To appear in: *Personality and Social Psychology Bulletin* [Volume/Issue not yet assigned]

Citation: Walker, A. C., Collins, R. N., Walker, H. E. K., Fugelsang, J. A., & Mandel, D. R. (In Press). Everyone I don't like is biased: Affective evaluations and the bias blind spot. *Personality and Social Psychology Bulletin*.

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 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. All data has been made available at the following link: https://osf.io/czvgy/. APA ethical standards were followed in the conduct of all studies, and we obtained informed consent from all participants following an explanation of the nature and possible consequences of the study they participated in. We declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Correspondence concerning this article should be addressed to Alexander C. Walker, 190 Thayer Street, Providence, RI, USA, 02912. Email: alexandercwalker@icloud.com

Abstract

People commonly exhibit a bias blind spot (BBS), judging themselves as less susceptible to bias than the "average other." However, less is known about how people attribute bias to familiar others who evoke strong affect. We examined whether attributions of bias are sensitive to 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 whom they liked. Experiments 2 and 3 examined the BBS in politically polarized groups of Democrats and Republicans. While participants judged themselves as somewhat less biased than co-partisans, they viewed themselves as much less biased than their political opponents. In all experiments, the effect of other target selection on the BBS was mediated by affective evaluations. We discuss the theoretical implications of affective evaluations guiding how people attribute bias to familiar others.

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

Introduction

Psychologists have documented 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 the same biases in themselves (Pronin et al., 2002). This *bias blind spot* often undermines judgment quality, 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 have been attributed to two complementary mechanisms (Pronin et al., 2004; Pronin, 2007, 2008): (1) people's belief that they objectively perceive reality (i.e., naïve realism; Griffin & Ross, 1991; Ross & Ward, 1996), and (2) self-other asymmetries in the value[1] 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 even 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 tend to value

[1] 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.

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 people's *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. Much evidence demonstrates the significant role of affect in human judgment (Lerner et al., 2015; Loewenstein et al., 2001; 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). Thus, independent of potential self-other asymmetries in individuals' cognitive strategies for assessing bias, when evaluating the biases of familiar others, readily available affective impressions may shape 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 toward a disliked group facilitate the view that members of that group are biased? The current research addresses these questions.

There are multiple plausible mechanisms by which affect may modulate attributions of bias and the bias blind spot. Consistent with an affect heuristic (Finucane et al., 2000; Slovic et al., 2007), we propose that readily available affective impressions function as heuristic cues,

allowing people to quickly and efficiently attribute bias to familiar others. When assessing bias, positive feelings toward well-liked others may be treated as evidence of objectivity, whereas negative feelings toward disliked others may be taken as diagnostic of bias. On this account, affective impressions inform judgments of bias by serving as an accessible and influential cue. Alternatively, people may unconsciously replace the complex question of "How biased is this person?" with the simpler question "How much do I like this person?", allowing the answer to this affect-based question to inform their bias assessments (Kahneman & Frederick, 2002). In this case, affective impressions do not merely serve as a readily accessible cue but replace the intended evaluation. A third possibility is that affective feelings shape bias attributions through motivated reasoning (Baumeister & Newman, 1994; Kunda, 1990). People are motivated to maintain favorable views of themselves and valued others while justifying negative views of disliked others. These motivations may guide how evidence is searched for, evaluated, and weighted. For example, when considering the biases of a well-liked other, individuals may preferentially recall information supporting their objectivity and scrutinize or discount evidence suggestive of bias. Furthermore, these motivations may lead individuals to use different cognitive strategies when evaluating the biases of well-liked versus disliked others, potentially making them more willing to consider the thoughts and intentions of liked others. Taken together, these accounts illustrate multiple pathways through which affective impressions may shape attributions of bias and modulate the bias blind spot in everyday interpersonal contexts, highlighting the potentially important role of affect in this consequential meta-bias.

The Current Research

The tendency for people to view themselves as less biased than an "average other" is well-established (Pronin & Hazel, 2023). However, less is known about how people attribute

bias to familiar others with whom they have positive or negative feelings. 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.[2] Consequently, we anticipated that participants' bias blind spots—capturing the extent to which they view themselves as less biased than a target other—would be reduced or even eliminated when randomly assigned to evaluate the biases of a well-liked individual. Thus, we hypothesized that the tendency to view oneself as uniquely unbiased would not extend to the positively-valenced relationships common in everyday social life, as individuals would extend favorable attributions of objectivity to well-liked others.

Like interpersonal contexts, intergroup conflict 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 toward one another and positive sentiments toward co-partisans (Iyengar et al., 2019) to assess the role of affect in attributions of bias within a polarized group context. Specifically, we recruited

[2] We did not preregister a directional hypothesis comparing evaluations of disliked well-known others with those of the average survey respondent for two reasons. First, we anticipated—and observed (see Table 1)—that negative affect toward disliked targets would be weaker in magnitude than positive affect toward liked targets, potentially attenuating the amplification of the bias blind spot in the Disliked Well-Known Other condition. Moreover, because unfamiliar neutral targets are already perceived as substantially more biased than the self (Mandel et al., 2022; Pronin et al., 2002), both unfamiliarity and negative affect may lead to similar conclusions—that a target other is considerably more biased than oneself. Nevertheless, consistent with the proposed role of affect in bias evaluations, we explored whether participants evaluating a disliked well-known other would exhibit larger bias blind spots than those evaluating the average survey respondent.

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 members were the focal "other." Additionally, we hypothesized that affective feelings toward 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 toward 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 collected a sample size of 660 participants for each

[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.

experiment with the goal of retaining data from at least 600 participants following pre-registered exclusions. This sample size was chosen on the basis of prior work on the bias blind spot (Mandel et al., 2022) with sensitivity power analyses using G*Power (Faul et al., 2007) revealing that this sample provided at least 80% power to detect a minimum effect size of $\eta_p^2 = .113$ for the conducted mixed ANOVAs and $d = 0.28$ for follow-up independent samples *t*-tests. 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; $M_{age} = 38.76$, $SD_{age} = 11.37$; 81% British, 7% Canadian, 3% American, 9% Other; 78% College educated). All experiments were reviewed and received ethics approval from the Defence Research and Development Canada Human Research Ethics Committee.

Measures and Materials

Well-Known Other Selection. Participants were randomly assigned to assess the biases of themselves and either the average survey respondent, a liked well-known other, or a disliked well-known other. Those assigned to evaluate the biases of a well-known other were instructed to "think about a person you know well and personally who you feel [positively/negatively] towards." After bringing a specific person to mind, participants indicated that person's relationship to them (parent, spouse, sibling, relative, friend, work colleague, or "Other"). Those who selected "Other" described this relationship in a free-entry text box. To ensure that participants evaluated the same individual throughout the bias blind spot task, they indicated their relationship with the individual they selected both before and during the task. We excluded data from participants who provided discrepant responses to relationship categorization questions. The distribution of relational categories across well-known other conditions, along with the mean familiarity and affect ratings for each category, are presented in Table 1.

Table 1

Experiment 1: Well-Known Other Selection

	Liked Well-Known Other			Disliked Well-Known Other		
Relation	***n* (%)**	**Familiarity**	**Affect**	***n* (%)**	**Familiarity**	**Affect**
Parent	26 (12%)	2.08 (0.84)	2.42 (0.81)	19 (9%)	2.05 (0.78)	-1.32 (1.70)
Spouse	87 (41%)	2.44 (0.71)	2.76 (0.46)	11 (5%)	2.09 (0.83)	1.00 (1.34)
Sibling	19 (9%)	1.74 (1.05)	2.68 (0.48)	10 (5%)	2.20 (0.92)	-1.70 (1.57)
Relative	12 (6%)	2.17 (0.83)	2.83 (0.39)	26 (12%)	1.19 (1.23)	-1.54 (1.48)
Friend	56 (27%)	1.89 (0.62)	2.38 (0.70)	37 (18%)	1.05 (1.37)	-1.03 (1.59)
Colleague	7 (3%)	1.00 (1.91)	1.86 (1.35)	83 (40%)	0.64 (1.27)	-1.59 (1.04)
Other	3 (1%)	1.67 (1.15)	2.67 (0.58)	23 (11%)	1.65 (1.27)	-2.09 (0.95)

Note. The distribution of relationship categories for selected well-known others in the Liked and Disliked Well-Known Other conditions is shown above. Familiarity was assessed with the item, "How familiar would you say you are with the thoughts, feelings, and behaviors of the person you imagined?" and was rated on a scale that ranged from -3 (Not at all familiar) to +3 (Very familiar). Affect was assessed with the item, "How do you feel about the person you know well and personally?" and was rated on a scale that ranged from -3 (Extremely negative) to +3 (Extremely positive). Means are presented with standard deviations in parentheses.

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 (see Table 2) 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*). Consistent with prior work (Scopelliti et al., 2015), this 14-item measure demonstrated good internal consistency ($\alpha = 0.82$), supporting item aggregation. We calculated a bias blind spot score for each participant by subtracting their mean Self judgment from their mean Other judgment. As such, positive scores reflect the belief that a target other exhibits the described biases more than oneself.

Table 2

Study Materials: Bias Blind Spot Task

Item	Description	Bias
1	People tend to judge a harmful action as worse than an equally harmful inaction. ...	**Action-inaction Bias**
2	People tend to do or believe a thing only because many other people believe or do that thing, to feel safer or to avoid conflict.	**Bandwagon Effect**
3	People tend to show a "disconfirmation" tendency in the way they evaluate research about dangerous habits. That is, they are more critical and skeptical in evaluating evidence that an activity is dangerous when they engage in that activity than when they do not.	**Disconfirmation Bias**
4	People tend to make irrational decisions to justify actions they have already taken. ...	**Escalation of Commitment**
5	People tend to exhibit an aversion to learning about potential losses. For example, people may try to avoid bad news by ignoring it.	**Ostrich Effect**
6	People tend to show a "self-interest" effect in the way they view political candidates. That is, people's ... judgments about the extent to which particular candidates would pursue policies good for the American people ... tend to be influenced by their feelings about whether the candidates' policies would serve their own particular interests.	**Self-Interest Bias**
7	People tend to show a "self-serving" tendency in the way they view their job performance. That is, they tend to take credit for success but deny responsibility for failure. ...	**Self-Serving Bias**
8	People tend to react to counter-evidence by strengthening their beliefs. ...	**Confirmation Bias**
9	People tend not to help in an emergency situation when other people are present. ...	**Diffusion of Responsibility**
10	People tend to make "overly dispositional inferences" in the way they view victims of assault crimes. That is, they are overly inclined to view the victim's plight as one he or she brought on by carelessness, foolishness, misbehavior, or naivetë.	**Fundamental Attribution Error**
11	People tend to show a "halo" effect in the way they form impressions of attractive people. For instance, when it comes to assessing how nice, interesting, or able someone is, people tend to judge an attractive person more positively than he or she deserves.	**Halo Effect**
12	People tend to possess an unconscious, automatic tendency to be less generous to people of a different race than to people of their race. ...	**Ingroup Favoritism**
13	People tend to underestimate the impact or the strength of another person's feelings. For example, people who have not been victims of discrimination do not really understand a victim's social suffering and the emotional effects of discrimination.	**Projection Bias**
14	Gender biases tend to lead people to associate men with technology and women with housework.	**Stereotyping**

Note. Bias Blind Spot items. Participants read descriptions of psychological biases and judged the extent to which they and a target other exhibited each bias. Some item descriptions are shortened within this table for presentation purposes. The exact wording of each item can be viewed in the Supplementary Materials (Part A). Participants viewed all 14 items in Experiment 1, but only items 1-7 in Experiments 2 and 3.

Bias Assessment Strategy. Using a methodology adapted from prior work (Pronin & Kugler, 2007),[4] we assessed the extent to which participants endorsed using 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 ("How I [think and feel/believe others think and feel/believe the person I selected thinks and feels]"), one based on behavior observation ("What I see in [my own/others'/the selected person's] behavior"), and another based on declarative knowledge ("Knowledge that I learned from an expert source").

Affect. We assessed participants' affective feelings toward 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.[5] 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

[4] Relative to prior work, methodological differences included instructing participants to select the strategy they relied on most (rather than have them rate the extent to which they used each strategy), minor wording adjustments to the introspection and behavioral observation strategy options, and the addition of a declarative knowledge option.
[5] This question regarding participants' meta-perceptions (perceptions of others' perceptions) was motivated by prior work demonstrating the importance of individuals' meta-perceptions in the political realm (Fernbach & Van Boven, 2022). Specifically, this research finds that political partisans tend to overestimate the amount of dislike rival partisans feel toward them, with negative meta-perceptions being linked with ideological extremity and anti-democratic attitudes (Lees & Cikara, 2020; Moore-Berg et al., 2020).

participants assessed the biases of the average survey respondent, a liked well-known other, or a disliked well-known other during the Other block.

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).[6] These exclusions were consistent with pre-registered criteria. After exclusions, 214 participants were randomly assigned to evaluate the biases of the average survey respondent, 210 participants a liked well-known other, and 209 participants a disliked well-known other.

Transparency and Openness

For all experiments, we collected our full sample prior to data analyses and report how we determined our sample size, 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).[7] Data and analyses scripts (i.e., annotated R scripts) pertaining to each experiment are available on Open Science Framework (https://osf.io/czvgy/). Data were analyzed using R, version 4.3.1 (R Core Team, 2023). All experiments were preregistered through Open Science Framework. These pre-registrations can be viewed via the following links (Exp. 1: https://osf.io/cpxgw/; Exp. 2: https://osf.io/d3tm5/; Exp. 3: https://osf.io/rhzqk/). We describe minor deviations from these pre-registrations in the supplementary materials (Part B).

Results and Discussion

[6] Note that one participant exhibited multiple grounds for exclusion.
[7] For the purpose of brevity, some measures are reported exclusively in the supplementary materials (Part A).

Participants demonstrated a bias blind spot, judging themselves (M = 3.56, SD = 0.90) as less biased than a target other (M = 4.31, SD = 1.03), $t(632) = 17.14$, $p < .001$, d = 0.68, 95% CI [0.59, 0.77]. This pattern was observed consistently at the item-level: participants exhibited a bias blind spot on 13 of the 14 bias blind spot items[8] (all $ps < .001$, all $ds > 0.17$). Critically, we assessed the degree to which participants' bias blind spots differed as a function of the target other they evaluated. This allowed us to test whether people view themselves as less biased not only than the hypothetical average person, but also relative to familiar individuals within their real-world social networks. 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, $F(2, 630) = 104.63$, $p < .001$, $\eta_p^2 = .249$ (see Figure 1). While we observed small differences in bias evaluations of the self across Other Reference Group conditions (all $ds < 0.33$), 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 [1.08, 1.51], and the average survey respondent (M = 4.72, SD = 0.77), $t(422) = 14.71$, $p < .001$, d = 1.43, 95% CI [1.21, 1.64].[9] 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 $d > 1.02$ in both cases). A follow-up independent samples t-test revealed that participants exhibited modestly larger bias blind spots when evaluating a disliked well-known other compared with the average survey

[8] Participants did not endorse exhibiting less *diffusion of responsibility* bias than a target other, $t(632) = 1.61$, p = .109, d = 0.06, 95% CI [-0.01, 0.14], and thus did not exhibit a bias blind spot for this item. A full breakdown of exploratory item-level analyses for each experiment is provided in Part E of the supplementary materials.

[9] 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 $p < .001$).

respondent, $t(421) = 2.58$, $p = .010$, $d = 0.25$, 95% *CI* [0.06, 0.44]. Conversely, participants did not display 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, 0.12], as they extended favorable attributions of objectivity to these individuals. Together, these results suggest that the widely documented bias blind spot, while potentially amplified in hostile relationships, does not extend to a meaningful class of social comparisons—those involving well-liked familiar others. As such, in many real-world relationships, people may not see themselves as uniquely objective, instead viewing themselves and well-liked others as similarly unbiased.

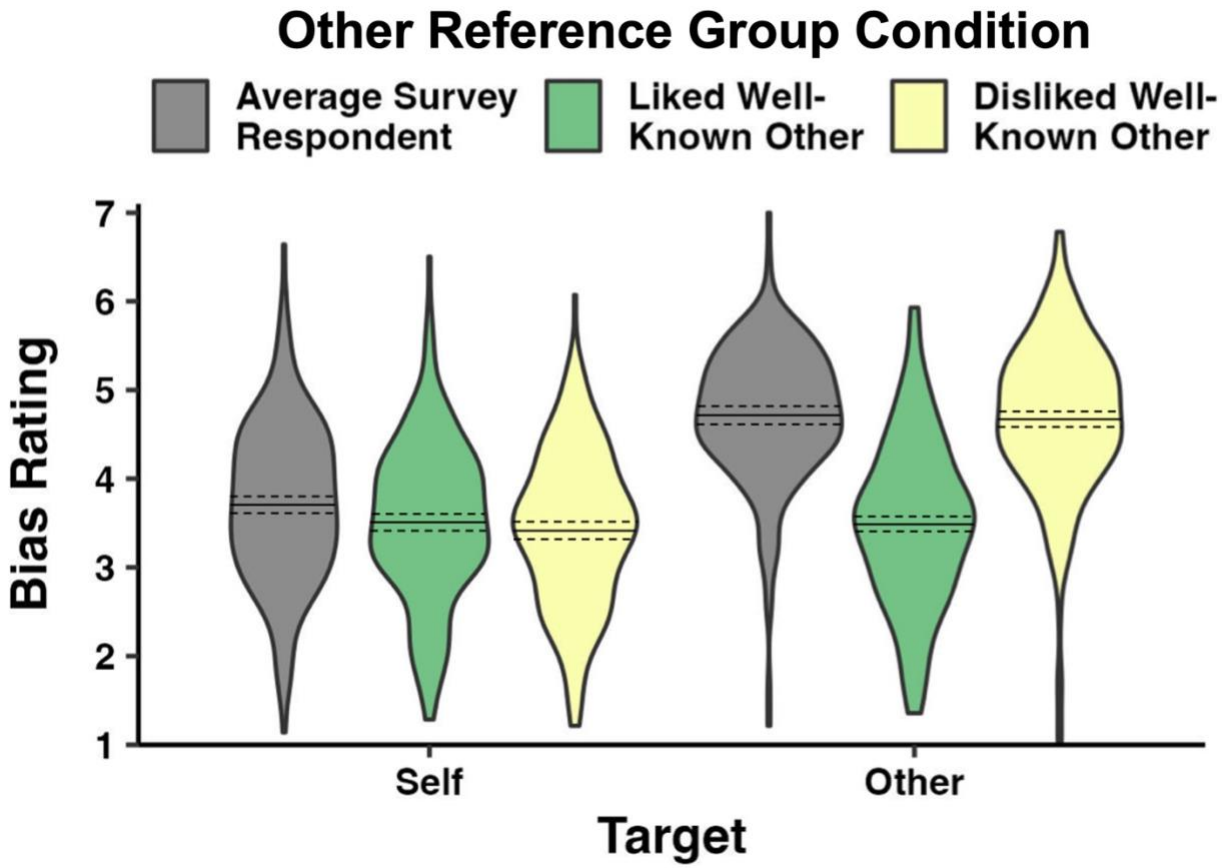

Figure 1. Experiment 1 Results: Bias Attributions. This figure displays the distribution of participants' mean bias ratings for each Target and Other Reference Group condition. The x-axis denotes the Target being evaluated (left = Self; right = Other). Colors indicate the Other Reference Group condition to which participants were randomly assigned (average survey

respondent, liked well-known other, disliked well-known other). Solid lines represent condition means for each target while dashed lines represent the corresponding 95% confidence intervals.

Affective Evaluations Are Linked to Bias Attributions

Next, we tested whether participants' affective evaluations were associated with their attributions of bias. As expected, we observed a strong negative correlation between participants' affective evaluations of a target other and the magnitude of their bias blind spot, $r(631) = -.52$, $p < .001$. Notably, this negative association was observed not only across conditions but also *within* each experimental group: Average survey respondent, $r(212) = -.23$, $p < .001$; Liked well-known other, $r(208) = -.18$, $p = .010$; Disliked well-known other, $r(207) = -.31$, $p < .001$. Regardless of the target other being evaluated, the more positively participants felt toward the target, the less bias they ascribed to them, $r(631) = -.50$, $p < .001$, and consequently, the smaller their bias blind spots. Furthermore, participants' *perceptions* of how a well-known other felt about *them* were also negatively correlated with their bias blind spot scores, $r(417) = -.46$, $p < .001$. Specifically, the more participants believed a well-known other viewed them negatively, the more bias they attributed to this familiar individual, $r(417) = -.52$, $p < .001$. Finally, we observed a small correlation between participants' feelings toward themselves and their self-rated bias, $r(631) = -.10$, $p = .013$, with those reporting more positive feelings toward themselves tending to view themselves as less biased. 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 rather than oneself that modulates the bias blind spot.

How Do Affective Feelings Influence Bias Attributions?

The results of Experiment 1 demonstrate that people do not perceive themselves as less biased than individuals who are familiar to them and whom they like. This finding, along with

the observed associations between affective feelings and attributions of bias, suggest that affective impressions shape how people perceive bias in others. However, the mechanism by which affect exerts this influence remains unclear. A prominent account of the bias blind spot—known as the *introspection illusion hypothesis*—proposes that individuals rely on introspective evidence when evaluating their own biases but focus more on observable behavior when judging the biases of others (Pronin, 2007, 2008). This self-other asymmetry in bias assessment strategies is thought to underlie, at least in part, self-other asymmetries in bias attributions. Accordingly, we examined whether participants' affective impressions of target others were related to the strategy they endorsed using to assess those individuals' biases.[10] If people are more inclined to consider the presumed thoughts and intentions of well-liked others when evaluating their susceptibility to bias, this may explain *why* affective impressions shape bias attributions.

Participants' affective evaluations of a target other were linked to the strategy they reported using to judge this target's biases. Specifically, participants reported more positive feelings toward a target when they endorsed an introspection-based strategy ($M = 0.86$, $SD = 1.73$) compared to a behavioral observation strategy ($M = 0.38$, $SD = 2.14$), $t(580) = 3.05$, $p = .002$, $d = 0.24$, 95% CI [0.08, 0.40]. However, contrary to the idea that affective impressions shape attributions of bias by shifting assessment strategies, participants were more likely to report using an introspection strategy when attributing bias to the average survey respondent (57.84%) than a liked well-known other (34.76%), $\chi^2 = 21.26$, $p < .001$—despite attributing significantly *more bias* to the former. More generally, participants' self-reported strategies were largely inconsistent with the introspection illusion account. For example, while a majority of participants (61.51%) endorsed a behavior observation strategy when assessing the biases of a

[10] Analyses examining participants' bias assessment strategies removed instances in which participants endorsed a declarative knowledge strategy as a) this strategy was endorsed by only a handful of participants (Exp. 1: $n = 12$; Exp 2: $n = 41$) and b) was not pertinent to the claims assessed here.

target other, $\chi^2 = 32.93$, $p < .001$, they did not favor an introspection strategy (Behavior: 51.03% vs. Introspection: 48.97%) when attributing bias to themselves, $\chi^2 = 0.27$, $p = .604$. Relatedly, use of an introspection strategy was not associated with attributions of bias regarding oneself or a target other, $r < .04$, $p > .320$.

Taken together, these findings suggest that affective evaluations do not shape bias attributions by altering the strategy people use to assess bias in others, with participants' endorsed strategy sharing no relation to their bias attributions. Rather, we find that participants' affective evaluations, particularly toward target others, predict the strength of their bias blind spots independent of the strategy used to assess bias (see Table 3). Therefore, the results of Experiment 1 are consistent with a novel account of the bias blind spot, in which readily available affective impressions act as heuristic cues that guide attributions of bias, leading people to perceive themselves and individuals they like as objective while viewing others, particularly disliked others, as biased.

Table 3

Affective Evaluations Predict Bias Blind Spot Scores

Predictor	b	95% CI	β	*t*	*p*
Intercept	0.72	[0.58, 0.86]	-	10.01	< .001
Affect (Self)	0.08	[0.03, 0.13]	0.11	3.32	< .001
Affect (Other)	-0.29	[-0.33, -0.25]	-0.53	-15.56	< .001
Strategy (Self)	0.14	[-0.01, 0.28]	0.06	1.84	.066
Strategy (Other)	0.13	[-0.03, 0.28]	0.06	1.64	.102

Note. $N = 621$; $df = 616$. Results from a multiple linear regression predicting participants' bias blind spot scores from their affective evaluations of the self and a target other, as well as their self-reported bias assessment strategy for each target. Strategy use was dummy-coded such that participants who reported using an introspection-based strategy were coded as 0, and those using a behavioral observation strategy were coded as 1. Model Summary: $R^2 = 0.30$, $F(4, 616) = 65.37$, $p < .001$.

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 toward in-group members and negative affect toward 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, and thus display 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 toward one's in-group to be associated with partisans viewing in-group members as less biased and stronger negative affect toward 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; 68% College educated) to be analyzed.

Measures and Materials

Bias Blind Spot. 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 D 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 2A). Item-level analyses showed significant partisan differences for 7 (out of 14) bias blind spot items.[11] Based on these findings, we selected seven bias blind spot items (Items 1-7 in Table 2) for inclusion in Experiment 2, each of which was rated as similarly "bias-like" and negative by Democrats and Republicans. Aggregated responses across these seven items revealed no significant partisan differences in bias categorization ($p = .605$, $d = 0.10$) or affect judgments ($p = .091$, $d = 0.34$; Figure 2B). Furthermore, the resulting 7-item, non-partisan bias blind spot measure showed acceptable internal consistency ($\alpha = 0.75$), supporting item aggregation.

[11] 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.

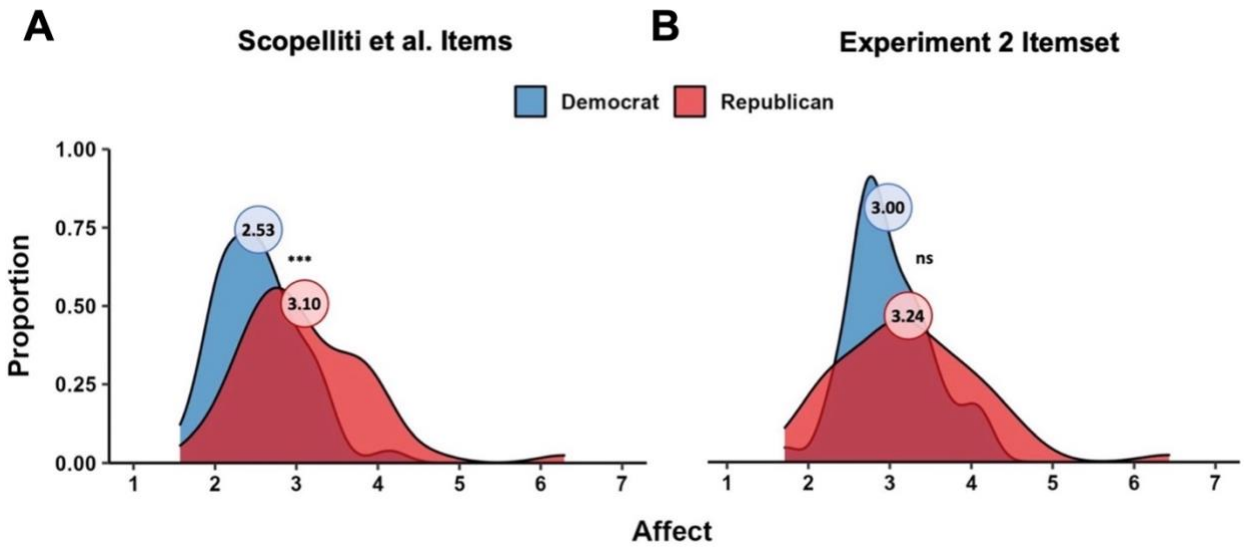

Figure 2. Item Selection Pre-Test: Affect. This figure displays the distribution of participants' (N = 100) mean affective judgment of 14 psychological biases described within bias blind spot items adapted from Scopelliti and colleagues (2015; Panel A). Participants' responses within this pilot study 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 panel B. 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*).

Political Identity Strength. We measured the strength of participants' identification with their preferred political party using a 5-item scale adapted from Leach and colleagues (2008). Participants judged how much they agreed with statements probing their identification with their political party (e.g., "I feel a bond with [Democrats/Republicans]") using a 7-point scale that ranged from 1 (*Strongly Disagree*) to 7 (*Strongly Agree).* Scale items showed excellent internal consistency ($\alpha = .93$) and thus were aggregated to form a composite measure of political identity strength.

Design 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 assessment strategy and affect-related questions, as in Experiment 1. 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. Unlike items in the bias blind spot task, these items asked participants to attribute bias to themselves and the average partisan *without* reference to specific psychological biases, and thus measured participants' bias attributions independent of the psychological biases selected.

Data Preparation

We excluded data from 34 participants who indicated that English was not their first language (n = 4), self-reported an age 61 years or older (n = 1), 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. Following exclusions, 309 participants were randomly assigned to evaluate the biases of the average Democrat, and 317 were assigned to evaluate the biases of the average Republican. This assignment resulted in 306 participants evaluating the biases of a typical in-group member and 320 evaluating the biases of a typical out-group member.

Results and Discussion

Participants exhibited a bias blind spot, judging themselves (M = 3.66, SD = 1.09) as less biased than a target other (M = 4.67, SD = 1.07), $t(625) = 19.06$, $p < .001$, $d = 0.76$, 95% CI [0.67, 0.85]. This effect generalized across biases, with participants rating themselves as less

biased than a target other on all seven bias blind spot items (all *ps* < .001, all *ds* > 0.29). 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. This analysis revealed 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* [0.94, 1.28].[12] 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 ($M = 1.59$, $SD = 1.38$), $t(624) = 12.31$, $p < .001$, $d = 0.98$, 95% *CI* [0.82, 1.15]. There was no main effect of Participant Partisanship, $F(1, 622) = 2.65$, $p = .104$, $\eta_p^2 = .004$, 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 ($p < .001$, $d = 0.41$), they displayed a larger bias blind spot when attributing bias to their political opponents ($p < .001$, $d = 0.75$).

[12] 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 $p < .001$).

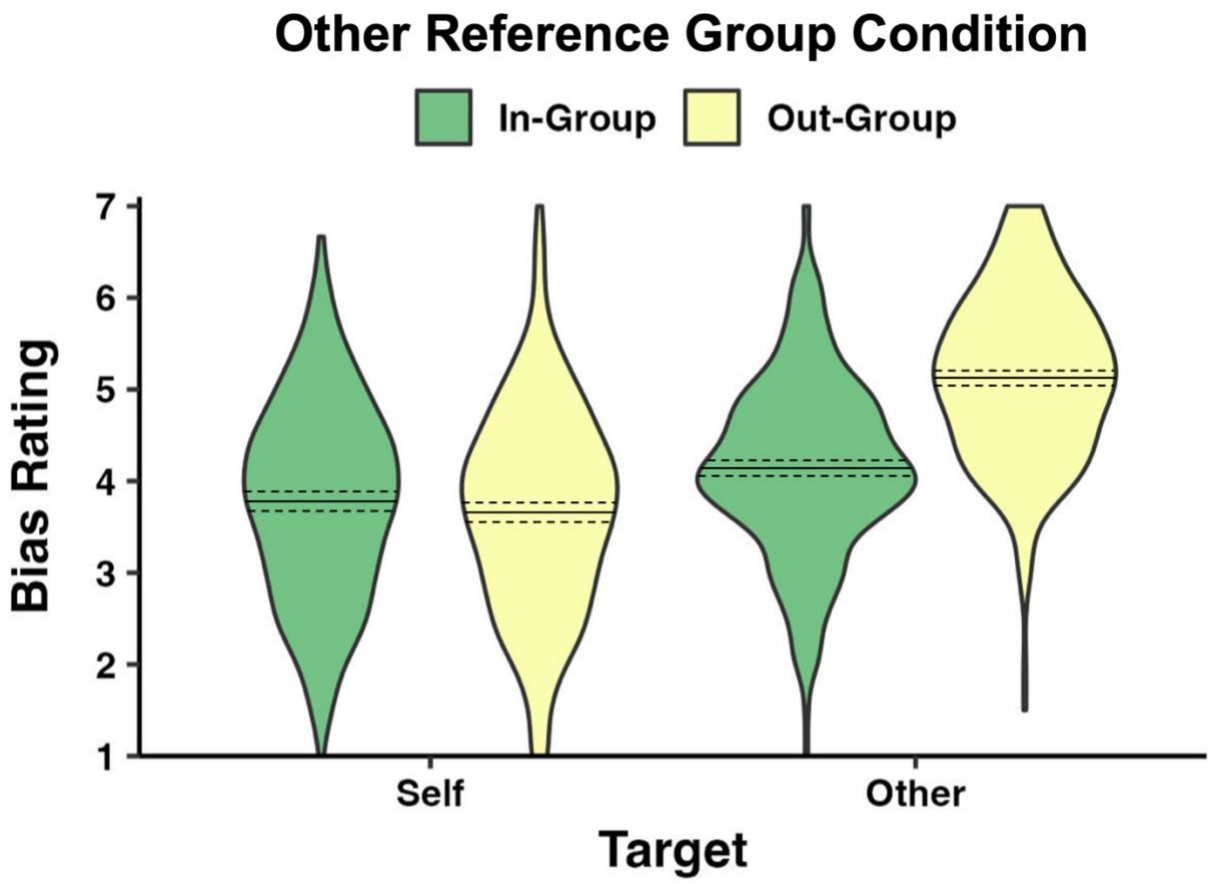

Figure 3. Experiment 2 Results: Bias Attributions. This figure displays the distribution of participants' mean bias ratings for each Target and Other Reference Group condition. The x-axis denotes the Target being evaluated (left = Self; right = Other). Colors indicate the randomly assigned Other Reference Group condition (in-group member vs. out-group member). Solid lines represent condition means for each target while dashed lines represent the corresponding 95% confidence intervals.

Affective Evaluations Are Linked to Bias Attributions Within and Across Groups

Next, we examined whether participants' attributions of bias were sensitive to their affective evaluations. Replicating the results of Experiment 1, negative feelings toward a target other were associated with viewing this target other as more biased, $r(624) = -.55$, $p < .001$, and consequently, larger bias blind spots, $r(624) = -.47$, $p < .001$. Importantly, these negative associations were observed not just across but also *within* each experimental condition. Participants' affective evaluations of a target other were negatively correlated with the

magnitude of their bias blind spot regardless of whether the target other they evaluated was a political in-group, $r(304) = -.30$, $p < .001$, or out-group member, $r(318) = -.23$, $p < .001$. Given the robust association between affective evaluations and bias blind spot scores, we conducted mediation analyses to test whether affective judgments accounted for differences in participants' bias blind spots across conditions. These analyses revealed that the effect of Other Reference Group on bias blind spot scores was partially mediated by participants' affective evaluations of target others ($b = 0.62$, 95% *CI* [0.40, 0.80]; see Supplementary Materials Part C), suggesting that heightened attributions of bias toward political out-group members were driven, in part, by participants' affective feelings toward each party. Taken together, these findings suggest that while affect contributes to intergroup differences in bias attribution (consistent with in-group favoritism), it also shapes judgments within group boundaries. Thus, these results are consistent with affective impressions serving as a general heuristic cue guiding bias attributions both across and within groups.

We also examined the relationship between participants' other affective judgments and their bias blind spot scores. Consistent with Experiment 1, participants' *perceptions* of how a target other felt about *them* was negatively correlated with their bias blind spot, $r(624) = -.47$, $p < .001$, such that the more participants believed a target other viewed them negatively the more they tended to judge this target other as biased, $r(624) = -.52$, $p < .001$. Additionally, participants who reported more positive feelings toward themselves 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, $z = -8.21$, $p < .001$, again suggesting that it is affective feelings toward others, not oneself, that modulates the bias blind spot.

Strongly Identified Partisans Attribute Less Bias to Their Political In-Group

The design of Experiment 2 permitted a test of whether participants' level of identification with their preferred political party was related to their bias blind spot score. Among participants assigned to evaluate the biases of the average out-group member, political identity strength was not correlated with bias blind spot scores, $r(318) = .09$, $p = .111$. However, among those judging the biases of the average in-group member, stronger identification with one's political in-group was associated with viewing the average in-group member as less biased, $r(304) = -.15$, $p = .009$, and relatedly, with *smaller* bias blind spots, $r(304) = -.30$, $p < .001$. This association remained when controlling for affective evaluations of the average in-group member, $b = -0.14$, 95% *CI* [-0.25, -0.02], $t(303) = -2.29$, $p = .023$, itself a significant predictor of bias blind spot scores, $b = -0.18$, 95% *CI* [-0.33, -0.02], $t(303) = -2.28$, $p = .024$. In sum, rather than strongly identified partisans attributing more bias to political opponents and exhibiting larger bias blind spots, political identity strength was linked to viewing in-group members as less biased—mirroring how individuals viewed themselves. This effect was not simply due to strong partisans' more positive feelings toward their in-group, as political identity strength continued to predict bias blind spot scores after controlling for affective evaluations.

Affect Predicts Bias Attributions Beyond Its Influence on Bias Assessment Strategies

We next examined whether participants' affective evaluations of target others were associated with the strategies they reported using to assess those targets' biases. A greater proportion of participants endorsed an introspection strategy when evaluating the biases of the average in-group member (62.20%) compared to out-group member (38.46%), $\chi^2 = 33.00$, $p < .001$. Moreover, participants reported more positive feelings toward target others' when endorsing an introspection strategy ($M = 0.53$, $SD = 1.62$) as opposed to one focused on a

target's behavior (M = -0.33, SD = 1.81), $t(594) = 6.18$, $p < .001$, $d = 0.50$, 95% CI [0.34, 0.67]. Together, these results suggest that affective evaluations are tied to individuals' bias assessment strategies, with more positive feelings toward a group corresponding to greater reliance on the presumed internal states—rather than observable behavior—of its members.

Next, we assessed whether affective impressions shaped attributions of bias by shifting participants' bias assessment strategies—specifically, whether positive feelings increased reliance on introspection, thereby reducing attributions of bias, consistent with the introspection illusion account. Participants' self-reported strategies again undermined key predictions of this account. For example, participants favored a behavioral observation strategy (65.95%) when assessing their own biases, $\chi^2 = 61.24$, $p < .001$. Moreover, they did not favor a behavioral observation strategy when assessing the biases of a target other (Behavior: 49.91% vs. Introspection: 50.09%), $\chi^2 < 0.01$, $p = .968$. Finally, there was no association between participants' self-directed bias assessment strategies and bias attributions, $r(600) < .01$, $p = .935$. However, endorsing an introspection strategy was associated with lower attributions of bias when ascribing bias to others, $r(601) = -.16$, $p < .001$. Endorsing an introspection strategy when evaluating others' biases was also associated with smaller bias blind spot scores—even after controlling for participants' affective evaluations (see Table 4). Thus, these results provide some evidence that people rely more heavily on introspection-based strategies when evaluating well-liked others and that such strategies are linked to viewing others as less biased. Nevertheless, affective feelings, particularly those toward target others, remained the strongest and most consistent predictor of participants' bias blind spots, suggesting that affective impressions primarily shape attributions of bias independent of their influence on bias assessment strategies.

Table 4

Affective Evaluations Predict Bias Blind Spot Scores in a Polarized Group Context

Predictor	b	95% CI	β	*t*	*p*
Intercept	0.81	[0.61, 1.00]	-	8.04	< .001
Affect (Self)	0.12	[0.06, 0.19]	0.14	3.92	< .001
Affect (Other)	-0.35	[-0.41, -0.30]	-0.47	-12.57	< .001
Strategy (Self)	0.03	[-0.17, 0.23]	0.01	0.31	.754
Strategy (Other)	0.27	[0.07, 0.46]	0.10	2.70	.007

Note. N = 585; df = 580. Results from a multiple linear regression predicting participants' bias blind spot scores from their affective evaluations of the self and a target other, as well as their self-reported bias assessment strategy for each target. Strategy use was dummy-coded such that participants who reported using an introspection-based strategy were coded as 0, and those using a behavioral observation strategy were coded as 1. Model Summary: R^2 = 0.26, $F(4, 580)$ = 49.99, $p < .001$.

Experiment 3

Experiment 3 provided a conceptual replication of Experiment 2, with a primary modification being that participants evaluated the biases of both the average Democrat and Republican. Therefore, Experiment 3 allowed us to assess whether the observed tendency for individuals to view themselves and liked-others (political in-group members) as less biased than more disliked-others (rival partisans) would persist in a within-subjects design in which potential comparisons between the average in-group and out-group member were more salient.

Method

Participants

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; 64% College educated).

Measures and Materials

All measures and materials described in Experiment 2 were featured in Experiment 3, with two exceptions. First, Experiment 3 did not include bias assessment strategy items. Second, Experiment 3 used a modified measure of political identity strength: one item was removed and four new items were added. These new items aimed to capture the extent to which *not* belonging to one's political out-group was an important part of one's identity (e.g., "The fact that I am not a [Republican/Democrat] is an important part of my identity"). The resulting 8-item scale demonstrated good internal consistency ($\alpha = 0.87$) and was aggregated to form a composite measure of political identity strength. Finally, the 7-item non-partisan bias blind spot measure used in Experiment 2 again showed acceptable internal consistency ($\alpha = 0.70$) in Experiment 3.

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 indicated that English was not their first language ($n = 1$), self-reported an age 61 years or older ($n = 1$), 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 and Discussion

Participants displayed a bias blind spot, judging themselves (M = 3.66, SD = 1.07) as less biased than target others (M = 4.62, SD = 0.70), $t(635) = 22.97$, $p < .001$, d = 0.91, 95% CI [0.82, 1.00]. This effect, once again, generalized across biases, with participants rating themselves as less biased than target others on all seven bias blind spot items (all $ps < .001$, all $ds > 0.31$). Critically, 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 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, $F(1,634) = 415.88$, $p < .001$, $\eta_p^2 = .396$. Participants exhibited a bias blind spot ($p < .001$) when evaluating the biases of the average in-group member (M = 0.43, SD = 1.00), but 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 [0.70, 0.88].[13] Thus, the tendency to view oneself and political in-group members as less biased than rival partisans persisted in a within-subjects design where direct comparisons between in- and out-group members were made salient.

This analysis also revealed 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), reflecting their greater attribution of bias to political out-group members. That is, Democrats and Republicans did not differ in the extent to which they attributed bias to themselves ($p = .338$) or the average in-group member ($p = .149$). However, Democrats attributed significantly more bias to the average out-group member (M =

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

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* [0.35, 0.67]. As a result, Democrats displayed larger bias blind spots than Republicans when their self-attributions of bias were compared to their attributions of bias toward the average out-group ($p < .001$, $d = 0.38$), but not in-group member ($p = .651$, $d = 0.04$).

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 ($p < .001$, $d = 0.74$). Moreover, Democrats attributed more bias to the average out-group member than Republicans, $t(634) = 3.93$, $p < .001$, $d = 0.31$, 95% *CI* [0.16, 0.47].

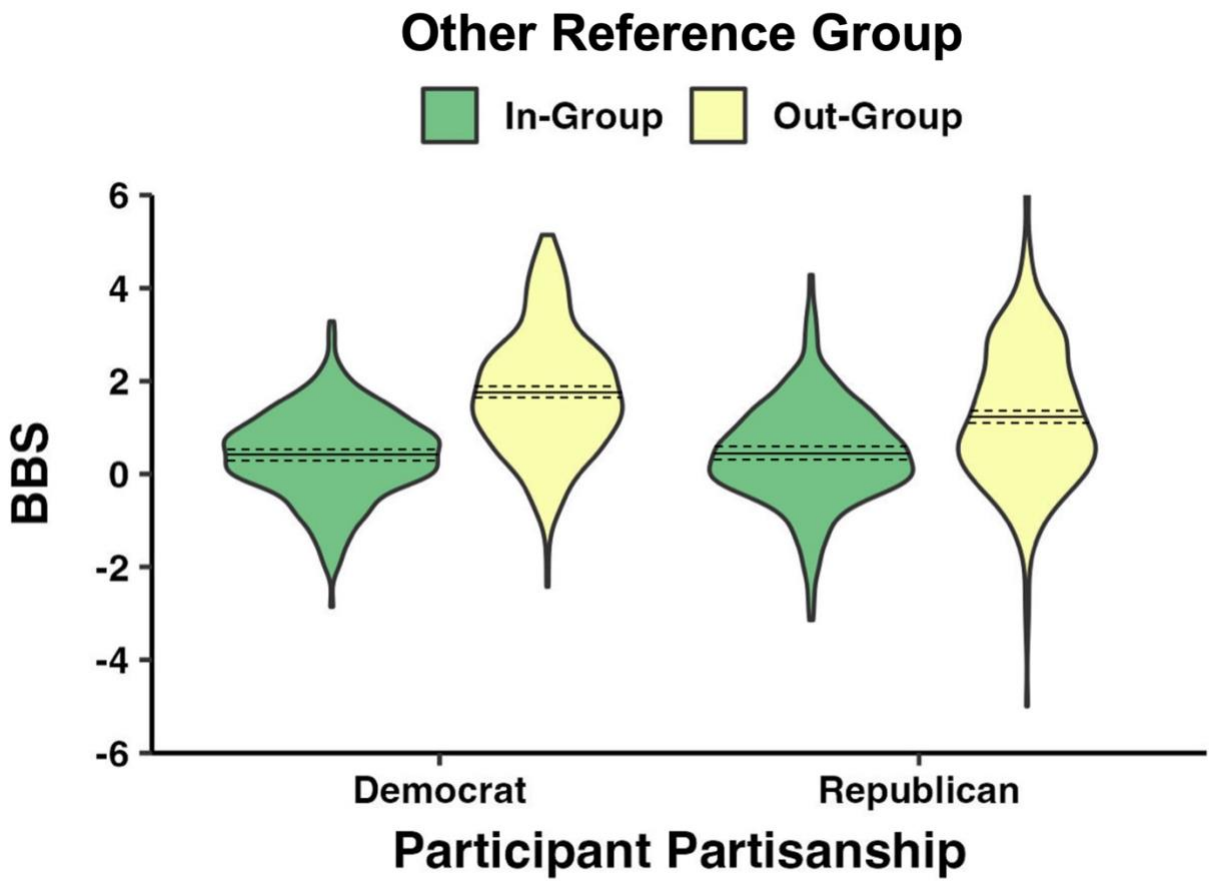

Figure 4. Experiment 3 Results: Bias Blind Spots. This figure displays the distribution of participants' bias blind spots (other bias rating – self bias rating) as a function of Participant Partisanship and Other Reference Group. The x-axis denotes participants' self-reported political affiliation (left = Democrat; right = Republican). Colors indicate the target other evaluated (in-group member vs. out-group member). Solid lines reflect the mean bias blind spot score for each grouping while dashed lines represent the corresponding 95% confidence interval.

Affective Evaluations Are Linked to Bias Attributions Within and Across Groups

Next, we examined whether participants' affective evaluations were associated with their attributions of bias. Reporting more positive feelings toward target others was correlated with judging those target others as less biased (in-group member: $r(634) = -.35$, $p < .001$; out-group member: $r(634) = -.41$, $p < .001$), and consequently, with smaller bias blind spots (in-group member: $r(634) = -.29$, $p < .001$; out-group member: $r(634) = -.32$, $p < .001$). Thus, as in Experiments 1 and 2, affective evaluations showed robust associations with attributions of bias *within* each target other category. We then conducted a mediation analysis to test whether affective evaluations accounted for differences in bias blind spot scores across target other categories. This analysis revealed that affective evaluations of target others fully mediated the effect of Other Reference Group on participants' bias blind spot scores ($b = -0.98$, 95% *CI* [-1.14, -0.82]; see Supplementary Materials Part C), suggesting that polarized attributions of bias toward in-group and out-group members were driven by corresponding differences in affective feelings toward these groups. Together, these findings provide further support for the significant role of affect in shaping attributions of bias both *between* and *within* group boundaries. Finally, we examined the association between participants' other affective judgments and their bias blind spot scores. Replicating the results of Experiments 1 and 2, the *perception* that target others possessed negative feelings toward *oneself* was associated with judging target others as more biased (in-group member: $r(634) = -.33$, $p < .001$, out-group member: $r(634) = -.40$, $p < .001$) and larger bias blind spots (in-group member: $r(634) = -.22$, $p < .001$, out-group member: $r(634)$

= -.31, $p < .001$). 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, $r(634) = -.07$, $p = .069$.

Strongly Identified Partisans Attribute Less Bias to Their Political In-Group

Lastly, we examined whether individual differences in political identity strength were associated with participants' bias blind spot scores. Replicating the results of Experiment 2, stronger identification with one's preferred political party was associated with attributing less bias to the average in-group member, $r(634) = -.25$, $p < .001$, and with smaller bias blind spots when the average in-group member served as the focal other, $r(634) = -.24$, $p < .001$. This association with bias blind spot scores remained when controlling for affective evaluations of the average in-group member, $b = -0.08$, 95% *CI* [-0.16, -0.01], $t(633) = -2.08$, $p = .038$, itself a strong predictor of participants' bias blind spots, $b = -0.23$, 95% *CI* [-0.32, -0.13], $t(633) = -4.74$, $p < .001$. Therefore, strong partisans again exhibited a tendency to view the average in-group member as less susceptible to a host of psychological biases, displaying smaller bias blind spots within an in-group context. Additionally, the relation between political identity strength and in-group bias attribution was not explained solely by strong partisans' more positive feelings toward their in-group, suggesting that other facets of a strong partisan identity contribute to ascribing less bias to co-partisans.

Unlike in Experiment 2, possessing a stronger partisan identity was associated with attributing more bias to the average out-group member, $r(634) = .26$, $p < .001$, and with larger bias blind spots when the average out-group member was the focal other, $r(634) = .18$, $p < .001$. However, this association with bias blind spot scores was no longer significant after controlling for affective evaluations of the average out-group member, $b = 0.05$, 95% *CI* [-0.05, 0.15], $t(633)$

= 1.06, p = .289, which again was a strong predictor of bias blind spot scores, b = -0.33, 95% *CI* [-0.42, -0.24], $t(633) = -7.02$, $p < .001$. Thus, strongly identified partisans' tendency to attribute more bias to political opponents appeared to be driven by their negative feelings toward the out-group.

General Discussion

Across three experiments, we find that people's tendency to attribute bias to others while denying it in themselves is shaped by 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. It was not participants' access to individuating information that eliminated their bias blind spot, as they viewed a familiar but disliked individual as considerably more biased than themselves. Rather, attributions of bias were closely linked to affective evaluations of others—both *across* and *within* target other conditions. The more positively participants felt toward a target other, the less bias they ascribed to them. By comparing individuals' self-evaluations of bias with their evaluations of familiar others, the current study demonstrates that the bias blind spot—a well-documented meta-bias considered to be a pervasive feature of social cognition—may rarely extend to the positively-valenced relationships that characterize much of social life. Thus, in many interpersonal contexts, people may extend favorable attributions of objectivity to others and, as a result, seldom perceive themselves as less biased than those they know and interact with.

Failure to attribute bias to individuals we like, while readily attributing bias to those we dislike has important theoretical and practical implications. 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, 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 negative affect and consequently, greater attributions of bias. Such negative feedback loops may be most pertinent in group contexts characterized by dislike and disagreement, such as political contexts featuring competing groups with distinct values.

Much work has documented the rise of affective 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 may amplify attributions of bias across political divides which, as we have discussed, can escalate conflict and lead to even greater partisan animus in the future. Rising partisan animosity, insofar as it increases the extent to which partisans view their political opponents as biased, may also influence how people consume political information and participate in politics. For example, perceiving political out-group members as biased can lead partisans to self-select into political echo chambers and disengage from political discussions that feature individuals with opposing views. Similarly, increased perceptions of bias across political lines may diminish individuals' trust in bipartisan institutions and lead people to disengage from democratic processes. While secondary analyses provide preliminary support for these claims,[14] future studies should

[14] In Experiment 3, attributing more bias to the average political out-group member was associated with a stronger preference for ideologically congruent media, $r(634) = .24$, $p < .001$ (see Supplementary Materials Part E).

investigate the extent to which affective evaluations of political in-group and out-group members shape how citizens consume political information and engage in democratic processes.

Experiments 2 and 3 examined the link between affect and attributions of bias in an intergroup context. We show that partisan Americans view the average member of their political out-group as far more susceptible to a host of psychological biases than themselves and the average member of their political in-group. Furthermore, we again observed a strong relationship between affective evaluations and attributions of bias. The more positively participants felt toward their political in-group, the less bias they attributed to the average in-group member. Similarly, more negative feelings toward one's political out-group predicted greater attributions of bias toward opposing partisans. Even *perceptions* of affect were associated 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. Thus, even when lacking negative affect toward rival partisans, heightened perceptions of partisan animosity—even if exaggerated[15]—may lead partisans to more readily impute bias to their political opponents, exacerbating partisan conflicts.

Research on affective polarization shows that negative feelings toward political out-groups shape individuals' political attitudes and behaviors (Abramowitz & Webster, 2016; Finkel et al., 2020; Kingzette et al., 2021). The present research contributes to this literature by linking the polarized feelings of American partisans to non-political judgments of political in-group and out-group members. That is, partisans' affective evaluations of the average Democrat and Republican predicted the extent to which they saw members of each group as prone to a set of domain-general psychological biases, most of which were not political in nature. Thus,

[15] 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 E).

partisans who harbor strong animosity toward their political opponents may not only distrust their judgments in political contexts (e.g., their assessment of political events) but also question their objectivity more broadly. These results are consistent with the idea that affective impressions serve as a domain-general heuristic cue guiding attributions of bias. By highlighting the central role of "hot" affective evaluations over "cold" cognitive strategies, our results also contribute to ongoing debates about the origins of partisan bias (Celniker & Ditto, 2024; Druckman & McGrath, 2019). Although our experiments did not assess the evidentiary basis for partisans' bias attributions, the robust association between affective evaluations and judgments of bias suggests that partisans' polarized affective feelings played a central role in their tendency to view in-group members as objective and out-group members as biased.

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 participants endorsed a behavioral observation strategy when assessing the biases of a target other in Experiment 1, this preference was not observed 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 infrequently observed and (b) bias assessment strategies were not reliably associated with attributions of bias.

Participants' affective evaluations of familiar others predicted 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. Similarly, in a polarized group context, affective impressions of the "average" Democrat and Republican mediated 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 central role in human judgment (Lerner et al., 2015; Loewenstein et al., 2001; Slovic et al., 2007). The present findings are consistent with people using readily available affective impressions as heuristic cues when assessing the biases of others. Thus, independent of individuals' strategies for assessing bias, affective impressions may function as readily accessible signals of objectivity or bias, leading people to attribute objectivity to those they like and bias to those they dislike. This account emphasizes the primacy of "hot" affective evaluations over "cold" cognitive strategies and helps explain how participants attributed bias to familiar others in the present study, including the observed elimination of the bias blind spot when participants evaluated the biases of themselves and a well-liked other. When judging the biases of familiar others who elicit affective reactions, these reactions may come to dominate assessments of bias either by shaping downstream cognitive processes (e.g., how evidence is searched for, evaluated, and weighted) or by acting as heuristic cues that allow people to quickly

and efficiently attribute bias to familiar others without engaging in more effortful evaluative processes.

Implications for Debiasing the Bias Blind Spot

Viewing oneself as objective while perceiving bias in others has been linked with poor judgment and negative interpersonal interactions (Kennedy & Pronin, 2008; Scopelliti et al., 2015). The present research has implications for interventions aimed at mitigating this important meta-bias, particularly in affect-rich interpersonal and intergroup contexts. Our findings suggest that fostering more positive affective evaluations of disliked individuals can reduce the extent to which people see themselves as less biased than these disliked others. Accordingly, interventions that target individuals' "hot" affective reactions—as opposed to their "cold" cognitive strategies—may help attenuate the large bias blind spots that characterize social contexts marked by animosity. By softening negative affect, such approaches may reduce the interpersonal strain and conflict-escalating actions associated with seeing oneself as uniquely objective. At the same time, affect-based interventions are unlikely to help people recognize their own (or well-liked others') susceptibility to bias. For this aim, cognitively oriented interventions encouraging individuals to question their objectivity may be better suited to helping them recognize and correct the psychological biases that shape their decision-making.

Study Limitations

We find that peoples' affective evaluations of familiar others guide how they attribute bias to these individuals, and consequently, the extent to which they view themselves as less biased than others. This finding was supported across distinct contexts (interpersonal and intergroup), study designs (within-subjects and between-subjects), and biases. Nevertheless, the present work is not without limitations. First, although Experiment 1 randomly assigned some

participants to evaluate the biases of a relatively neutral target (i.e., the average survey respondent), Experiments 2 and 3 did not include an analogous control condition. Consequently, we were unable to determine whether the level of bias attributed to political in-group and out-group members exceeded, matched, or fell below baseline evaluations of the average person. Future research could incorporate both partisan and neutral targets within the same design to directly compare bias attributions toward political partisans with baseline perceptions of the average person. Second, the current study manipulated affect by randomly assigning participants to a) assess the biases of a liked or disliked well-known other and b) assess the biases of the average political in-group and out-group member. This methodology was high in ecological validity, taking advantage of participants' naturally occurring affective impressions. However, target others may have differed not only with regards to participants' affective evaluations, but in additional ways that influenced bias attributions. While we find that affective evaluations *within* each target category share a negative relation with attributions of bias, future work may seek to manipulate affective impressions in a way that eliminates the potential influence of extraneous variables (e.g., disagreement) on bias attributions.

Constraints on Generality

While Experiment 1 primarily included non-American participants, Experiments 2 and 3 leveraged the polarized affective feelings of Democrats and Republicans to investigate how affective evaluations influence attributions of bias in an intergroup context. While this context—marked by strong positive affect toward political in-group members and negative affect toward out-group members—provided a compelling testbed for our hypotheses, its U.S.-centric focus may raise questions about the generalizability of our findings to other intergroup contexts and political systems. However, consistent with the findings of Experiment 1, prior research has

documented individuals' bias blind spots (Chandrashekar et al., 2021; Niszczota et al., 2023) and reliance on an affect heuristic (Finucane et al., 2000; Keller et al., 2006) in non-U.S. populations. Moreover, affective polarization is not unique to the United States; comparable affectively charged intergroup divisions have been observed globally, including in multi-party systems (Reiljan, 2020; Wagner, 2021). On this basis, we expect our findings—particularly the role of affect in bias attribution—to generalize across intergroup contexts characterized by polarized affective evaluations of in-group and out-group members. Additionally, consistent with the results of the current 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 D) provides one example of this practice. 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 familiar 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 relatively objective while viewing disliked individuals as considerably more biased. These findings have theoretical implications for current theories of bias attribution, highlighting the primacy of "hot" affective

evaluations over “cold” cognitive strategies. They 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

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of US elections in the 21st century. *Electoral Studies*, *41*, 12-22. https://doi.org/10.1016/j.electstud.2015.11.001

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

Celniker, J. B., & Ditto, P. H. (2024). Of preferences and priors: Motivated reasoning in partisans' evaluations of scientific evidence. *Journal of Personality and Social Psychology*, *127*(5), 986-1011. https://doi.org/10.1037/pspa0000417c

Chandrashekar, S. P., Yeung, S. K., Yau, K. C., Cheung, C. Y., Agarwal, T. K., Wong, C. Y. J., ... & Feldman, G. (2021). Agency and self-other asymmetries in perceived bias and shortcomings: Replications of the Bias Blind Spot and link to free will beliefs. *Judgment and Decision Making*, *16*(6), 1392-1412. https://doi.org/10.1017/S1930297500008470

Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, *9*(2), 111-119. https://doi.org/10.1038/s41558-018-0360-1

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

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

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1-17. https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

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

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Keller, C., Siegrist, M., & Gutscher, H. (2006). The role of the affect and availability heuristics in risk communication. *Risk Analysis*, *26*(3), 631-639. https://doi.org/10.1111/j.1539-6924.2006.00773.x

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

Kingzette, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). How affective polarization undermines support for democratic norms. *Public Opinion Quarterly*, *85*(2), 663-677. https://doi.org/10.1093/poq/nfab029

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498.

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., ... & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*(1), 144-165. https://doi.org/10.1037/0022-3514.95.1.144

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, *4*(3), 279-286. https://doi.org/10.1038/s41562-019-0766-4

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

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267-286. https://doi.org/10.1037/0033-2909.127.2.267

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

Moore-Berg, S. L., Ankori-Karlinsky, L. O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences*, *117*(26), 14864-14872. https://doi.org/10.1073/pnas.2001263117

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

Niszczota, P., Pawlak, M., & Białek, M. (2023). Bilinguals are less susceptible to the bias blind spot in their second language. *International Journal of Bilingualism*, *27*(5), 569-585. https://doi.org/10.1177/13670069221110383

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

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reiljan, A. (2020). 'Fear and loathing across party lines'(also) in Europe: Affective polarisation in European party systems. *European Journal of Political Research*, *59*(2), 376-396. https://doi.org/10.1111/1475-6765.12351

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

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, *69*, 102199. https://doi.org/10.1016/j.electstud.2020.102199

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