Everyone I don't like is biased: Affective evaluations and the bias blind spot

Alexander C. Walker[1,2], Robert N. Collins[3], Heather E. K. Walker[1], Jonathan A. Fugelsang[1], and David R. Mandel[3]

[1] HumanSystems Inc., Guelph, ON, Canada

[2] Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA

[3] Defence Research and Development Canada, Toronto, ON, Canada

Author Note

Alexander C. Walker https://orcid.org/0000-0003-1431-6770

Robert N. Collins https://orcid.org/my-orcid?orcid=0000-0002-1714-7215

Heather E. K. Walker https://orcid.org/0000-0002-4109-2365

Jonathan A. Fugelsang https://orcid.org/0000-0002-6342-7023

David R. Mandel https://orcid.org/0000-0003-1036-2286

This research was supported by Canadian Safety and Security Program project CSSP-2018-TI-2394 under the direction of the last author. Correspondence concerning this article should be addressed to Alexander C. Walker, 190 Thayer Street, Providence, RI, USA, 02912. Email: alexander_walker1@brown.edu

Abstract

People commonly exhibit a bias blind spot, judging themselves to be less susceptible to bias than the "average other." However, less is known about how people attribute bias to familiar others who evoke strong affect. The present work (N = 1,980) examines the degree to which participants' attributions of bias are sensitive to their affective impressions of others. In Experiment 1, participants viewed themselves as considerably less biased than the average survey respondent and a personally-known disliked other but not less biased than a familiar individual who they liked. Experiments 2 and 3 examined the bias blind spot in politically polarized groups of Democrats and Republicans. While participants judged themselves to be somewhat less biased than the average co-partisan, they viewed themselves as much less susceptible to various psychological biases compared to their political opponents. Notably, in all experiments, the effect of other target selection (e.g., in-group vs. out-group member) on the bias blind spot was mediated by affective evaluations of target others. We discuss the theoretical implications of people using readily available affective evaluations as heuristic cues when attributing bias to familiar others. Additionally, we consider the practical implications of people failing to attribute bias to individuals they like, while readily attributing bias to those they dislike, including the potential for such attributions to exacerbate interpersonal and intergroup conflicts.

Keywords: bias blind spot, affect, interpersonal perception, social comparison, intergroup conflict, political polarization

Public Significance Statement

This study finds that readily available affective impressions guide how people attribute bias to others, leading people to view themselves and individuals they like as less prone to psychological biases than the average person or individuals they dislike. This result was demonstrated across both interpersonal and intergroup contexts and suggests that initial feelings of dislike can promote the view that a disliked individual or group is biased, with accusations of bias having the potential to hinder compromise and escalate conflict.

Transparency and Openness

For all experiments, we collected our full sample prior to data analyses and report all data exclusions, all manipulations, and all measures used. All measures and materials presented in the current study can be viewed in the supplementary materials (Part A).[1] Data and analyses scripts (i.e., annotated R scripts) pertaining to each experiment are available on Open Science Framework (osf.io/czvgy). Experiments 1, 2, and 3 were preregistered through Open Science Framework. These pre-registrations can be viewed via the following links (Exp. 1: osf.io/cpxgw; Exp. 2: osf.io/d3tm5; Exp. 3: osf.io/rhzqk).

[1] For the purpose of brevity, some measures are reported exclusively in the supplementary materials (Part A).

Introduction

Psychologists have revealed numerous biases that distort human judgment, both within the social (Nisbett & Ross, 1980; Nisbett & Wilson, 1977) and cognitive domain (Tversky & Kahneman, 1974). While people readily recognize these biases in others, they often fail to perceive bias in themselves (Pronin et al., 2002). This *bias blind spot* hinders judgment, as people fail to correct for biases that they do not know they have (Scopelliti et al., 2015). Furthermore, viewing oneself as objective while accusing others of bias can promote negative interpersonal interactions, facilitating feelings of distrust and impeding conflict resolution (Kennedy & Pronin, 2008; Pronin, 2008). Thus, the bias blind spot, in which people judge themselves as exhibiting less bias than their peers, represents an important meta-bias, one that can hinder judgment and provoke conflict.

Bias blind spots are said to arise as a result of two complementary mechanisms (Pronin et al., 2004; Pronin, 2007, 2008): (1) peoples' belief that they objectively perceive reality (i.e., naïve realism; Griffin & Ross, 1991; Ross & Ward, 1996), and (2) self-other asymmetries in the value[2] given to introspective evidence (Pronin & Kugler, 2007). Believing that we objectively perceive reality leads us to assume that the beliefs of others will correspond with our own "objective" assessments. When this assumption is proven false, people tend not to question the objectivity of their judgments, instead attributing a lack of information, or more damning, bias, to individuals' with opposing views. Self-other asymmetries regarding the value given to introspective evidence support the perception of an objective self in a world of biased others. The cognitive processes that engender psychological biases are thought to be nonconscious and, as such, inaccessible to introspection (Wilson & Brekke, 1994; Wilson et al., 2002). Still, people

[2] Relevant to this self-other asymmetry in *value* is the self-other asymmetry in *access* to introspective evidence, with people able to directly probe their own but not others' introspections.

tend to value introspective evidence when assessing their own—but not others—biases. Failing to find evidence of bias when consulting introspection, people are inclined to conclude that their judgments were unaffected by bias. Conversely, when assessing bias in others', people tend to use a different strategy, examining peoples' *behavior* for evidence of bias (Pronin et al., 2004). Due to the presumed nonconscious nature of many biases, this self-other asymmetry in *strategies* for assessing bias has been theorized to produce self-other asymmetries in *attributions* of bias that reflect a bias blind spot.

Much of the literature surrounding the bias blind spot has investigated the tendency for people to rate themselves as less biased than a hypothetical "average other" (e.g., the average American). Therefore, the extent to which people view themselves as less biased than their friends, family members, or enemies is less clear. Affective feelings towards familiar others may guide attributions of bias, with people viewing themselves as similarly biased to well-liked others (e.g., friends and family) and considerably less biased than individuals they dislike. Much evidence demonstrates the significant role of affect in human judgment (Lerner et al., 2015; Slovic et al., 2007). Affective reactions are generated quickly and automatically (Bargh et al., 1992; Zajonc, 1980), and guide how people attend to, encode, and interpret various events (Baumeister & Newman, 1994). As such, independent of potential self-other asymmetries in individuals' *strategies* for assessing bias, when evaluating the biases of familiar others, readily available affective impressions may exert a primary role in bias attributions. Nevertheless, the role of affect in attributions of bias, and consequently, the bias blind spot, is poorly understood. Do people view themselves as less biased than individuals they like? Do negative feelings towards a disliked group facilitate the view that members of this group are biased? The current study assessed these questions.

The Current Study

The tendency for people to view themselves as less biased than an "average other" (e.g., the average American) is well-established (Pronin & Hazel, 2023). However, less is known about how people attribute bias to familiar others who they have positive or negative feelings towards. In Experiment 1, participants judged the extent to which they and either a liked well-known other, a disliked well-known other, or the average survey respondent exhibited various psychological biases. We hypothesized that participants' attributions of bias would be sensitive to their affective evaluations of a target other. Therefore, we predicted that participants would attribute less bias to a liked well-known other than a disliked well-known other or the average survey respondent and consequently, exhibit smaller bias blind spots when randomly assigned to evaluate the biases of a well-liked individual.

Like interpersonal contexts, inter-group conflicts can evoke strong affective judgments typified by in-group love and out-group hate (Finkel et al., 2020). The extent to which affective evaluations guide attributions of bias is an important question, as viewing out-group members as biased can exacerbate group conflict (Kennedy & Pronin, 2008; Pronin et al., 2006). In Experiments 2 and 3, we leveraged Democrats' and Republicans' negative feelings towards one another and positive sentiments towards co-partisans (Iyengar et al., 2019) to assess the role of affect in attributions of bias within a polarized group context. Specifically, we recruited Democrats and Republicans and had them judge the extent to which they and the average Democrat and Republican exhibited different psychological biases.[3] We hypothesized that participants would attribute more bias to the average political out-group member than in-group member, resulting in participants exhibiting larger bias blind spots when political out-group

[3] Note that these biases were (largely) non-political in nature such that they pertained to general psychological tendencies (e.g., action-inaction bias) as opposed to, for example, the partisan treatment of political information.

members were the focal "other." Additionally, we hypothesized that affective feelings towards target others would guide attributions of bias. As such, we predicted that more positive affective evaluations of in-group members would be associated with partisans' failing to perceive bias in their political in-group, whereas more negative affective evaluations of out-group members would be associated with partisans' attributing more bias to their political opponents. Taken together, the present work investigates the extent to which affective impressions guide attributions of bias towards familiar others, and consequently, impact the magnitude of individuals' bias blind spots within interpersonal and intergroup contexts.

Experiment 1

Method

Participants

Six hundred and sixty participants were recruited from Prolific, an online research platform. To be eligible to participate in this study, participants were required to: (1) self-report an age between 18 and 60 years old, (2) self-report English as their first language, (3) reside in the United States, United Kingdom, Canada, Australia, or New Zealand, (4) have completed between 100 and 10,000 work submissions on Prolific, and (5) possess a Prolific submission approval rating greater than or equal to 95%. Participants received $4.50 USD upon completion of a 25-minute online questionnaire. We excluded data from 27 participants based on a set of pre-registered exclusion criteria (see *Data Preparation* below), leaving data from 633 participants (49% Female; M_{age} = 38.76, SD_{age} = 11.37) to be analyzed.

Measures and Materials

Bias Blind Spot. Participants completed a bias blind spot task adapted from Scopelliti and colleagues (2015). For this task, they were presented with descriptions of 14 biases (e.g.,

confirmation bias) and asked to evaluate the extent to which they (Self block) and a target other (Other block) exhibited each bias. Specifically, on each trial, participants were presented with a description of a bias (referred to as an effect or tendency) and asked, "To what extent do you believe that [you/the average survey respondent/the person you know well and personally] show[s] this effect or tendency?" Responses to this question ranged from 1 (*Not at all*) to 7 (*Very much*). We calculated the bias blind spot of each participant by subtracting their mean Self judgment from their mean Other judgment. As such, positive scores represented the belief that a target other exhibits the described biases more than oneself.

Bias Attribution Strategy. We assessed the extent to which participants used different strategies when judging the biases of themselves and others. Participants were asked: "When I rated the extent to which [I/the average survey respondent/the person I know well and personally] exhibited a psychological effect or tendency, I answered mainly based on:" and endorsed one of three presented strategies. These strategies included a strategy based on introspection, one based on behavior observation, and another based on declarative knowledge.

Affect. We assessed participants' affective feelings towards each target with the question: "How do you feel about [yourself/people in general/the person you know well and personally]?" We also measured participants' perceptions of how the average survey respondent (i.e., "people in general") or a well-known other feels about them. Responses to both questions were provided on a 7-point scale that ranged from -3 (*Extremely negative*) to +3 (*Extremely positive*).

Design and Procedure

Experiment 1 used a block design for which participants evaluated the extent to which they (Self block) and a target other (Other block) exhibited various biases, with block order counterbalanced. Each block began with 14 bias blind spot trials and concluded with participants

responding to strategy and affect-related questions. Random assignment determined whether participants assessed the biases of the average survey respondent, a liked well-known other, or a disliked well-known other during the Other block. Participants assigned to evaluate the biases of a well-known other were asked to think about a person they knew well and personally whom they felt either positively or negatively towards, depending on condition.

Data Preparation

We excluded data from 27 participants who indicated that English was not their first language (n = 2), reported problems viewing study materials (n = 7), selected a different imagined well-known other before and during the bias blind spot task (n = 9), or completed Experiment 1 in under 500 seconds (n = 10).[4] These exclusions were consistent with pre-registered criteria.

Results

Participants demonstrated a bias blind spot, $t(632) = 17.14$, $p < .001$, $d = 0.77$, 95% *CI* [0.65, 0.95], judging themselves (M = 3.56, SD = 0.90) as less biased than a target other (M = 4.31, SD = 1.03). Critically, we assessed the degree to which participants' bias blind spots differed as a result of the target other they evaluated. A mixed ANOVA featuring Target (Self vs. Other) as a within-subjects factor and Other Reference Group (Average survey respondent vs. Liked well-known other vs. Disliked well-known other) as a between-subjects factor revealed a Target by Other Reference Group interaction, $F(2, 627) = 104.35$, $p < .001$, $\eta_p^2 = .250$ (see Figure 1). While we observed small differences in bias evaluations of the self across Other Reference Group (all $d < 0.33$), considerably less bias was attributed to liked well-known others (M = 3.50, SD = 0.93) compared to disliked well-known others (M = 4.69, SD = 0.89), $t(417) = 13.25$, $p < .001$, $d = 1.29$, 95% *CI* [1.07, 1.55], and the average survey respondent (M = 4.72, SD

[4] Note that one participant exhibited multiple grounds for exclusion.

= 0.77), $t(422) = 14.71$, $p < .001$, $d = 1.43$, 95% *CI* [1.20, 1.68].[5] Consequently, participants evaluating the biases of the average survey respondent ($M = 1.00$, $SD = 0.97$) or a disliked well-known other ($M = 1.25$, $SD = 1.04$) displayed large bias blind spots ($p < .001$ and $d > 1.19$ in both cases). Conversely, participants did not demonstrate a bias blind spot when evaluating the biases of a liked well-known other ($M = -0.02$, $SD = 0.83$), $t(209) = -0.27$, $p = .784$, $d = -0.02$, 95% *CI* [-0.15, 0.11], viewing themselves as similarly biased as this liked individual.

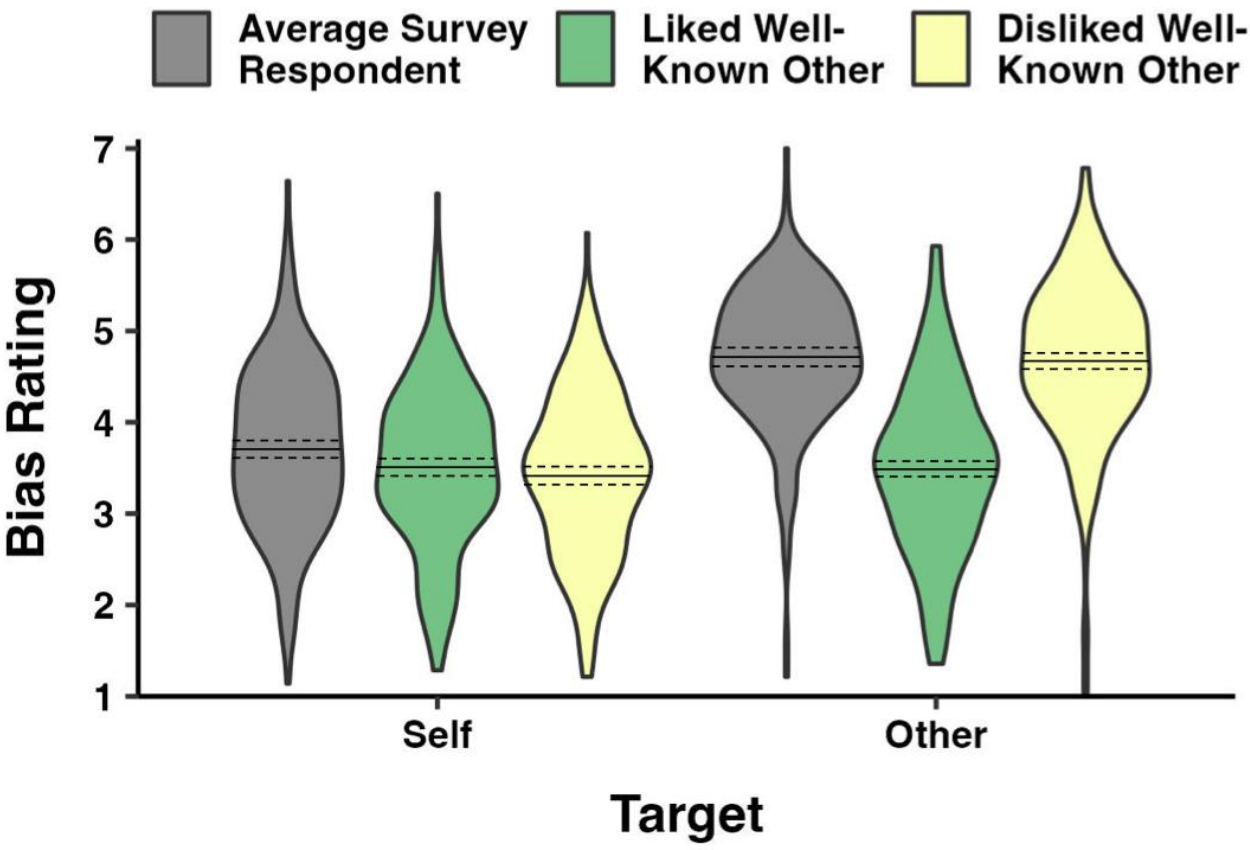

Figure 1. Experiment 1 Results: Bias Attributions. This figure displays the distribution of participants' mean bias ratings for each Target and Other Reference Group condition. Solid lines represent the mean bias rating within a condition while dashed lines represent the 95% confidence interval.

[5] This finding generalized across biases as, for all 14 bias blind spot items, liked well-known others were judged as exhibiting the described bias less than disliked well-known others (all $p < .013$) and the average survey respondent (all $p < .001$).

Participants' attributions of bias were sensitive to their affective evaluations. We observed a strong negative association between participants' affective evaluations of a target other and the magnitude of their bias blind spot, $r(631) = -.52$, $p < .001$. That is, viewing a target other negatively was associated with perceiving this target other as more biased, $r(631) = -.50$, $p < .001$, and consequently, with larger bias blind spots. Mediation analyses revealed that the effect of Other Reference Group on participants' bias blind spots was partially mediated by their affective evaluations of a target other ($b = 0.89$, 95% *CI* [0.57, 1.22]; see Supplementary Materials Part B). Furthermore, participants' *perceptions* of how a well-known other (but not "people in general") felt about *themselves* was negatively correlated with their bias blind spot, $r(417) = -.46$, $p < .001$, as the more participants perceived a well-known other as holding negative feelings towards them the more bias they attributed to this familiar individual, $r(417) = -.52$, $p < .001$. Finally, we observed a small correlation such that participants who reported more positive feelings towards themselves also tended to view themselves as less biased, $r(631) = -.10$, $p = .013$. This modest correlation with reference to the self was significantly smaller than the former correlation which referenced the other, $z = -7.52$, $p < .001$, suggesting that it is affective evaluations of others, and not oneself, that modulates the bias blind spot.

Bias Attribution Strategies and the Bias Blind Spot

Consistent with past work (Pronin et al., 2004), a majority of participants (61.51%)[6] in Experiment 1 endorsed a behavior observation strategy when assessing the biases of a target other, $\chi^2 = 32.93$, $p < .001$. However, participants did not favor an introspective strategy (Behavior: 51.03% vs. Introspection: 48.97%) when attributing bias to themselves, $\chi^2 = 0.27$, $p = .604$, nor was reliance on introspective evidence correlated with attributions of bias regarding

[6] These analyses removed instances in which participants endorsed a declarative knowledge strategy as a) this strategy was endorsed by only a handful of participants ($n = 12$) and b) was not pertinent to the claims assessed here.

oneself or a target other, $r < .04$, $p > .320$. Thus, the results of Experiment 1 were largely inconsistent with a prominent explanation of the bias blind spot that claims that this meta-bias emerges, at least in part, as a result of people using introspective evidence to probe their own biases, and behavior observation to assess bias in others (Pronin, 2007, 2008).

Experiment 2

Experiment 2 investigated the role of affective evaluations in attributions of bias within a polarized group context. Instead of manipulating affect directly, we utilized Democrats' and Republicans' positive affect towards in-group members and negative affect towards out-group members (Iyengar et al., 2019). We hypothesized that partisans would judge political out-group—compared to in-group—members as more susceptible to a host of psychological biases, displaying larger bias blind spots when assessing bias in rival partisans. Furthermore, we hypothesized that the effect of other reference group (in-group vs. out-group member) on partisans' bias blind spots would be mediated by their affective evaluation of a target other. Thus, we expected stronger positive affect towards one's in-group to be associated with partisans viewing in-group members as less biased and stronger negative affect towards one's out-group to be associated with partisans attributing more bias to their political opponents.

Method

Participants

We recruited a sample of 660 participants from Prolific using the same recruitment criteria as Experiment 1 with the exception that all participants were required to reside in the United States and self-report their political affiliation as either Democrat or Republican. Participants received $4.50 USD upon completion of a 25-minute online questionnaire. We

excluded data from 34 participants based on pre-registered exclusion criteria, leaving data from 626 participants (51% Female; M_{age} = 39.77, SD_{age} = 11.12; 50% Democrat) to be analyzed.

Materials

People are less likely to acknowledge exhibiting biases that they perceive as socially undesirable (Pronin et al., 2002). As such, we conducted a pilot study (see Supplementary Materials Part C for a full report) to ensure that Democrats and Republicans viewed all biases described in Experiment 2 as similarly "bias-like" and negative (i.e., socially undesirable). The results of this study revealed that Democrats viewed the described biases (adapted from Scopelliti et al., 2015) more negatively than Republicans ($p < .001$, $d = 0.85$; Figure 2). Item-level analyses showed significant partisan differences for 7 (out of 14) bias blind spot items.[7] Based on these findings, we selected seven bias blind spot items for inclusion in Experiment 2, all of which were viewed as similarly "bias-like" and negative by Democrats and Republicans.

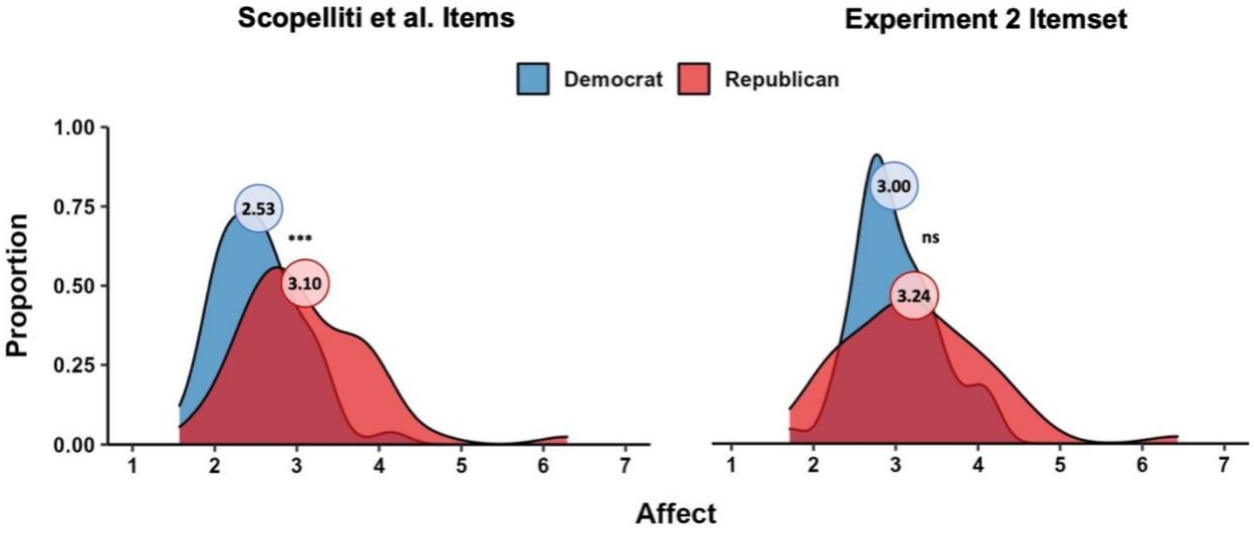

Figure 2. Item Selection Pre-Test: Affect. This figure displays the distribution of participants' (*N*

[7] Note that it may not have been the biases themselves that evoked different reactions from Democrats and Republicans, but rather how these biases were described. To provide one example, the in-group favoritism item, created by Scopelliti and colleagues (2015), for which we observed partisan differences, focuses on peoples' "unconscious and automatic tendency to be less generous to people of a different race than to people of their race." It is possible that describing in-group favoritism in a different manner may reduce or even eliminate the partisan differences observed here.

= 100) mean affective judgment of 14 psychological biases described within bias blind spot items adapted from Scopelliti and colleagues (2015). Participants' responses within this pilot study informed the selection of seven non-partisan bias blind spot items for inclusion in Experiment 2. The distribution of affective judgments regarding these seven remaining items is shown in the right panel. Results of a *t*-test comparing the affective judgments of Democrats and Republicans are depicted (*** $p < .001$, *ns* $p > .05$) as are the mean affective judgments of Democrats and Republicans within each item set. Affect: "How do you view this effect or tendency?" (1 = *Extremely Negative*, 7 = *Extremely Positive*).

Measures and Procedure

The methodology of Experiment 2 largely mirrored that of Experiment 1. Participants completed a bias blind spot task in which they were presented with descriptions of seven biases and evaluated the extent to which they (Self block) and a target other (Other block) exhibited each bias. Block order was counterbalanced. Following all bias blind spot items within a given block, participants responded to bias attribution strategy and affect-related questions, as in Experiment 1.[8] Random assignment determined whether participants evaluated the biases of the average Democrat or Republican during the Other block. Following the bias blind spot task, participants responded to three items, asking them how biased (*bias non-specific*) they viewed themselves, the average Democrat, and the average Republican.

Data Preparation

We excluded data from 34 participants who self-reported an age 61 years or older ($n = 1$), indicated that English was not their first language ($n = 4$), reported problems viewing study materials ($n = 7$), or completed Experiment 2 in under 450 seconds ($n = 22$). These exclusions were consistent with our pre-registered criteria.

Results

[8] These questions were slightly altered in Experiment 2 (see Supplementary Materials Part A) in order to accommodate Experiment 2's focus on political partisanship.

Participants exhibited a bias blind spot, $t(625) = 19.06$, $p < .001$, $d = 0.94$, 95% *CI* [0.82, 1.07], judging themselves ($M = 3.66$, $SD = 1.09$) as less biased than a target other ($M = 4.67$, $SD = 1.07$). To assess whether participants exhibited larger bias blind spots when evaluating the biases of the average member of their political out-group (as opposed to in-group), we conducted a mixed ANOVA with Target (Self vs. Other) as a within-subjects factor and Participant Partisanship (Democrat vs. Republican) and Other Reference Group (In-Group Member vs. Out-Group Member) as between-subjects factors. We observed a Target by Other Reference Group interaction, $F(1, 622) = 152.26$, $p < .001$, $\eta_p^2 = .197$ (see Figure 3). While bias evaluations of the self did not differ across Other Reference Group, $t(624) = 1.57$, $p = .116$, $d = 0.13$, 95% *CI* [-0.03, 0.28], participants attributed significantly more bias to the average member of their political out-group ($M = 5.18$, $SD = 0.91$) than in-group ($M = 4.14$, $SD = 0.96$), $t(624) = 13.91$, $p < .001$, $d = 1.11$, 95% *CI* [0.95, 1.29].[9] Consequently, while participants displayed a bias blind spot ($p < .001$) when evaluating the biases of the average co-partisan ($M = 0.41$, $SD = 0.98$), they exhibited larger bias blind spots when evaluating the biases of the average out-group member ($M = 1.59$, $SD = 1.38$), $t(624) = 12.31$, $p < .001$, $d = 0.99$, 95% *CI* [0.82, 1.17]. There was no main effect of Participant Partisanship, $F(1, 622) = 0.30$, $p = .582$, $\eta_p^2 < .001$, nor did we observe any additional interactions (all $p > .05$). Responses to items asking participants how biased they view themselves, the average Democrat, and the average Republican (bias non-specific), mirrored responses provided within the bias blind spot task (which focused on seven specific psychological biases). That is, while participants viewed themselves as less biased than the average member of their political in-group ($p < .001$, $d = 0.38$), they displayed a larger bias blind spot when attributing bias to their political opponents ($p < .001$, $d = 0.92$).

[9] This finding generalized across biases as, for all seven bias blind spot items, political in-group members were judged as exhibiting the described bias less than political out-group members (all $p < .001$).

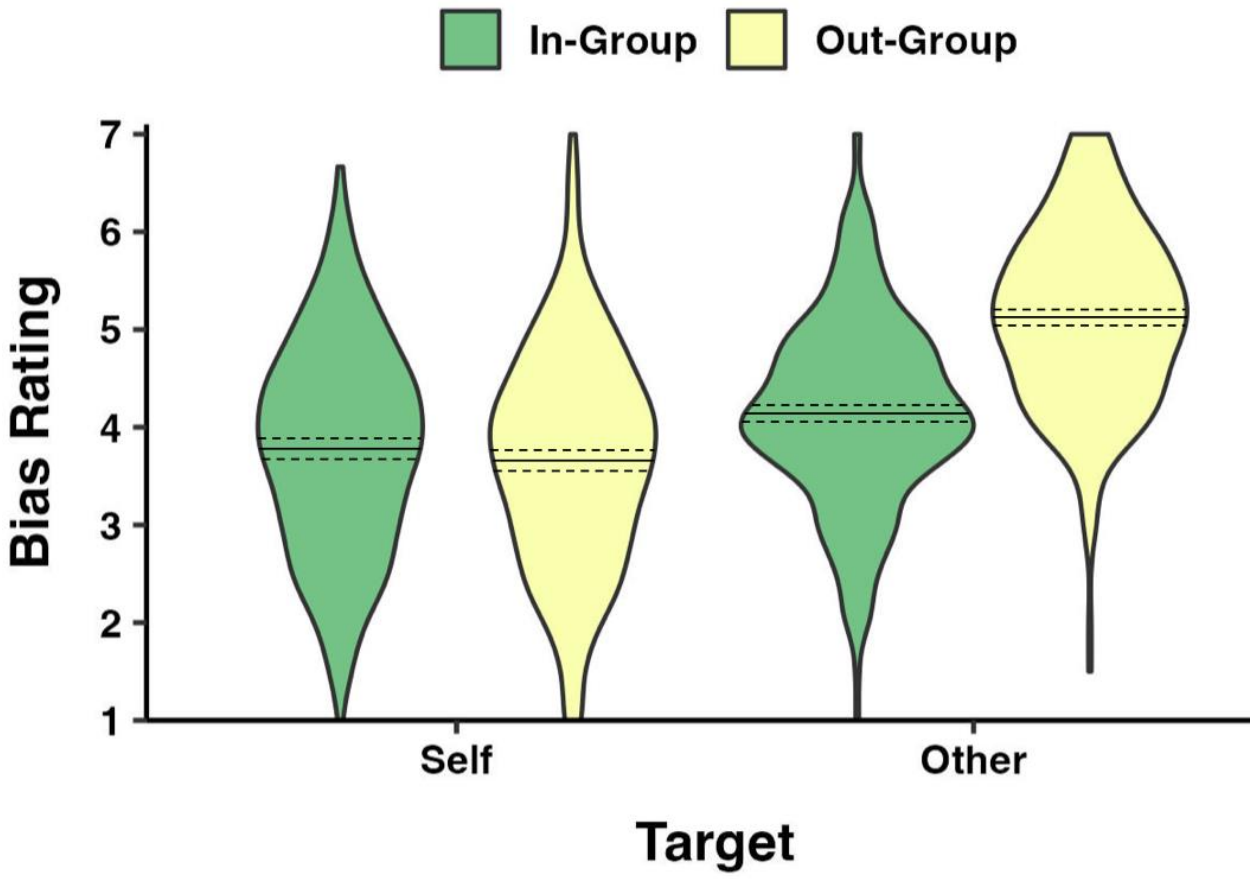

Figure 3. Experiment 2 Results: Bias Attributions. This figure displays the distribution of participants' mean bias ratings for each Target and Other Reference Group condition. Solid lines represent the mean bias rating within a condition while dashed lines represent the 95% confidence interval.

Participants' attributions of bias were once again sensitive to their affective evaluations. Replicating the results of Experiment 1, negative feelings towards a target other were associated with viewing this target other as more biased, $r(624) = -.55$, $p < .001$, and consequently, greater bias blind spots, $r(624) = -.47$, $p < .001$. Additionally, the effect of Other Reference Group on participants' bias blind spots was partially mediated by their affective evaluations of target others ($b = 0.62$, 95% *CI* [0.40, 0.80]; see Supplementary Materials Part B). Moreover, participants' perceptions of how a target other felt about themselves were negatively correlated with their bias blind spot, $r(624) = -.47$, $p < .001$, with participants reporting that a target other possessed negative feelings towards them tending to view this target other as more biased, $r(624) = -.52$, p

< .001. Finally, as in Experiment 1, participants who reported more positive feelings towards themselves also tended to view themselves as less biased, $r(624) = -.11$, $p = .006$. This modest correlation with reference to the self was significantly smaller than the former correlation referencing a target other, $z = -8.21$, $p < .001$.

Bias Attribution Strategies and the Bias Blind Spot

Inconsistent with Experiment 1 and an introspection-based account of the bias blind spot (Pronin, 2007, 2008), participants did not reliably favor a behavioral observation strategy when assessing the biases of a target other (Behavior: 50.08% vs. Introspection: 49.92%),[10] $\chi^2 < 0.01$, $p = .968$. Also inconsistent with this account was the finding that a majority of participants (65.95%) relied on a behavioral observation strategy when assessing their own biases, $\chi^2 = 61.24$, $p < .001$. Furthermore, we did not observe an association between participants' strategy for assessing bias and bias attributions when attributing bias to the self, $r(600) < .01$, $p = .935$. However, we did observe a small correlation between use of an introspection strategy and lower attributions of bias when attributing bias to others, $r(601) = -.16$, $p < .001$.

Experiment 3

Experiment 3 provided a replication of Experiment 2, with a primary modification being that in Experiment 3 participants evaluated the biases of both the average Democrat and Republican.

Method

Participants

[10] As in Experiment 1, analyses reported within this section removed instances in which participants ($n = 41$) reported relying on declarative knowledge when assessing bias in either themselves or a target other.

We recruited 660 participants from Prolific using the same criteria as Experiment 2. Based on pre-registered criteria, we excluded data from 24 participants, leaving data from 636 participants (49% Female; M_{age} = 38.26, SD_{age} = 10.80; 50% Democrat) to be analyzed.

Measures and Materials

All measures and materials described in Experiment 2 were featured in Experiment 3 with the exception that Experiment 3 no longer featured bias attribution strategy items.[11]

Design and Procedure

The design of Experiment 3 mirrored that of Experiment 2, with one exception. Participants in Experiment 3 completed an additional bias blind spot task block. That is, unlike in Experiment 2, all participants evaluated the extent to which both the average Democrat *and* the average Republican exhibited seven distinct biases. As in previous experiments, participants also evaluated the extent to which they exhibited these biases (Self block). The presentation of Self vs. Other blocks as well as Democrat vs. Republican Other blocks were counterbalanced.

Data Preparation

We excluded data from 24 participants who self-reported an age 61 years or older (n = 1), indicated that English was not their first language (n = 1), reported problems viewing study materials (n = 9), or completed Experiment 3 in under 450 seconds (n = 13). These exclusions were consistent with pre-registered criteria.

Results

Participants displayed a bias blind spot, $t(635) = 22.97$, $p < .001$, $d = 0.91$, 95% *CI* [0.81, 1.01], judging themselves (M = 3.66, SD = 1.07) as less biased than target others (M = 4.62, SD = 0.70). We assessed whether the magnitude of this bias blind spot increased when the target other was the average member of one's political out-group (as opposed to in-group). A mixed

[11] These items were removed on account of the null findings observed in Experiments 1 and 2.

ANOVA featuring Participant Partisanship (Democrat vs. Republican) as a between-subjects factor and Other Reference Group (in-group member vs. out-group member) as a within-subjects factor revealed a main effect of Other Reference Group, $F(1,634) = 415.88$, $p < .001$, $\eta_p^2 = .396$. That is, while participants exhibited a bias blind spot ($p < .001$) when evaluating the biases of the average in-group member ($M = 0.43$, $SD = 1.00$), they exhibited much larger bias blind spots when evaluating the biases of the average out-group member ($M = 1.48$, $SD = 1.45$), $t(635) = 19.91$, $p < .001$, $d = 0.79$, 95% CI [0.71, 0.87].[12] We also observed a main effect of Participant Partisanship, $F(1,634) = 9.10$, $p = .003$, $\eta_p^2 = .014$, and a Participant Partisanship by Other Reference Group interaction, $F(1,634) = 30.19$, $p < .001$, $\eta_p^2 = .045$ (see Figure 4). Democrats displayed larger bias blind spots ($M = 1.08$, $SD = 0.97$) than Republicans ($M = 0.83$, $SD = 1.11$). These effects resulted from Democrats attributing greater bias to their political opponents. That is, Democrats and Republicans in our sample did not differ in the extent to which they attributed bias to themselves ($p = .338$) or the average in-group member ($p = .651$). However, Democrats attributed significantly more bias to the average out-group member ($M = 5.38$, $SD = 0.84$) than did Republicans ($M = 4.92$, $SD = 0.94$), $t(634) = 6.41$, $p < .001$, $d = 0.51$, 95% CI [0.36, 0.67]. As such, Democrats displayed larger bias blind spots than Republicans when their self-attributions of bias were compared to their attributions of bias towards the average out-group ($p < .001$, $d = 0.38$), but not in-group member ($p = .651$, $d = 0.03$).

[12] As in Experiment 2, this effect of Other Reference Group generalized across biases (all $p < .001$).

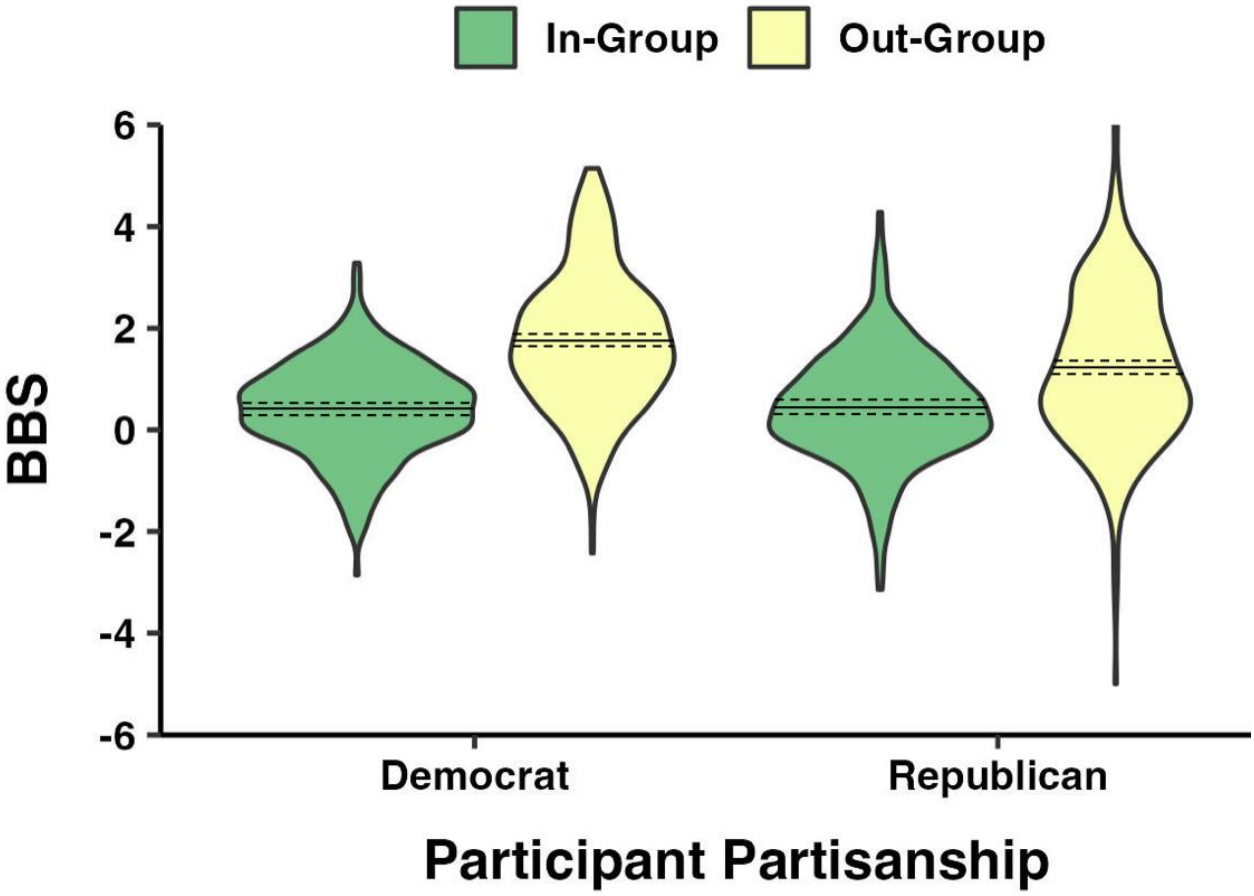

Figure 4. Experiment 3 Results: Bias Blind Spots. This figure displays the distribution of participants' bias blind spots (other bias rating – self bias rating) for each Other Reference Group. Solid lines represent the mean bias rating within a condition while dashed lines represent the 95% confidence interval.

Responses to questions asking participants to attribute bias to themselves and others in a bias non-specific manner replicated these findings, suggesting that our results did not depend on the specific psychological biases assessed. For example, within these single-item measures, participants viewed themselves as less biased than the average co-partisan ($p < .001$, $d = 0.37$) and displayed larger bias blind spots when attributing bias to the average out-group member ($p < .001$, $d = 0.89$). Moreover, Democrats attributed less bias to the average in-group member ($p = .012$, $d = 0.20$) and more bias to the average out-group member ($p = .016$, $d = 0.19$) than Republicans.

As in Experiments 1 and 2, negative affect towards a target other was associated with judging a target other as more biased, $r(634) = -.27$, $p < .001$, and consequently, with greater bias blind spots, $r(634) = -.20$, $p < .001$. Likewise, the *perception* that a target other possessed negative feelings towards *oneself* was associated with judging this target other as more biased, $r(634) = -.28$, $p < .001$, and larger bias blind spots, $r(634) = -.16$, $p < .001$. Affective evaluations of target others fully mediated the effect of Other Reference Group on the magnitude of participants' bias blind spots ($b = -0.98$, 95% *CI* [-1.14, -0.82]; see Supplementary Materials Part B). Unlike in Experiments 1 and 2, we did not observe an association between affective judgments and bias attributions when the target of evaluation was oneself, $r(634) = -.07$, $p = .069$. This non-significant correlation with reference to the self was smaller than the former correlation which referenced the other, $z = -3.87$, $p < .001$. In combination with the findings of Experiments 1 and 2, this finding provides strong support for our hypothesis that, comparatively speaking, it is one's affect towards others—rather than oneself—that shapes the bias blind spot.

General Discussion

Across three experiments, we find that people's tendency to attribute bias to others, while denying bias in oneself, is sensitive to their affective evaluations of targets, particularly target others. Consistent with past examinations of the bias blind spot (Mandel et al., 2022; Pronin et al., 2002), participants in Experiment 1 attributed more bias to the average survey respondent than themselves. However, when asked to assess the biases of a liked well-known other, participants no longer displayed a bias blind spot, viewing themselves and this liked individual as similarly biased. Notably, it was not participants' access to individuating information that eliminated their bias blind spot, as they attributed considerably more bias to a familiar disliked individual than themselves. Rather, affective evaluations of target others were strongly correlated

with attributions of bias, such that participants attributed less bias to target others the more positively they felt towards them. Thus, in many interpersonal contexts, people may seldom perceive themselves as less biased than the people they know and interact with.

Failure to attribute bias to individuals we like, while readily attributing bias to those we dislike has theoretical and practical implications. For instance, the association between affect and attributions of bias can promote "bias-perception conflict spirals," previously observed within the context of disagreement (Kennedy & Pronin, 2008). In the case of affect, initial feelings of dislike can facilitate the view that a disliked individual is biased. This attribution of bias can then encourage individuals to engage in conflict-escalating actions (as has been shown in prior work; Kennedy & Pronin, 2008; Pronin et al., 2006) that, over time, lead to stronger feelings of dislike and consequently, greater attributions of bias. Such negative feedback loops may be most pertinent to group contexts characterized by dislike and disagreement, such as political contexts featuring competing groups with distinct values. Much work has documented the rise of political polarization and partisan animosity in the United States (Finkel et al., 2020) and other countries (Boxell et al., 2022). Given peoples' proclivity to attribute bias more readily to individuals they dislike, rising partisan animosity can be expected to facilitate greater attributions of bias across political divides which, as we have discussed, can escalate conflict and lead to even greater partisan animus in the future.[13]

Experiments 2 and 3 examined the link between affect and attributions of bias in an intergroup context. Here, we demonstrate that partisan Americans view the average member of

[13] Rising partisan animosity, insofar as it increases the extent to which partisans view their political opponents as biased, may also shape partisans' consumption of political information. Perceiving political out-group members as biased may lead partisans to self-select into political echo chambers and disengage from political discussions featuring individuals with opposing views. Consistent with these claims, attributing more bias to the average political out-group member was associated with a stronger preference for ideologically congruent media in Experiment 3, $r(634) = .24$, $p < .001$ (see Supplementary Materials Part D).

their political out-group as far more susceptible to a host of psychological biases than themselves or the average member of their political in-group. Furthermore, we once again observed an association between participants' affective evaluations and their attributions of bias. Strong positive feelings towards one's political in-group were associated with viewing the average in-group member as less biased, while strong negative evaluations of one's political out-group predicted greater attributions of bias towards opposing partisans. Even *perceptions* of affect were correlated with attributions of bias. For example, participants perceiving more negative feelings between Democrats and Republicans attributed more bias to the average member of their political out-group and exhibited larger bias blind spots. Therefore, even when lacking negative affect towards rival partisans, heightened perceptions of partisan animosity—even if exaggerated[14]—may lead partisans to more readily impute bias to their political opponents.

Affect and the Bias Blind Spot: Theoretical Implications

A prominent account of the bias blind spot states that this meta-bias results, at least in part, from peoples' tendency to value their own introspective evidence, while dismissing the introspective evidence of others (Pronin, 2007, 2008). Distrusting the introspections of others, people instead focus on others' behavior when assessing their level of bias (Pronin et al., 2004). Given that the cognitive processes that produce many psychological biases are inaccessible to introspection (Wilson & Brekke, 1994; Wilson et al., 2002), this self-other asymmetry in *strategies* for assessing bias has been theorized to produce the self-other asymmetry in *attributions* of bias that characterize the bias blind spot. Contrary to this account, a majority of participants in Experiments 1 and 2 endorsed a behavioral observation—as opposed to introspection—strategy when assessing their own biases. Likewise, while a majority of

[14] Consistent with prior work on "false polarization" (Fernbach & Boven, 2022; Lees & Cikara, 2021), participants in Experiment 2 overestimated how biased members of their political out-group viewed the average member of their political in-group ($p < .001$, $d = 0.23$; see Supplementary Materials Part D).

participants endorsed a behavioral observation strategy when assessing the biases of a target other, this preference was small and statistically unreliable in Experiment 2. Additionally, reliance on introspective evidence was weakly correlated with lower attributions of bias in Experiment 2, with no such relation being observed in Experiment 1. Thus, while participants exhibited bias blind spots in Experiments 1 and 2, this meta-bias was not readily explained by self-other asymmetries in reliance on introspective evidence, as (a) such asymmetries were inconsistently observed and (b) bias assessment strategies were not reliably associated with attributions of bias.

In the current study, participants' affective evaluations of familiar others mediated how much bias they attributed to familiar others and consequently, the extent to which they viewed themselves as less biased than individuals they know and interact with. Furthermore, in a polarized group context, affective impressions of the "average" Democrat and Republican predicted the extent to which American partisans attributed different psychological biases to political in-group and out-group members. Affective impressions are generated quickly and automatically (Bargh et al., 1992; Zajonc, 1980) and play a significant role in human judgment (Lerner et al., 2015; Slovic et al., 2007). The present findings are consistent with participants using readily available affective feelings as heuristic cues when assessing the biases of others. Thus, independent of individuals' strategies for assessing bias, affective feelings may prompt them to attribute objectivity to people they like and bias to those they dislike, consistent with the observed elimination of the bias blind spot when participants evaluated bias in themselves and a well-liked other.

Constraints on Generality

The present work primarily recruited participants from the United States, limiting our ability to generalize our findings to other cultures. Nevertheless, we expect the present findings to generalize across cultures. Relatedly, consistent with the results of the present study, we expect our findings to generalize across psychological biases. However, researchers should ensure that the biases they assess are viewed as socially undesirable and are endorsed as biases by participants, particularly if wishing to make comparisons between groups with distinct values (e.g., Democrats and Republicans) or recruiting participants from previously unassessed cultures. The pilot study discussed in Experiment 2 and reported in full in the Supplementary Materials (Part C) provides one example of this practice. Speaking to our investigation of bias attributions in a polarized group context (Experiments 2 and 3), we expect our findings to generalize to other intergroup contexts where individuals' possess positive affective evaluations of in-group members and negative affective evaluations of out-group members. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Conclusion

People exhibit a bias blind spot, perceiving themselves as less susceptible to bias than the "average other" (Pronin & Hazel, 2023). The present investigation provides insight into how people attribute bias to themselves and a variety of target others who evoke significant affect. Across interpersonal and intergroup contexts, we show that attributions of bias are sensitive to affective evaluations of others, with people perceiving themselves and individuals they like as objective while viewing disliked individuals as considerably biased. The present findings have theoretical implications for current theories of bias attribution, suggesting that affective evaluations can serve as a heuristic cue that guides how people attribute bias to others. These findings also have practical implications, suggesting that initial feelings of dislike can facilitate

the perception that a disliked individual or group is biased, encouraging conflict-escalating actions and hindering compromise. In a political domain characterized by rising partisan animosity, understanding how affective feelings shape attributions of bias is of considerable importance, as effective governance is not possible when people view their political opponents as hopelessly biased.

References

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, *62*(6), 893-912. https://doi.org/10.1037/0022-3514.62.6.893

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin*, *20*(1), 3-19. https://doi.org/10.1177/0146167294201001

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2022). Cross-country trends in affective polarization. *Review of Economics and Statistics*, 1-60. http://www.nber.org/papers/w26669

Fernbach, P. M., & Van Boven, L. (2022). False polarization: Cognitive mechanisms and potential solutions. *Current Opinion in Psychology*, *43*, 1-6. https://doi.org/10.1016/j.copsyc.2021.06.005

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., ... & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533-536. https://doi.org/10.1126/science.abe1715

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. *Advances in Experimental Social Psychology*, *24*, 319-359. https://doi.org/10.1016/S0065-2601(08)60333-0

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129-146. https://doi.org/10.1146/annurev-polisci-051117-073034

Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin*, *34*(6), 833-848. https://doi.org/10.1177/0146167208315158

Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B*, *376*(1822), 20200143. https://doi.org/10.1098/rstb.2020.0143

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799-823. https://doi.org/10.1146/annurev-psych-010213-115043

Mandel, D. R., Collins, R. C., Walker, A. C., Fugelsang, J. A., & Risko, E. F. (2022). Hypothesized drivers of the bias blind spot—cognitive sophistication, introspection bias, and conversational processes. *Judgment and Decision Making*, *17*(6), 1-40. https://doi.org/10.1017/S1930297500009475

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250-256. https://doi.org/10.1037/0022-3514.35.4.250

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, *11*(1), 37-43. https://doi.org/10.1016/j.tics.2006.11.001

Pronin, E. (2008). How we see ourselves and how we see others. *Science*, *320*(5880), 1177-1180. https://doi.org/10.1126/science.1154199

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*(3), 781-799. https://doi.org/10.1037/0033-295X.111.3.781

Pronin, E., & Hazel, L. (2023). Humans' bias blind spot and its societal significance. *Current Directions in Psychological Science*, *32*(5), 402-409. https://doi.org/10.1177/0963721423117874

Pronin, E., Kennedy, K., & Butsch, S. (2006). Bombing versus negotiating: How preferences for combating terrorism are affected by perceived terrorist rationality. *Basic and Applied Social Psychology*, *28*(4), 385-392. https://doi.org/10.1207/s15324834basp2804_12

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, *43*(4), 565-578. https://doi.org/10.1016/j.jesp.2006.05.011

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381. https://doi.org/10.1177/0146167202286008

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In T. Brown, E. S. Reed, & E. Turiel (Eds.), *Values and knowledge* (pp. 103-135). Hillsdale, NJ: Erlbaum.

Scopelliti, I., Morewedge, C. K., McCormick, E., Min, H. L., Lebrecht, S., & Kassam, K. S. (2015). Bias blind spot: Structure, measurement, and consequences. *Management Science*, *61*(10), 2468-2486. https://doi.org/10.1287/mnsc.2014.2096

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333-1352. https://doi.org/10.1016/j.ejor.2005.04.006

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117-142. https://doi.org/10.1037/0033-2909.116.1.117

Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In D. Griffin & T. Gilovich (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 185-200). New York: Cambridge.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151-175. https://doi.org/10.1037/0003-066X.35.2.151