https://theintercept.com/2019/08/02/anti-abortion-violence-kansas/>.
- 22 Chait J. 2016. Yes, most Donald Trump supporters are deplorable and irredeemable. *Intelligencer.* [accessed 2023 Dec 15]. <https://nymag.com/intelligencer/2016/09/yes-most-donald-trump-supporters-are-deplorable.html>.
- 23 Brady WJ, Jackson JC, Lindström B, Crockett MJ. 2023. Algorithm-mediated social learning in online social networks. *Trends Cogn Sci.* 27:947–960.
- 24 Baliotti S, Getoor L, Goldstein DG, Watts DJ. 2021. Reducing opinion polarization: effects of exposure to similar people with differing political views. *Proc Natl Acad Sci U S A.* 118:e2112552118.
- 25 Kteily N, Bruneau E, Waytz A, Cotterill S. 2015. The ascent of man: theoretical and empirical evidence for blatant dehumanization. *J Pers Soc Psychol.* 109:901–931.
- 26 De Freitas J, Cikara M. 2018. Deep down my enemy is good: thinking about the true self reduces intergroup bias. *J Exp Soc Psychol.* 74:307–316.
- 27 Hartman R, Hester N, Gray K. 2023. People see political opponents as more stupid than evil. *Pers Soc Psychol Bull.* 49:1014–1027. <https://doi.org/10.1177/01461672221089451>
- 28 Bullock JG, Gerber AS, Hill SJ, Huber GA. 2013. Partisan bias in factual beliefs about politics. *Quart J Polit Sci.* 10:519–578.
- 29 Ferraro KF, Wilmoth JM. 2000. Measuring morbidity: disease counts, binary variables, and statistical power. *J Gerontol Ser B.* 55:S173–S189.
- 30 APA. 2016. 1 in 4 employees negatively affected by political talk at work this election. <https://www.apa.org>. [accessed 2021 Dec 8]. <https://www.apa.org/news/press/releases/2016/09/employees-political-talk>.
- 31 Dunn J, Ruedy NE, Schweitzer ME. 2012. It hurts both ways: how social comparisons harm affective and cognitive trust. *Organ Behav Hum Decis Process.* 117:2–14.
- 32 Malhotra D, Lumineau F. 2011. Trust and collaboration in the aftermath of conflict: the effects of contract structure. *Acad Manage J.* 54:981–998.
- 33 Carney DR, Jost JT, Gosling SD, Potter J. 2008. The secret lives of liberals and conservatives: personality profiles, interaction styles, and the things they leave behind. *Polit Psychol.* 29:807–840.
- 34 Pettigrew TF, Tropp LR. 2006. A meta-analytic test of intergroup contact theory. *J Pers Soc Psychol.* 90:751–783.
- 35 Bail CA, et al. 2018. Exposure to opposing views on social media can increase political polarization. *Proc Natl Acad Sci U S A.* 115:9216–9221.
- 36 Frimer JA, Skitka LJ. 2020. Are politically diverse thanksgiving dinners shorter than politically uniform ones? *PLoS One.* 15:e0239988.
- 37 Guffler K, Wagner U. 2017. Backfire of good intentions: unexpected long-term contact intervention effects in an intractable conflict area. *Peace Confl J Peace Psychol.* 23:383–391.
- 38 Prior M, Sood G, Khanna K. 2015. You cannot be serious: the impact of accuracy incentives on partisan bias in reports of economic perceptions. *Q J Polit Sci.* 10:489–518.
- 39 Peterson E, Iyengar S. 2021. Partisan gaps in political information and information-seeking behavior: motivated reasoning or cheerleading? *Am J Polit Sci.* 65:133–147.
- 40 Flynn DJ, Nyhan B, Reifler J. 2017. The nature and origins of misperceptions: understanding false and unsupported beliefs about politics. *Polit Psychol.* 38:127–150.
- 41 Shaffer SD. 1981. Balance theory and political cognitions. *Am Polit Q.* 9:291–320.
- 42 Peterson E, Iyengar S. 2022. Partisan reasoning in a high stakes environment: assessing partisan informational gaps on COVID-19. *HKS Misinfo Rev.* 3(2). <https://doi.org/10.37016/mr-2020-96>.
- 43 Malka A, Adelman M. 2023. Expressive survey responding: a closer look at the evidence and its implications for American democracy. *Perspect Polit.* 21:1198–1209.
- 44 Pennycook G, Thompson VA. 2016. Base-rate neglect. In: Pohl RF, editor. *Cognitive illusions: intriguing phenomena in thinking, judgment, and memory.* 2nd ed. London: Psychology Press. p. 44–61.
- 45 Huff C, Tingley D. 2015. Who are these people? evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Res Polit.* 2:2053168015604648.
- 46 Boxell L, Gentzkow M, Shapiro JM. 2024. Cross-country trends in affective polarization. *Rev Econ Stat.* 106:557–565.
- 47 Kubin E, Puryear C, Schein C, Gray K. 2021. Personal experiences bridge moral and political divides better than facts. *Proc Natl Acad Sci U S A.* 118:e2008389118.
- 48 Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R. 2015. Tweeting from left to right: is online political communication more than an echo chamber? *Psychol Sci.* 26:1531–1542.