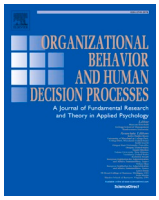
Organizational Behavior and Human Decision Processes 170 (2022) 104147

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

journal homepage: www.elsevier.com/locate/obhdp



You can't handle the truth! Conflict counterparts over-estimate each other's feelings of self-threat☆

Charles A. Dorison [a,*], Julia A. Minson [b]

[a] Kellogg School of Management, Northwestern University, United States
[b] Harvard Kennedy School, Harvard University, United States

A B S T R A C T

Attitude conflict—interpersonal disagreement on deeply-held, identity relevant issues—is common in personal, professional, and policy settings. Understanding one's counterpart is critical to success in such contexts. Although prior literature focuses on misperceptions of counterpart *cognitions*, people often rely on *affect* to explain others' behavior. Here, we examine the accuracy of individuals' assessments of others' affective states during attitude conflict. Specifically, we examine one affective state that has been theorized to play a central role in such situations: self-threat (i.e., threat to the integrity of an individual's self-concept). In four pre-registered studies (N = 1,707), individuals systematically over-estimated the levels of threat reported by conflict counterparts, which in turn increased confidence in persuasion. The effect was mediated by "naïve realism," an excessive faith in the objectivity of one's views. The present studies document a novel barrier to effective communication and extend our understanding of how affect drives behavior during attitude conflict.

1. Introduction

Attitude conflict—interpersonal disagreement on deeply-held, identity relevant issues (Judd, 1978; Minson et al., 2019; Minson & Dorison, 2021)—is common in personal, professional, and policy settings. Managing such conflict is, in turn, foundational to successful relationships, organizational performance, and even democratic functioning. Research on conflict management has thus received decades of sustained attention across social science fields (Cronin & Weingart, 2007; De Dreu & Weingart, 2003; Ross, 1993; Thomas, 1992; Tjosvold et al., 2014; Van Kleef & Côté, 2018).

An important insight emerging from this body of work is the central role that understanding one's counterpart plays in navigating interpersonal conflict (e.g., Bruneau & Saxe, 2012; Ickes, 1993; Galinsky & Moskowitz, 2000; Galinsky & Mussweiler, 2001; Galinsky et al., 2008; Neale & Bazerman, 1983). Yet, a large literature has also shown that people systematically fail at this task, misjudging other's motives, intentions, evaluations, and situational construals (e.g., Epley et al., 2006; Epley, Keysar, et al., 2004; Epley et al., 2004).

Although the prior literature largely focuses on misperceptions of counterpart *cognitions*, in their attempts to explain the behavior of others, people often invoke *affective* explanations. People learn about the social world in part through interpreting each other's emotions (Van Kleef, 2009; Van Kleef et al., 2010). In conflict specifically, emotions carry information about the counterparts' reactions and behavioral intentions (e.g., Van Kleef et al., 2004; Van Kleef & Cote, 2007). For example, when a counterpart appears angry, we might expect them to walk away from the negotiating table unless concessions are made. In this manner, inferences about a counterpart's affect suggest specific approaches to continuing the interaction.

In the present manuscript, we examine the accuracy of individuals' assessments of others' affective states (i.e., affective perspective taking) during attitude conflict. We focus on one negative affective state in particular that has been theorized by prior research to be central in such situations: self-threat (i.e., threat to the integrity of an individual's self-concept; Steele & Liu, 1981, 1983) because of the pivotal role attributed to this phenomenon by prior research. Multiple scholars have argued that exposure to contradictory ideas—an inevitable feature of disagreement—is aversive specifically *because* of the feelings of self-threat that such experiences produce (e.g., Collins et al., 2017; Festinger, 1957; Frimer et al., 2017; Matz & Wood, 2005; Nam et al., 2013; Steele & Liu, 1981, 1983; Webb et al., 2013; for review, see Hart et al., 2009). This theorizing has been supported by studies showing that affirming the self-concept increases willingness to engage with opposing views, by reducing self-threat (Badea & Sherman, 2019; Binning et al., 2010; Cohen & Sherman, 2014; Sherman et al., 2017; Sherman et al., 2020). Importantly, however, prior research has not measured the experience of self-threat or tested the accuracy of threat perceptions. Instead, the

☆ The authors appreciate feedback from members of Minson Lab. Both authors contributed equally to this work.
* Corresponding author.
E-mail address: charles.dorison@kellogg.northwestern.edu (C.A. Dorison).

https://doi.org/10.1016/j.obhdp.2022.104147
Received 1 August 2019; Received in revised form 9 February 2022; Accepted 5 March 2022

presence of self-threat is inferred from changes in behavior when affirmation manipulations are deployed.

We present four primary studies (and three supplementary studies) testing a self-other difference in perceived versus self-reported feelings of threat during attitude conflict. Our work makes three contributions. First, we contribute to the literature on perspective taking by assessing a key *affective* (rather than cognitive) mis-prediction during attitude conflict. Second, we contribute to the literature on self-threat by testing whether individuals systematically over-estimate the level of self-threat that counterparts feel. Finally, we examine a novel barrier to conflict resolution.

In the following sections, we provide the theoretical background for our hypotheses. First, we overview research on self-threat in conflict, with a particular focus on measurement challenges. Second, we discuss the literature on failures of perspective taking generally, and affective perspective taking specifically. Finally, we turn to the literature on naïve realism (i.e., the illusion of personal objectivity; Ross, 2018) to predict that parties in conflict will over-estimate the feelings of threat experienced by counterparts because they are overly certain regarding the objectivity and evidentiary strength of their own beliefs.

1.1. Prior research on self-threat in attitude conflict

Research on self-threat holds deep roots in cognitive dissonance theory, a theory that has made more impact on psychology than perhaps any other (Festinger, 1957). The core insight is now familiar to psychologists and non-psychologists alike: namely, the presence of contradicting cognitions produces in individuals the experience of "dissonance" – a state of aversive affective arousal that people are motivated to avoid. And while dissonance research has traditionally focused on *intrapersonal conflict* between beliefs and behavior, both classic and modern researchers have argued that cognitive dissonance also underpins the negative affect experienced during *interpersonal conflict*. As far back as the original articulation of dissonance, Festinger argued that: "The open expression of disagreement in a group leads to the existence of cognitive dissonance in its members" (Festinger, 1957, p. 261-262; Matz & Wood, 2005).

Later formulations of dissonance theory (e.g., Steele & Liu, 1981, 1983) worked toward greater precision as to why exactly contradictory ideas might lead to negative affect. This work proposed that the presence of arguments against one's core beliefs poses a threat to the integrity of an individual's self-concept, a phenomenon termed "self-threat." Specifically, Steele and colleagues theorized that the belief in a moral, intelligent, reasonable and agentic self is a fundamental psychological need, and that expressions of disagreement on important, identity-relevant issues may threaten this belief. This work proposed that exposure to arguments for opposing views may lead one to feel that one's self concept is under threat because questioning deeply held convictions may cause one to question one's own intelligence and morality.

Importantly, classic research on the role of self-threat did not *directly* measure the self-threat experience. Instead, the presence of threat was inferred when the relevant manipulations produced predicted results. For example, when self-affirmation manipulations theorized to bolster the self-concept increased participants' willingness to compromise, researchers inferred that self-threat caused the reluctance to compromise observed in the control condition (e.g., Cohen et al., 2000). Thus, the question of how exactly self-threat ought to be measured to allow for a precise comparison between self-reported measures and observer inferences remains open.

Precise measurement of self-threat in the face of attitude conflict poses several challenges. What specific questionnaire items might induce lay participants to honestly report the relevant psychological experience? It is unclear how well-understood terms such as "self-concept integrity" are to a prototypical experimental participant. This problem is further exacerbated by the fact that even if the terms were well-understood, individuals may be unwilling to answer honestly. To address this issue, we use a variety of measures to triangulate on the self-reported and counterpart-inferred levels of threat during attitude conflict. Our approach is informed by prior work suggesting that a key cause of self-threat in attitude conflict is the possibility that one's core beliefs might be proven false (Steele & Liu, 1981, 1983).[1] Thus our theorizing centers on the extent to which participants feel anxious or uneasy about how well their beliefs and arguments will stand up to scrutiny. Our work dovetails with current theorizing regarding the measurement of cognitive dissonance, which over time has come to equate dissonance with feelings of anxiety, unease, and discomfort (Jonas et al., 2014).[2]

Importantly, attitude conflict is likely to elicit multiple types of anxiety, beyond threat to one's self-concept. For example, individuals may feel anxious about the impact of any given interaction on their relationship with their counterpart. Especially in situations characterized by power asymmetries, individuals may feel anxiety regarding important life outcomes such as one's career or social standing. Finally, even people who are fully confident in their attitudes may feel anxious about their ability to argue effectively, especially in the face of an assertive counterpart. Although all these sources of anxiety have important behavioral consequences, they have not been theorized to threaten one's self concept to the extent that being proven incorrect in one's core beliefs has been. In other words, whereas an argument with one's boss might make one anxious about career advancement, it is unlikely to lead to a questioning of one's morality and rationality in the same manner as questioning the accuracy of one's convictions might.

To ensure that we are rigorously measuring self-threat, we employ several different approaches. Our measures range in the extent to which they require participants to explicitly reflect on threat vs. other conceptually related states. We also juxtapose self-threat with both a different negative affective state (anger) and other types of anxiety that are unrelated to threats to one's self-concept. In one experiment, we also employ financial incentives for truthful reporting.

1.2. Affective perspective taking in attitude conflict

An extensive literature on "perspective taking" makes it clear that individuals frequently fail to predict others' cognitions by anchoring on their own thoughts and preferences and adjusting insufficiently for apparent differences (e.g., Epley et al., 2006; Epley, Keysar, et al., 2004; Epley et al., 2004). By contrast, much less empirical work has examined the ability to predict counterpart feelings (i.e., affective perspective taking). Yet, emotional expressions provide information to counterparts, in turn triggering (1) inferential processes and (2) reciprocal affective reactions (Van Kleef, 2009). Despite the central role that perceived emotions play in regulating social behavior, and prior research suggesting that individuals are likely to make systematic errors in this domain, little research has focused on it.

Importantly, a large literature on affective forecasting demonstrates that individuals systematically mis-predict how events are likely to influence their *own* affective reactions. For example, studies have found that receiving tenure does not confer the sustained happiness that junior academics anticipate; relatedly, failing an important exam does not carry the affective sting that students fear (Gilbert et al., 1998; Wilson et al., 2000; Wilson & Gilbert, 2003, 2005). Given that individuals make systematic errors in predicting their own affect (which they have vast experience with), it seems likely that they might make interpersonal

[1] In other contexts outside of attitude conflict there are other sources of self-threat, for example failing to achieve one's academic or professional goals or failing to live up to one's moral ideals (Steele & Liu, 1981, 1983).

[2] For example, Jonas and colleagues (2004, p. 237) essentially equate measuring dissonance with measuring anxiety: "It was only when researchers began to zero in on [Behavioral Inhibition System]-specific anxious arousal… that the consciously reportable affective consequences [of dissonance] became clear.".

errors, as well.

A second, related line of research suggests that individuals fail to accurately predict how others' emotional responses will influence judgments and decisions. For example, Van Boven and Loewenstein (2005) describe empathy gaps in emotional perspective taking, in which people under-estimate the effect of emotional situations on others' preferences and decisions (for related work, see Campbell et al., 2014; O'Brien & Ellsworth, 2012). While this prior work has examined predictions about how others' emotions influence choices, we take on a related but distinct question: namely, how accurate are people in their predictions about others' affective states. Although several literatures suggest that individuals may be poor at this, this problem may be even more acute in conflict (Epley & Kardas, 2020), and remains understudied.

We address this gap in the literature by examining individuals' inferences and predictions about the feelings of threat experienced by their conflict counterparts. Specifically, we predict that individuals systematically over-estimate the extent to which attitude conflict induces feelings of self-threat in holders of opposing views. To do so, we compare self-reports to counterpart inferences. Below, we review prior research that gives rise to this specific prediction.

1.3. Naïve realism

Extensive research has demonstrated that individuals believe that their views and opinions, even those around highly contentious issues, stem from an accurate and unbiased assessment of reality. This phenomenon, referred to as "naïve realism" or "the illusion of personal objectivity" (e.g., Griffin & Ross, 1991; Pronin et al., 2004; Robinson et al., 1995; Ross & Ward, 1995, 1996) has been linked to a variety of psychological barriers to conflict resolution.

Partly as a consequence of perceiving their own views as reasonable and appropriate to the situation, people see the judgments of disagreeing others as having been contaminated by cognitive or motivational biases (Liberman et al., 2012; Pronin et al., 2004; Ross & Ward, 1995, 1996; Ross, 2018). When encountering disagreement, individuals do not sensibly ask whether they, their counterpart, or both are in error. Instead, they typically proceed from the assumption that their beliefs and responses are correct, and therefore that the other is wrong.

We build on three related literatures under the broader umbrella of naïve realism research to predict that people will infer that counterparts in disagreement are experiencing a higher level of self-threat than they do themselves. First, prior research has demonstrated that people generally view the evidence and arguments for opposing views to be of lower quality than evidence and arguments for their own beliefs. For example, in classic experiments on biased assimilation (Lord, Ross & Lepper, 1979; Lord, Lepper & Preston, 1984) participants provided with fictitious scientific evidence for and against the death penalty saw the evidence supporting their perspective as more sound than that supporting the other side. More recently, Minson et al. (2019) found that partisans found arguments in support of their beliefs on border security to be superior to those opposing their beliefs. Thus, we theorize that because people believe opponents to have weaker evidence for their views, they will also expect opponents to experience greater levels of self-threat when engaging with the opposing perspective.

Second, work on the false consensus effect demonstrates that people over-estimate the number of others who agree with their choices and perspectives (e.g., Ross et al., 1977). This work shares theoretical roots with the naïve realism tradition because the confidence in one's own objectivity and intelligence underpins people's belief that reasonable others will agree with them (Liberman et al., 2012). Importantly, this phenomenon has implications for dispute settings. For example, Babcock and Lowenstein (1997) demonstrated that participants induced to take on the perspective of a party in a lawsuit overestimated the likelihood that an objective judge will side with them. Because agreement is treated as evidence of the correctness of one's beliefs (Deutsch & Gerard, 1955), the conviction that a greater number of reasonable others side with the self rather than the disagreeing party should again lead to estimates of higher levels of threat for the disagreeing other than the self.

Finally, in addition to the considerations above, people also believe disagreeing others to be more prone to a host of specific psychological biases. Research on the so-called "bias blind spot" has shown that although people recognize several important biases in their own reasoning, they believe others, and especially disagreeing others, to suffer from them to a greater degree (Pronin, Gilovich & Ross, 2004; Pronin, Lin & Ross, 2002). Most relevant to our present concerns, Pronin and colleagues (2002) demonstrated that individuals believe others to experience a greater level of cognitive dissonance in consumer contexts, and thus to engage in more dissonance reduction strategies. Although this work examines a consumer decision (i.e., purchase regret) rather than attitude conflict, it broadly supports our hypothesis that individuals will over-estimate the level of threat produced by interpersonal attitude conflict, attributing a counterpart's unwillingness to change their mind to "defensiveness.".

In summary, prior research finds that people believe the reasons behind their convictions to be relatively stronger, their supporters to be relatively more numerous, and their minds to be relatively less biased, compared to disagreeing others. Here, we theorize that these tendencies, borne of naïve realism, lead individuals to conclude that they have less cause for concern in conflictual conversations than disagreeing counterparts. In other words, because disagreeing others have weaker evidence, fewer supporters and suffer from more biased reasoning, they must at some level, feel threatened by the potential error of their convictions. Because both parties hold these beliefs, the resulting pattern of data we predict is one of lower self-reported threat and higher threat inferred for conflict counterparts.

2. Methodological overview

2.1. Study overview

In four primary studies (collective N = 1,707; all pre-registered) and three supplemental studies (collective N = 1,002; two pre-registered), we examine individuals' inferences regarding the level of self-threat experienced by counterparts in attitude conflict. After generating items from a pilot study (Supplemental Study 1), we compare levels of threat reported for the self to levels of threat forecasted for a disagreeing counterpart. In Study 1, we investigate the over-estimation of threat in a within-subjects design during a virtual debate in the run-up to the 2020 United States presidential election. Study 2 tests the over-estimation in a between-subjects design in the context of political speeches. We also assess whether the over-estimation extends equally to a different negative affective state (anger). Study 3 compares threat forecasted for the self to threat forecasted for a disagreeing counterpart in a future argument. Study 3 also investigates the over-estimation of threat with a new measure, to assesses whether the over-estimation extends equally to general feelings of anxiety, and tests naïve realism as a mediator of the over-estimation. Supplemental Studies 2–3 use a similar paradigm to further assess robustness across multiple measures of threat and other types of anxiety, including with financial incentives for truthful reporting. Finally, Study 4 tests whether de-biasing the over-estimation of threat reduces confidence in one's own persuasion abilities. Together, the present studies investigate a novel barrier to effective communication and extend our understanding of how affect drives judgment and choice during attitude conflict.

2.2. Open science statement

We report how we determined our sample size, all data exclusions, all manipulations, and all measures. We did not analyze the data before reaching our predetermined sample size. Data, code, preregistrations, and materials are available here: https://researchbox.

org/577&PEER_REVIEW_passcode=JNLRLC. We conducted pilot studies to create an initial data point for our estimated effect size. In all cases, the effect size was a Cohen's d > 0.50. To obtain 95% power to detect this effect size in a between-subjects design, we needed 105 participants per experimental cell. We increased the sample size per cell to 200 to yield a smaller amount of uncertainty around the detected effect size.

3. Study 1

Study 1 provides an initial test of the hypothesis that individuals systematically over-estimate the level of self-threat experienced by conflict counterparts. We tested this hypothesis by recruiting participants for a synchronous debate on a hot-button political topic: the 2020 United States presidential election. After the debate, participants indicated their own level of self-threat experienced during the debate and reported their inferences regarding their counterpart's level of threat, in a counterbalanced order. We hypothesized that participants would over-estimate the level of threat reported by counterparts.

3.1. Method

We recruited participants in the United States through a third-party market research firm, ROI Rocket. We pre-registered to collect at least 300 conversations and ended up with 318. Conversations took place in the two-month period prior to the 2020 United States presidential election. The study consisted of two parts: a pre-survey (used for screening purposes) and the main survey (completed the following day).

The pre-survey began with an attention check and asked whether the participant would be available at the time of the main survey the following day. Participants then reported who they were planning to vote for in the election. They also indicated how strongly they supported their candidate, how strongly they opposed the other candidate, and how much they cared about politics – all on 5-point Likert scales anchored at 1 (Not at all) and 5 (Extremely). Finally, participants indicated their age and gender.

Participants were invited to complete the main survey if they met the following conditions: (1) they strongly supported their candidate (>=3 on the 5-point scale); (2) they strongly opposed the other candidate (>=3 on the 5-point scale); and (3) they were available and at the time of the main survey the following day.

The next day, eligible participants completed the main survey. To begin, we confirmed participants' voting preferences and that they were willing to have a 10-minute debate with another participant. Then, participants read the following instructions, with their preferred candidate (and the opposing candidate) inserted in place of the italics: "You reported that you intended to vote for the *Democratic Candidate/Republican Candidate* in the upcoming presidential election. In this survey, you will be paired with someone who intends to vote for the *Republican Candidate/Democratic Candidate*. You will talk to this person for ten minutes. Your goal in this conversation is to use the entire 10 min to discuss your beliefs about who is the best candidate. For example, the candidates differ on their approaches to immigration, response to COVID-19, and environmental policy. These are just some of the topics that you could cover. Your partner has received the same instructions.".

After correctly answering a few simple comprehension check questions (those who answered incorrectly twice in a row were precluded from completing the survey), participants read the following final instructions: "On the next page, you will be paired with another participant who is also completing this study to have a 10-minute chat-based conversation about the upcoming presidential election. You will receive a bonus of $3.00 at the end of this study if third-party coders determine that you remained on topic and engaged in the conversation with your counterpart for the full 10 min." We added this financial bonus to incentivize participants to fully engage in what is typically considered an aversive activity (Dorison et al., 2019).

Table 1
Full List of Threat Items Developed From Pilot Study (Used in Studies 1–4).

Self	Other
"To what extent, if at all, were you afraid of feeling uninformed?"	"To what extent, if at all, would they be afraid of feeling uninformed?"
"To what extent, if at all, were you scared that your opinions are not supported by facts?"	"To what extent, if at all, would they be scared that their own opinions are not supported by facts?"
"To what extent, if at all, were you worried that XXX might be right?"	"To what extent, if at all, would they be worried that XXX might be right?"
"To what extent, if at all, were you anxious about the idea that if you're wrong about this, you might be wrong about other things as well?	"To what extent, if at all, would they be anxious about the idea that if they're wrong about this, they might be wrong about other things as well?"

Note: XXX represents different targets used in different studies. Wording varied slightly across study contexts. Specific wording for each study can be found in our survey materials.

Participants then completed a 10-minute debate using the ChatPlat software (e.g., Huang et al., 2017; Wolf et al., 2016;). After completing the 10-minute debate, participants responded to dependent measures used for this study as well as for another, unrelated study.

Our dependent variables of interest were participants' reported level of self-threat during the debate and their inferred level of self-threat for their debate counterpart, answered in a counterbalanced order. The four items were generated based on the open-ended responses from the pilot study (Supplemental Study 1) and are presented in Table 1. All four items were presented in a randomized order and were answered on 9-point Likert scales from 0: "Not even the slightest bit" to 8: "More strongly than ever before." Items were adapted for each study context.

3.2. Results

Sample and descriptive statistics. A total of 4,344 participants completed the pre-survey over the two-month period in September and October 2020. 1,561 participants opened the main survey and 636 participants (318 dyads) were matched for a conversation with a supporter of the opposing presidential candidate. Of these 636 participants who were matched, 505 completed the entire survey (65% female, M_{age} = 54.17, SD_{age} = 13.88, 250 Republicans and 255 Democrats). Of these 505 participants, a third-party coder who reviewed all transcripts identified 367 who remained on topic for the entirety of the debate. These 367 participants served as our final dataset for analysis.

On average, participants wrote 142 words per conversation. The median word count per participant was 130 and the maximum was 520. The standard deviation was 71 words.

Over-estimation of threat. In both conditions, the items tapping participants' levels of threat were highly correlated (*alphas* > 0.86). Thus, we averaged the four items to create an index.

We next turned to testing our key hypothesis: whether in synchronous conversation, individuals over-estimate the level of threat felt by a conflict counterpart. In our study, participants served in both the role of "self" (in giving self-reports) and "other" (in making interpersonal predictions). Given the dyadic nature of the study, we thus used mixed effects models specifying target (i.e., self vs. other) as a fixed effect and including a random effect for dyad to account for multiple observations of the same conversation (one from each conversation counterpart).

As depicted in Fig. 1, results supported our hypothesis: participants inferred higher levels of threat for their counterparts than that which they reported for themselves (M_{self} = 2.00, SD_{self} = 1.40 vs. M_{other} = 2.65, SD_{other} = 1.72, b = 0.65, SE = 0.11, *95% CI* [0.44, 0.87], p <.001). Indeed, 48.2% of participants inferred higher levels of threat for their partner than that which they reported themselves compared to just 21.4% did the opposite (the remaining 30.4% forecasted equal levels of threat for self and partner). We did not find evidence that the misprediction was significantly different for Biden versus Trump supporters (b = 0.24, SE = 0.22, *95% CI* [-0.19, 0.67], p >.27).

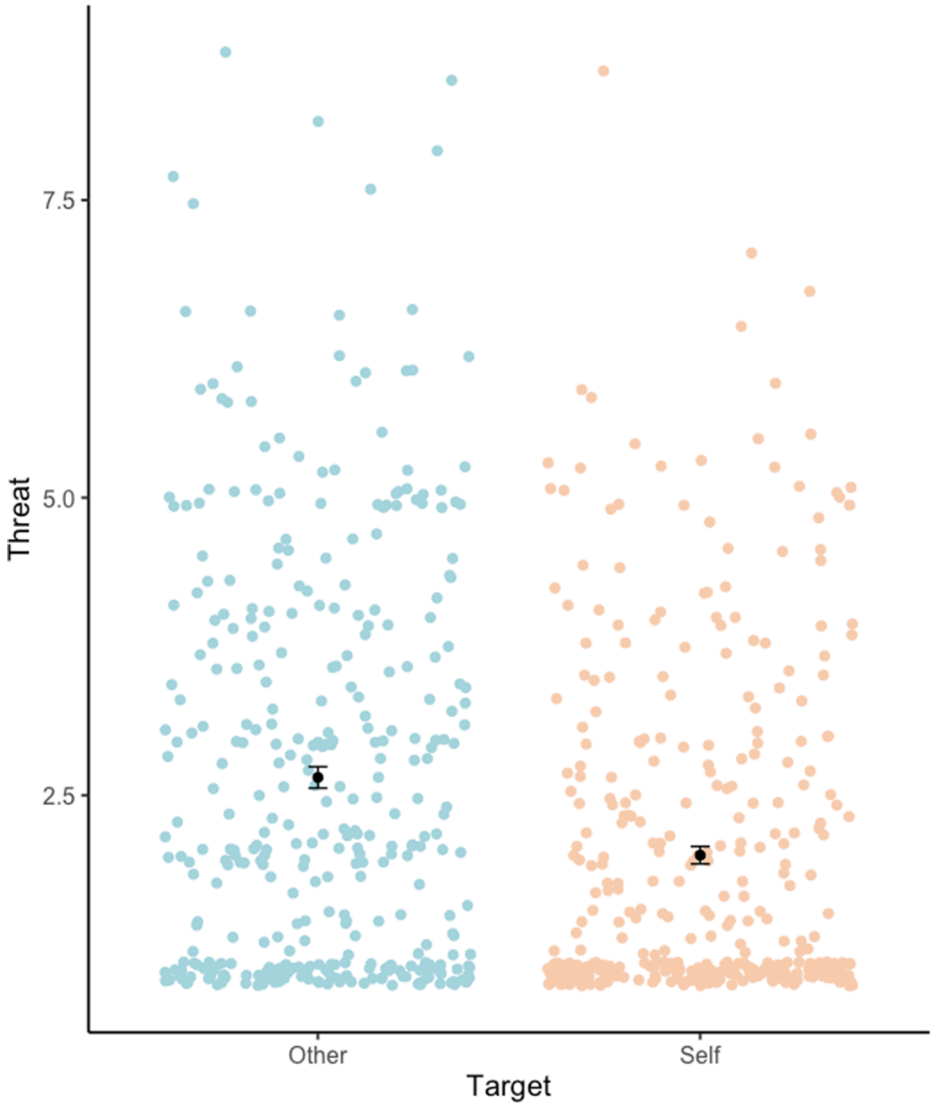


Fig. 1. *Over-estimation of Self-threat After Engaging in a Synchronous Debate in the Weeks Before the 2020 United States Presidential Election. Note.* Participants engaged in a 10-minute debate in the run-up to the 2020 United States Presidential Election (Study 1). They systematically over-estimated the level of threat experienced by their ideological opponents, as compared to the levels those opponents reported. Error bars represent one standard error and colored dots represent raw data.

Linguistic analyses. The conversations provided a rich set of text that allowed us to conduct an initial set of exploratory analyses related to the over-estimation of threat (full details available in our online materials). While future research could analyze the data in greater depth, we examined a small subset of relevant linguistic dimensions. First, we used the politeness R package (Yeomans et al., 2018), which was designed specifically to examine language in conflict. This package uses pre-trained natural language processing models to calculate a set of syntactic and social markers from natural language (e.g., words and phrases that express gratitude, apologies, acknowledgment). From the politeness package, we also assessed the number of positive and negative emotion words used in the conversation.

Results revealed that participants' use of positive emotion words was inversely related to self-reported feelings of threat, and also inversely related to inferences regarding the counterpart's feelings of threat, although this latter relationship was marginally significant. In other words, participants who used more positive emotion words perceived the exchange as less threatening for both themselves and their partner. However, the partner's use of positive emotion words was uncorrelated with the focal participant's perceptions of own and partner's level of threat. Surprisingly, none of these relationships were significant for negative emotion words.

In a final set of exploratory linguistic analyses, we assessed whether the proportion of words written by one partner in the conversation could predict feelings of threat. We found that the proportion of total words written was negatively correlated with self-reported feelings of threat and was unrelated to perceptions of a counterpart's threat. Taken together, these results suggest that linguistic analyses could provide a fruitful future path for understanding the dynamics of threat in conversation, but more work is needed before firm conclusions can be drawn.

3.3. Discussion

Study 1 provided initial evidence for the over-estimation of self-threat in a live, synchronous disagreement. Although both parties had the opportunity to offer their best arguments in support of their candidate, participants inferred that their opponents felt more threatened than they themselves were. Just as prior research makes clear that individuals make systematic errors in understanding the cognitions of others (i.e., failures of perspective taking), Study 1 reveals that they make systematic errors in understanding the affect experienced by others.

One obvious alternative explanation for these results is that people are reluctant to report their own negative affect. We explore this possibility in Study 2 by testing whether the documented over-estimation is specific to threat or general to other negative affective states (e.g., anger).

4. Study 2

Study 2 had three aims. First, we assessed whether the over-estimation of threat felt by conflict counterparts replicates in a between-subjects (rather than within-subjects) design. Second, we tested whether the over-estimation would persist when the participant did not generate the arguments featured in the study. Specifically, we examined affective predictions and reactions to speeches by professional politicians rather than those generated by the participants themselves. Third, and of most central theoretical concern, we examined whether the over-estimation would extend to a different negative affective state: anger. We selected anger because prior work on moral conviction has identified it as a key affective reaction to disagreements on moralized attitudes (e.g., Mullen & Skitka, 2006; Skitka, 2014; Skitka & Wisneski, 2011; Skitka et al., 2021; Wiltermuth & Flynn, 2013). Examining the extent to which people report anger and recognize it in others in the midst of attitude conflict also allows us to rule out the possibility that the documented threat misprediction is simply driven by individuals' reluctance to report their own negative affect while inferring it in others.

Table 2
Full List of Anger Items Developed From Pilot Study Uused in Studies 2–3).

Self	Other
"To what extent, if at all, were you angry that XXX was wasting your time?"	"To what extent, if at all, would they be angry that XXX is wasting their time?"
"To what extent, if at all, were you irritated that XXX doesn't change his mind in the face of good evidence?"	"To what extent, if at all, would they be irritated that XXX doesn't change his mind in the face of good evidence?"
"To what extent, if at all, were you frustrated that XXX wasn't using better critical thinking?"	"To what extent, if at all, would they be frustrated that XXX isn't using better critical thinking?"
"To what extent, if at all, were you mad that XXX might influence others who don't know better?"	"To what extent, if at all, would be they mad that XXX might influence others who don't know better?"

Note. XXX represents different targets used in different studies. Wording varied slightly across study contexts. Specific wording for each study can be found in our survey materials.

4.1. Method

We recruited 400 MTurk workers (179 female, 219 male, 2 non-binary/other, $M_{age} = 36.76$, $SD_{age} = 11.44$) for a "survey about political opinions." After an attention check, participants indicated their political ideology on a 7-point Likert scale from 1 (Very liberal) to 7 (Very conservative).

Participants were then randomly assigned to one of two between-subjects experimental conditions. In the "Self" condition, participants watched a video by a senator advocating for the opposing ideology and reported their own levels of threat and anger (described below). In the "Other" condition, participants watched a video clip of a senator advocating for the participant's *own* political ideology and forecasted the levels of threat and anger for an MTurker who holds the opposing political ideology. We used Senator Bernie Sanders and Senator Ted Cruz as our Liberal and Conservative target senators, respectively because at the time they held the most liberal and conservative voting records in the Senate and were highly familiar to the participants in our pool. Video clips were the most recent speeches uploaded to the YouTube channels of the respective senators at the time of the study (for a similar methodology, see Dorison et al., 2019).

Both threat and anger were measured using four Likert items, presented in a randomized order and answered on 9-point scales from 0: "Not even the slightest bit" to 8: "More strongly than ever before." The measurement of threat was identical to Study 1 with wording slightly adapted for the new experimental context. We created the anger measures in the same way as the threat measures: by adapting open-ended responses from the pilot study (Supplementary Study 1). Table 2 presents the full wording for all anger items. The average levels of threat and anger again served as our primary dependent variables. At the end of the study, all participants indicated their age and gender.

4.2. Results

In line with our pre-registration, we excluded from analysis anyone who (a) failed the attention check or (b) reported being "middle of the road" in their political ideology. We also excluded one participant who had missing data on multiple affect items. These exclusion criteria left us with a total of 319 participants.

Over-estimation of threat. Both the threat (alpha = 0.89) and anger (alpha = 0.86) scales achieved high levels of reliability. Additionally, an exploratory factor analysis indicated a two-factor solution, with the four threat items loading onto the first factor and the four anger items loading onto the second factor, leading us to create two indices by averaging the relevant items. From here on, the terms "threat" and "anger" refer to these indices.

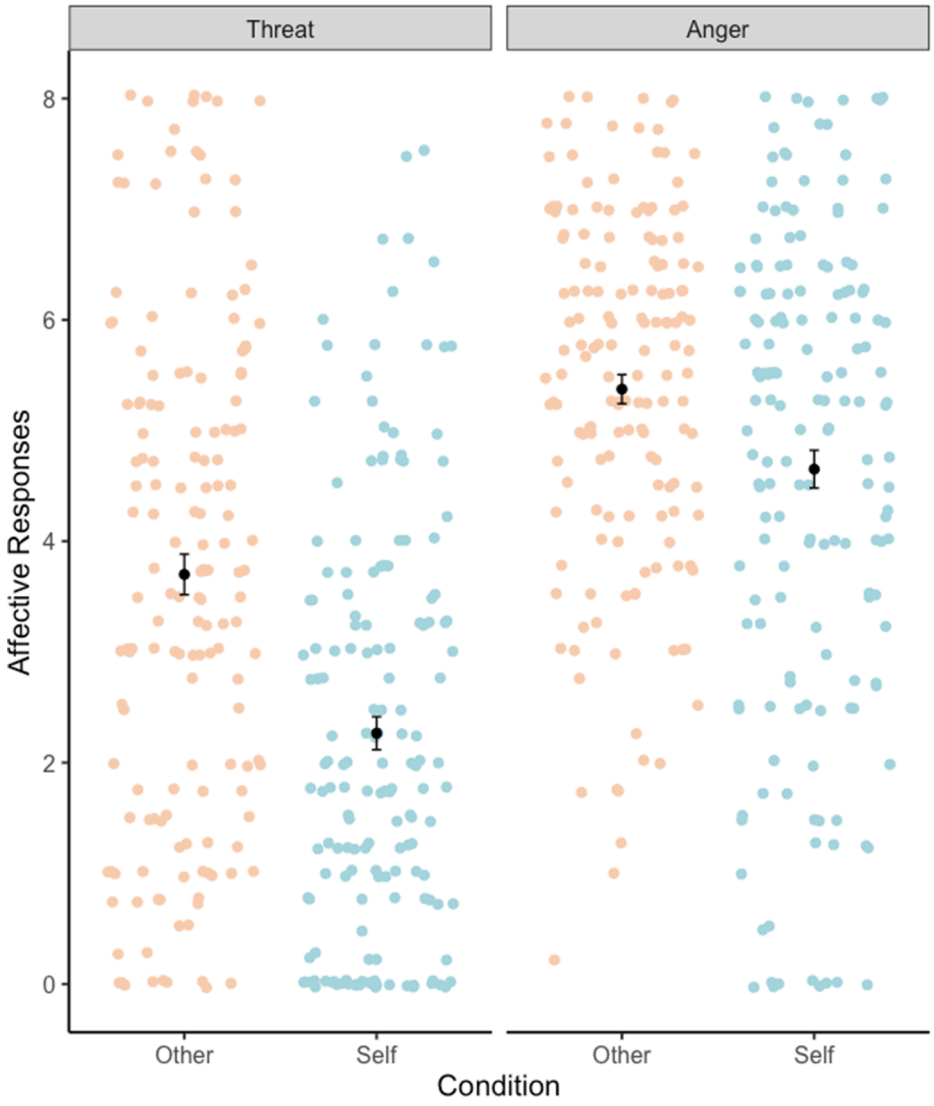


Fig. 2. *Over-estimation of Threat (More So Than Anger) After Watching Video Clips of Speeches by Professional Politicians. Note.* Participants watched a speech by a United States Senator (Study 2). They systematically over-estimated the level of threat experienced by their ideological opponents to a greater extent than they over-estimated anger. Error bars represent one standard error and colored dots represent raw data.

As depicted in Fig. 2, our data again support our hypothesis. Participants in the Other condition systematically over-estimated the levels threat that participants in the Self condition reported ($M_{Other} = 3.70$, $SD_{Other} = 2.29$ vs. $M_{Self} = 2.27$, $SD_{Self} = 1.91$, $t(302) = 6.07$, $p < .001$, *Cohen's d* $= 0.68$). To put these results in perspective, we ran a simulation in which we randomly drew 10,000 pairs of participants, one participant from each condition. Participants in the Other condition forecasted higher levels of threat than their randomly-selected match from the Self condition actually reported 65.8% of the time, and the reverse just 30.4% of the time (the remaining 3.8% of pairs indicated equal levels of threat).

Participants in the Other condition also tended to over-estimate the levels of anger that participants in the Self condition reported ($M_{Other} = 5.37$, $SD_{Other} = 1.64$ vs. $M_{Self} = 4.65$, $SD_{Self} = 2.18$, $t(300) = 3.36$, $p < .001$, *Cohen's d* $= 0.37$). However, a 2 (within: anger, threat) × 2 (between: self, other) mixed ANOVA provided evidence for a significant interaction ($F(1, 317) = 5.15$, $\eta^2 = 0.02$, $p = .024$), providing evidence that that the over-estimation of threat was significantly greater than the over-estimation of anger.

4.3. Discussion

Study 2 replicated the over-estimation of threat in a between-subjects design, and demonstrated that this over-estimation did not extend to the same degree to another negative affective state (anger). This result suggests that people are not simply reluctant to report negative emotions, and that they are not simply over-estimating negative affect for their counterparts. Rather, the experience of disagreement leads to a specific pattern of affective perspective taking where individuals overestimate the level of threat counterparts will experience when confronted with opposing arguments.

We also found that the over-estimation persisted when the participant did not generate the arguments on their own, suggesting that people do not simply believe that their own arguments should inspire threat, but rather that arguments associated with their side in a disagreement were superior to arguments on the other side.

5. Study 3

Study 3 had three goals. First, we assess the generalizability of the

over-estimation of self-threat using a new measure. Attitude conflict presents a relatively complex domain for self-threat research because concern about the correctness of one's beliefs serves as just one (of many) possible sources of anxiety-related states. In Studies 1–2, our measures included short descriptors of what exactly a person might be anxious about to ensure that when responding, participants were reflecting on the type of anxiety implied in prior research on self-threat, and not anxiety about other things (for example, anxiety regarding having an awkward social interaction). However, one weakness of this methodological approach comes from comparisons across affective states. That is, when we compare self-threat to anger in Study 2, it may not be the affective states per se that are driving differences in mispredictions, but rather differences related to the descriptors. Thus, in Study 3, we take a different approach by introducing a new set of items that more closely match the lay terminology used to describe feelings of self-threat without relying on more extensive descriptors. Specifically, we use the words "vulnerable" and "insecure" to capture the feeling of a general concern with one's self concept. This allows us to more tightly contrast self-threat with other negative affective states, but also to test robustness of the over-estimation with a new measure.

As argued above, attitude conflict is rife with multiple sources of anxious arousal, and there is no reason to predict that all of them would appear to apply to one's counterpart to a greater extent than to the self. For example, people might expect that both they and their counterpart are equally uncomfortable with interpersonal awkwardness (irrespective of topic), and people in low power positions might feel more anxious in arguments than high power counterparts, even if they are convinced of the correctness of their position. This logic suggests that if participants are presented with general anxiety measures that do not specify exactly what the target is anxious about, the pattern we document should be attenuated because different participants will be focusing on different sources of anxious arousal. To test this prediction, we included several items measuring general anxiety based on the same affective terms used in the earlier studies but without the accompanying descriptors (i.e., anxious, worried, scared, and nervous). Thus, Study 3 has three dependent measures: self-threat measured using descriptors (as in Studies 1–2), self-threat measured without descriptors (i.e., vulnerable and insecure), and general anxiety (anxiety, worried, scared, and nervous; measured without descriptors, as well).

Finally, the third goal of Study 3 is to test the psychological mechanism driving our effect. While Studies 1–2 provided consistent evidence that participants over-estimate the self-threat reported by conflict counterparts, the question of *why* such over-estimation occurs remains untested. We propose that the cause of this misprediction is naïve realism – individuals' conviction that their own views reflect an objective reality and are thus fundamentally reasonable and supported by sound evidence. Research on naïve realism has argued that to the extent that people view their own views as fundamentally objective, they infer bias and error in the disagreeing views of others. We predict that these expectations of superior personal objectivity will lead individuals to infer a greater level of threat in disagreeing counterparts because engagement with opposing views is likely to expose the error of those counterparts' convictions.

To achieve these three goals, we conducted a study in which participants imagined a future argument with a peer. We then compared threat (and other negative affective states) forecasted for the self to threat forecasted for a disagreeing counterpart. We predicted (1) that participants would over-estimate the threat forecasted by counterparts; (2) that such effects would hold across both measures of threat (i.e., both with and without descriptors); (3) that over-estimation of threat would be greater than the over-estimation of general anxiety; and (4) that the overestimation of threat for conflict counterparts would be mediated by naïve realism.

5.1. Method

We solicited participation by 400 MTurk workers for a 5-minute study of political opinions (220 female, 175 male, 5 = nonbinary/other, $M_{age} = 40.26$, $SD_{age} = 12.68$).

We prompted all participants to think of any topic, political or otherwise, that they held a very strong opinion about. After indicating the topic in a text box, participants were told to "Imagine a situation in which you are discussing this topic with someone who also has a very strong opinion but holds the opposite views from your own. Imagine that both you and this other person are making arguments for your points of view and trying to change each other's mind. Please imagine this situation as vividly as possible." These instructions ensured that the imagined scenario was perfectly symmetric for the participant and their imagined counterpart: both parties were engaged in persuasion regarding a topic on which they held strong attitudes.

As in Study 2, we then randomly assigned participants to either the Self condition or the Other condition in a between-subjects design. In the Self condition, participants answered three sets of questions regarding the emotions they themselves would feel during this conversation (described below). In the Other condition, participants answered the same three sets of questions, but indicated how they expected *the other person* to feel. The first set of questions were identical to the self-threat items used in Studies 1–2, with minor modifications to adjust for a future social interaction. We refer to this set of items as "threat with descriptors." The second set of questions were also designed to measure self-threat, but without additional verbal descriptors of what exactly one was feeling threatened about. Specifically, participants indicated to what extent, if at all, they (or their counterpart, depending on condition) would feel "vulnerable" and "insecure" during this argument. These two items formed the "threat without descriptors" index. Finally, the third set of questions measured general anxiety (also without descriptors). Specifically, participants indicated to what extent, if at all, they (or their counterpart, depending on condition) would feel "anxious," "worried," "scared," and "afraid." Of note, these are the same four anxiety items used in the "threat with descriptors" index, but with the key difference that they do not include the additional information that ties them directly to feelings of self-threat. We refer to this as the "general anxiety" index. The three sets of questions were presented in a randomized order. Further, within each set, items were presented in a randomized order. All affect items were answered on 9-point Likert scales from 0: "Not even the slightest bit" to 8: "More strongly than ever before." The average predicted levels of self-threat with descriptors, self-threat without descriptors, and general anxiety served as our primary dependent variables.

After answering the three sets of affect questions, participants answered a new question based on research on naïve realism and used to measure participants' beliefs regarding the soundness of their own versus their counterpart's views. Specifically, participants read the following text: "People generally believe that their attitudes on important issues are reasonable, objective, and supported by evidence. In other words, that their attitudes reflect the 'way things really are' in the world. However, when we are faced with opposing arguments, we might question the accuracy and objectivity of our earlier beliefs." After reading the block of text, we asked participants in the Self condition: "In the situation you imagined, to what extent would the arguments your opponent makes lead you to question whether your beliefs on this issue are fundamentally correct and objective?" Participants gave their answer on a 5-point scale from 1: "Not at all" to 5: "Very much." Participants in the Other condition answered the same question on the same scale, but with regard to how they expected their counterpart would feel as a result of their own arguments. To the extent that people are "naïve realists," confident in the veracity and objectivity of their own views,

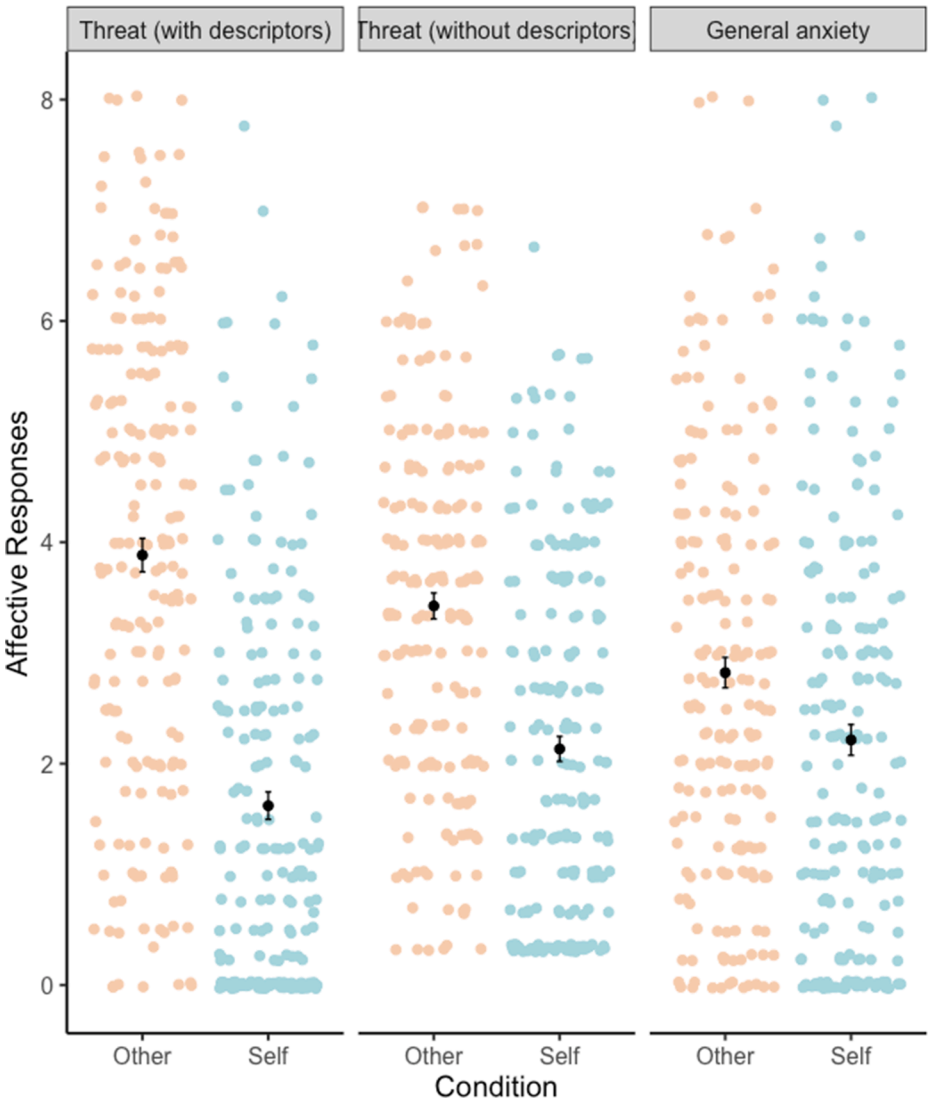


Fig. 3. *Over-estimation of Threat (Both With and Without Descriptors, and to a Greater Extent Than General Anxiety) When Imagining a Future Argument With a Peer.* ***Note.*** Participants imagined a future argument with a peer (Study 3). They systematically over-estimated the level of threat experienced by their ideological opponents (when threat was measured both with and without descriptors) to a greater extent than they over-estimated general anxiety. Error bars represent one standard error and colored dots represent raw data.

participants should report higher values on these items for disagreeing others versus for themselves. We predicted that this difference would mediate the difference in forecasted levels of threat (both with and without descriptors) for the self versus the disagreeing counterpart.

Finally, at the end of the survey, participants completed demographic measures, including age, gender, and ethnicity.

5.2. Results

In line with our pre-registration, we excluded from analysis one participant who failed the attention check, leaving us with a total of 399 participants.

Over-estimation of threat. We first assessed the internal reliability of the three indices. Both the threat with descriptors (*alpha* = 0.92) and general anxiety (*alpha* = 0.92) scales achieved high reliability. In addition, the two items in the threat without descriptors index (i.e., vulnerable and insecure) were highly correlated ($r = 0.83$). An exploratory factor analysis revealed a two-factor solution, in which the two threat sub-scales (with and without descriptors) loaded onto the first factor and the general anxiety items loaded onto the second factor. Given our theoretical interest in testing whether the over-estimation of threat persisted across the two different measurements, we kept them as separate indices (although effects persist across all threat items when analyzed jointly as a single scale).

We then conducted inferential analyses. Replicating our prior results, participants in the Other condition predicted significantly higher levels of threat with descriptors for their counterparts than participants in the Self condition predicted for themselves ($M_{other} = 3.88$, $SD_{other} = 2.13$ vs. $M_{self} = 1.61$, $SD_{self} = 1.77$, $t(397) = 11.60$, $p < .001$, *Cohen's d* = 1.16). Importantly, the same pattern held true, and was similarly large in magnitude, for the new measure of threat without descriptors. That is, even when measured without the additional information, participants over-estimated how much threat their counterpart would feel compared to participants' predictions for themselves ($M_{other} = 3.42$, $SD_{other} = 1.64$, vs. $M_{self} = 2.13$, $SD_{self} = 1.61$, $t(397) = 7.97$, $p < .001$, *Cohen's d* = 0.80).

Participants also over-estimated a counterparts' feelings of general anxiety ($M_{other} = 2.82$, $SD_{other} = 1.92$ vs. $M_{self} = 2.21$, $SD_{self} = 1.98$, $t(397) = 3.11$, $p = .002$, *Cohen's d* = 0.31). However, our key question was whether this over-estimation was smaller than the over-estimation of threat. To address this question, we conducted a 2 (between: self, other) × 2 (within: threat, general anxiety) mixed ANOVA. We repeated this analysis twice: once comparing threat with descriptors to general anxiety and one comparing threat without descriptors to general anxiety. In both cases, we found evidence that the over-estimation of threat was greater than the over-estimation of general anxiety (with descriptors: $F(1,397) = 84.60$, $\eta^2 = 0.18$, $p < .001$; without descriptors: $F(1,397) = 21.15$, $\eta^2 = 0.05$, $p < .001$). Results are depicted in Fig. 3.

Naïve realism and mediation. Based on previous work on naïve

realism, we predicted that individuals would expect their counterparts to question the correctness and objectivity of their beliefs more than they would question the correctness and objectivity of their own beliefs. This was indeed the case ($M_{other} = 2.80$, $SD_{other} = 1.14$ vs. $M_{self} = 1.72$, $SD_{self} = 0.95$, $t(397) = 10.35$, $p < .001$, *Cohen's d* = 1.04).

Of more central theoretical interest, we designed Study 3 to test whether the difference in naïve realism reported above was a driver of the over-estimation of threat in conflict counterparts. To address this question, we conducted two sets of between-subjects mediation analysis with the Lavaan package in R (Rosseel, 2012) with 10,000 bootstrapped samples. In both models, the independent variable was condition (1 = Other, 0 = Self), the mediating variable was naïve realism, and the dependent variable was feelings of threat. In the first model, threat was measured using the threat with descriptors index. In the second model, threat was measured using the threat without descriptors index (i.e., "vulnerable," "insecure"). Consistent with predictions, in both models, naïve realism mediated the self-other difference in predictions of self-threat during conflict (with descriptors: $b = 0.24$, *95% CI* [0.29, 0.19], $z = 9.23$, $p < .001$; without descriptors: $b = 0.27$, *95% CI* [-0.32, −0.21], $z = 9.36$, $p < .001$). Of note, naïve realism mediated 48% and 72% of the total effect of condition on threat for the "with descriptors" and the "no descriptors" indices, respectively.

5.3. Discussion

The results of Study 3 provide three key pieces of evidence. First, the over-estimation of threat persisted with a new measure, one that did not include descriptive information regarding the source of threat. Second, the over-estimation of self-threat did not extend equally to general anxiety. Finally, the over-estimation of self-threat was mediated by naïve realism.

In addition to Study 3, we also conducted two additional Supplemental Studies to further examine the robustness of our findings. Specifically, we address a key alternative explanation for our results based on the idea that people are simply unwilling to report feelings of self-threat. While we describe these studies in greater detail in the Supplementary Information, it is worth briefly summarizing the results here. In Supplementary Study 2, we ask participants about an additional source of anxiety: the anxiety they or their counterparts would feel if the policies supported by their opponents came to be implemented. While we again replicate the over-estimation of threat, we find a full reversal of our earlier results with regard to this new type of anxiety: participants report being more anxious about this possibility than they expect their counterparts to be. This study provides further evidence that participants are not simply reluctant to report anxiety-related states, rather (as suggested by Study 3) they are certain in the accuracy and objectivity of their beliefs and thus feel that disagreeing others have greater reason to feel threat.

Finally, in Supplementary Study 3, we use the Bayesian Truth Serum technique (Prelec, 2004; see also John et al., 2012) to financially incentivize participants to accurately report their affect. We find a pattern of results almost identical to our prior research: participants over-estimated others' levels of self-threat, but not others' levels of anger. Taken together, Study 3 (and Supplemental Studies 2–3) provided converging evidence regarding the methodological robustness of our effect across variations of measurement, incentive procedures, and comparison affective states.

6. Study 4

In Study 4, we turn to examining a behavioral consequence of the over-estimation of self-threat. Specifically, we test whether erroneous beliefs about how much threat one's counterparts are experiencing might lead individuals to engage in fruitless debates, falsely believing that they have a real chance at persuading the other side. Extensive evidence makes clear that individuals over-estimate their abilities across a wide variety of domains (for review, see Moore & Healy, 2008). Could it be the case that over-estimation of threat experienced by conflict counterparts leads individuals to harbor excessive confidence in their persuasion abilities?

6.1. Method

We solicited participation from 400 MTurk workers (160 female, 240 male, $M_{age} = 36.62$, $SD_{age} = 11.22$) for a 5-minute study of political opinions. All participants first completed an attention check asking about the purpose of the study.

After reporting their general political orientation on a 7-point Likert scale from 1: "Very Liberal" to 7: "Very Conservative," participants indicated their agreement with statements concerning five controversial policy topics: the death penalty, recreational marijuana, presidential job performance, illegal immigration, and gun control. Agreement was scored on a 7-point Likert scale from −3 (Strongly Disagree) to + 3 (Strongly Agree). Participants also ranked the five issues in terms of how strongly they felt about each. After providing their views, participants were told that we would attempt to match them with a counterpart for a debate on the issue that they felt most strongly about.

While the participant ostensibly waited for the experimental software to match them with a debate partner, they were asked to answer several questions. First, participants forecasted their level of self-threat during the upcoming debate by answering the four items used in Studies 1–3. Second, participants were told that they would be given the opportunity to bet up to $0.50 on whether they would be able to persuade their counterpart. They were told that if they persuaded their counterpart any money they bet would be doubled. However, if they did not persuade their counterpart, any money they bet would be lost.

Participants were then randomly assigned to one of two between-subjects conditions, which varied in the information they received about their debate partner. Participants were told that their partner answered the same emotion questions that they themselves had just answered. In the Realistic condition (i.e., the Treatment Condition), we presented participants with a counterpart who had purportedly indicated the *typical* levels of self-threat reported by participants in the Self conditions from previous studies. In the Imagined condition (i.e., the Control Condition), we presented participants with a counterpart who had purportedly indicated the levels of self-threat *forecasted* by participants in the Other conditions from previous studies. Thus, the Realistic condition represented an aggregated version of actual participant responses, while the Imagined condition represented an aggregated version of the self-threat levels forecasted by prior participants. To control for any possible effects of real or imagined political extremity, participants were told that counterparts were moderately conservative (if the participant was liberal) or moderately liberal (if the participant was conservative). After seeing the counterpart's (fictional) responses on the self-threat items, participants were reminded of the betting procedure and chose how much they would like to wager, if at all. This choice served as our pre-registered dependent variable.

Finally, participants were told that we could not match them with a partner and thanked for their time. We paid all participants the amount they would have earned if they had won the debate. At the end of the survey, participants completed demographic measures, including age, gender, and ethnicity.

6.2. Results

On average, participants bet 41% of their bonus on their ability to persuade their counterpart. 23% of participants bet their entire bonus, 18% of participants bet some of their bonus, and 59% of participants bet none of their bonus. Thus, rather than a normal distribution, our distribution depicted an inverted U-shape, with modal responses at 0% and 100%. Results are depicted in Fig. 4.

The percentage of the bonus money that participants chose to bet

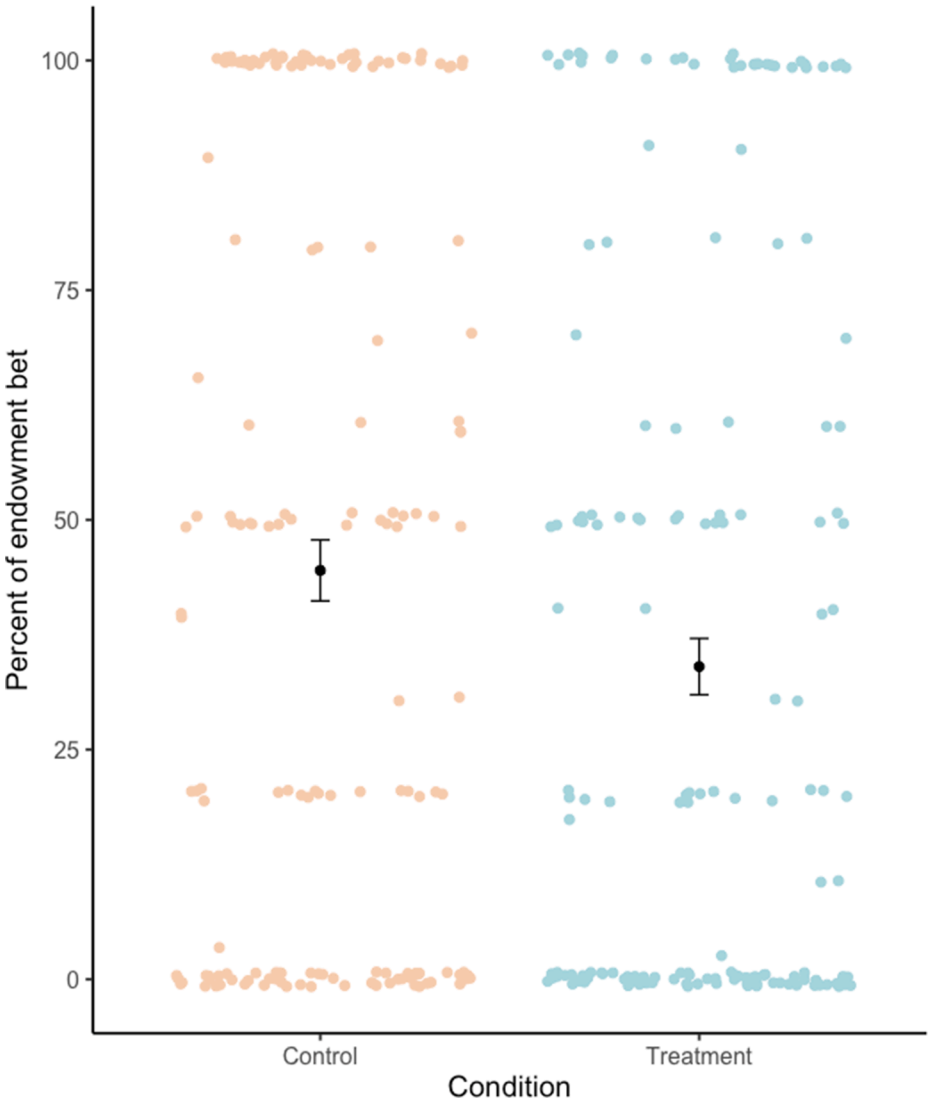


Fig. 4. *Willingness to Bet on One's Ability to Persuade a Counterpart Decreased When Participants' Beliefs Regarding the Counterpart's Level of Threat Were De-biased. Note.* Participants had the opportunity to bet on their ability to persuade a counterpart (Study 4). Their willingness to bet significantly decreased when they received an informational intervention that de-biased their beliefs about their counterpart's level of self-threat. Error bars represent one standard error and colored dots represent raw data.

differed by condition. Participants in the Realistic condition bet 34% of their bonus ($SD = 39\%$), whereas participants in the Imagined condition bet 44% of their bonus ($SD = 42\%$). A *t*-test confirmed that this difference was statistically significant ($t(317) = 2.26$, $p = .021$, *Cohen's d* = 0.26).

We pre-registered to analyze this variable using a *t*-test. However, the descriptive statistics above make clear that the distribution is not normal; rather, the modal responses are 0 and 100%. Thus, for robustness, we tested the effect of condition on betting using a beta regression, which does not make assumptions regarding the normality of the distribution of the dependent variable. The effect of condition on willingness to bet was robust to this new analysis: $b = 0.30$, $se = 0.15$, $z = 1.98$, $p = .047$.[3]

6.3. Discussion

Study 4 documents a downstream behavioral consequence of the over-estimation of threat in conflict counterparts. Specifically, participants were willing to bet greater amounts of real money when facing a counterpart who reported an unrealistically elevated level of threat associated with the possibility of being proven wrong. While we do not have a performance measure of persuasion in this study, this result suggests that the over-estimation of feelings of threat among parties in attitude conflict causally leads people to engage in persuasion attempts—attempts that prior research has shown to be largely fruitless (although see Kalla & Broockman, 2020).

7. General discussion

Conflict about attitudes – ranging from political beliefs, to family norms, to professional convictions – pervades daily life. Successfully navigating such conflict serves as a foundation for well-functioning relationships, organizations, and even democracies. To do so, individuals must accurately predict how their actions will impact others.

Four pre-registered studies revealed four key results. First, individuals systematically over-estimated the level of self-threat reported by conflict counterparts. Second, such over-estimation did not equally generalize to either a separate affective state (e.g., anger) or other types

[3] It could be the case that perceived threat increases the likelihood that someone bets at all. If so, then the percentage of individuals who bet any amount of money should be higher in the Imagined condition than in the Realistic condition. We find partial support for this hypothesis: Participants in the Realistic condition bet 54% of the time, whereas participants in the Imagined condition bet 65% of the time. A logistic regression showed that this difference was marginally significant ($z = 1.88$, $p = .061$).

of anxiety (e.g., generalized anxiety). Third, the over-estimation was underpinned by naïve realism: the belief that one's views are more objective and boast a greater evidentiary base than those of disagreeing others. Finally, the over-estimation of threat was substantial enough that it causally affected behavior with real financial consequences, leading individuals to harbor greater confidence in their persuasion abilities.

7.1. Contributions

Our work makes three theoretical contributions. First, our research contributes to the research literature on perspective taking and mis-prediction. An extensive body of prior work demonstrates that understanding one's counterpart is critical to successful conflict resolution (e.g., Bruneau & Saxe, 2012; Ickes, 1993; Galinsky & Mussweiler, 2001; Galinsky & Moskowitz, 2000; Galinsky et al., 2008; Neale & Bazerman, 1983). Traditionally, prior research has focused on failures of judging other's motives, intentions, evaluations, and situational construals (e.g., Epley et al., 2006; Epley, Keysar, et al., 2004; Epley et al., 2004). Here, we examine errors in interpersonal, affect-based judgments. Such an over-estimation is critical because emotions carry information about one's counterparts' reactions and behavioral intentions, in turn suggesting specific strategies and interventions for managing conflictual dialogue. Our work thus contributes to the research literature by suggesting a necessary complementary focus on affective perspective taking, especially during attitude conflict.

Indeed, outside of the domain of attitude conflict, prior research on affective forecasting has made clear that individuals mis-predict their own affective reactions across a wide variety of events, and that these faulty predictions can drive many sub-optimal decisions (e.g., Wilson & Gilbert, 2003, 2005; Morewedge & Buechel, 2013; Wilson et al., 2004; Dorison et al., 2019). Our work adds not only to this literature, but also to a growing body of research examining emotional perspective taking (Van Boven & Loewenstein, 2005; Campbell et al., 2014; Van Boven et al., 2013), in which individuals systematically mis-predict how others' affective reactions will in turn influence their behavior.

Importantly, a second contribution of our work is to the research literature on self-threat. Scholars have long theorized that the experience of self-threat underpins the negative affective consequences of attitude conflict. In turn, a growing body of empirical research has attempted to use self-affirmation as a conflict management strategy, driven by the hypothesis that affirming one specific aspect of the self-concept will reduce threat to the self-concept more broadly, and thus increase engagement with opposing views (Badea & Sherman, 2019; Binning, Sherman, Cohen, & Heitland, 2010; Cohen et al., 2007; Cohen & Sherman, 2014; Sherman, Brookfield, & Ortosky, 2017; Sherman, Lokhande, Müller, & Cohen, 2021). Importantly, however, prior research has not measured the experience of self-threat or tested whether parties in conflict are accurate in predicting how much threat their counterpart is *actually* experiencing. Here, we demonstrate that individuals (and perhaps even psychologists) systematically over-estimate the self-threat experienced by conflict counterparts.

Third, our work contributes to the influential research literature on naïve realism (i.e., the illusion of personal objectivity; Ross, 2018). Prior research under the broad umbrella of naïve realism has been linked to a variety of psychological barriers to conflict resolution (Griffin & Ross, 1991; Minson, Liberman, & Ross, 2011; Pronin, Gilovich, & Ross, 2004; Robinson, Keltner, Ward, & Ross, 1995; Ross & Ward, 1995, 1996). The present work connects this literature to the two mentioned above (perspective taking/mis-prediction and self-threat), yielding new insights about a novel barrier.

Our work also has applied contributions. For example, our work has implications for optimizing organizational and team performance because organizational life requires individuals to successfully navigate disagreement. Successfully forecasting how one's actions will influence a counterpart serves as a foundation for such endeavors. Importantly, our work also documents a barrier to successful conflict resolution that is exacerbated by the mis-prediction of counterparts' level of threat: excessive confidence in one's ability to persuade others. Although in our research we offered participants a single chance to bet on their success to measure such confidence, we suspect that in the world outside of the research laboratory several related phenomena would also emerge. For example, individuals might be more willing to enter an argument rather than walk away, or erroneously believe that their initial argument won the day (e.g., Conger, 1998). If people continue to believe in the correctness of their views, and mistakenly infer self-threat in their counterpart, they may dismiss any attempts to counter argue as "defensiveness." If this is the case, most arguments the counterpart could make, almost irrespective of quality, will fall on deaf ears. This dynamic provides some explanation for why most persuasion attempts around important, identity-relevant issues turn out to be futile.

7.2. Limitations and future directions

Multiple limitations of the present studies merit note and offer direction for future research. First, open questions remain regarding boundary conditions for over-estimation of threat in counterparts. While we found a robust over-estimation of threat across studies, the size of the effect varied from approximately a half standard deviation to over a standard deviation across studies and measurement. Future research is needed to assess when such effects are likely to be heightened or attenuated. For example, it could be the case that individuals in close relationships are more accurate when forecasting their counterpart's affective reactions. Additionally, this error may be less pronounced among individuals who come from cultural contexts in which attention to the psychological states of others is of greater social import (Markus & Kitayama, 1991). Furthermore, the over-estimation of threat that we document may have a temporal component. The passage of time may help individuals better distinguish between stronger and weaker arguments and become better calibrated with regard to the state of the world and to the mental state of people on the other side of any given debate. An experience sampling approach may be well-suited to assess this question.

Second, the present studies focused on two emotions of interest to theory and research on self-threat and disagreement: anxiety and anger. Future research could examine affective reactions and perspective taking for a host of negative (and positive) affective states, including but not limited to sadness, guilt, and shame. While recent work has identified 27 categories of emotion bridged by continuous variants (Cowen & Keltner, 2017), perhaps most theoretically interesting for the present investigation is the emotion of pride. The positive emotion of pride serves as a theoretically meaningful counterfactual to self-threat because it is associated with feelings of certainty (unlike threat, which is associated with uncertainty). Our studies suggest that individuals in conflict would under-estimate the pride felt by opposing partisans. Future research is needed to test not only this specific prediction, but also to broaden the scope of investigation to other affective states in attitude conflict.

Third, open questions remain regarding different types of conflict. While the present investigation examined both political and non-political attitude conflict and found concordance, future research could examine how our findings may vary by domain.

7.3. Conclusion

Social psychology has furnished the world with many examples of human inferential shortcomings in domains both familiar and novel. Yet none of us can claim lack of experience when it comes to conflict. We observe our counterparts' emotions through their words, their body language, their tone, and sometimes the objects they throw at us. Indeed, relational and organizational success requires effective conflict resolution on a daily basis. Yet, it seems that even in this familiar context and even with incentives for accuracy, people systematically misjudge

their counterparts' affect. When both people think that they are more accurate than the other side, one of them is likely to be wrong. In the case of predicting others' feelings of threat, our data suggest that they both are.

Funding.

The present work was supported by the authors' faculty research fund, Harvard's Foundations of Human Behavior Initiative, the Harvard Mind-Brain-Behavior Interfaculty Initiative, and the Harvard Program on Negotiation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives, 11*, 109–126.

Badea, C., & Sherman, D. K. (2019). Self-affirmation and prejudice reduction: When and why? *Current Directions in Psychological Science, 28*(1), 40–46.

Binning, K. R., Sherman, D. K., Cohen, G. L., & Heitland, K. (2010). Seeing the other side: Reducing political partisanship via self-affirmation in the 2008 presidential election. *Analyses of Social Issues and Public Policy, 10*(1), 276–292.

Bruneau, E., & Saxe, R. (2012). The Power of Being Heard: The Benefits of 'Perspective-Giving' in the Context of Intergroup Conflict. *Journal of Experimental Social Psychology, 48*(4), 855–866.

Campbell, T., O'Brien, E., Van Boven, L., Schwarz, N., & Ubel, P. (2014). Too much experience: A desensitization bias in emotional perspective taking. *Journal of Personality and Social Psychology, 106*(2), 272.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self. *Personality and Social Psychology Bulletin, 26*(9), 1151–1164.

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology, 65*, 333–371.

Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology, 93*(3), 415–430.

Collins, T. P., Crawford, J. T., & Brandt, M. J. (2017). No evidence for ideological asymmetry in dissonance avoidance: Unsuccessful close and conceptual replications of Nam, Jost, and van Bavel (2013). *Social Psychology, 48*(3), 123–134.

Conger, J. A. (1998). Qualitative research as the cornerstone methodology for understanding leadership. *The Leadership Quarterly, 9*(1), 107–121.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences, 114*(38), E7900–E7909.

Cronin, M. A., & Weingart, L. R. (2007). Representational gaps, information processing, and conflict in functionally diverse teams. *The Academy of Management Review, 32*(3), 761–773.

De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology, 88*(4), 741–749.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology, 51*(3), 629–636.

Dorison, C. A., Minson, J. A., & Rogers, T. (2019). Selective exposure partly relies on faulty affective forecasts. *Cognition, 188*, 98–107.

Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking: Reactive egoism in social interaction. *Journal of Personality and Social Psychology, 91*(5), 872.

Epley, N., & Kardas, M. (2020). Understanding the minds of others: Activation, application, and accuracy of mind perception. In P. Van Lange, T. Higgins, & A. Kruglanski (Eds.), Social Psychology: Handbook of Basic Principles (3nd Ed.).

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*(3), 327.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*(6), 760–768.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson.

Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology, 72*, 1–12.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology, 78*(4), 708–724.

Galinsky, A. D., & Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology, 81*(4), 657–669.

Galinsky, A. D., Magee, J. C., Gruenfeld, D. H., Whitson, J. A., & Liljenquist, K. A. (2008). Power reduces the press of the situation: Implications for creativity, conformity, and dissonance. *Journal of Personality and Social Psychology, 95*(6), 1450–1466.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*(3), 617–638.

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. *Advances in Experimental Social Psychology, 24*, 319.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*(4), 555.

Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of personality and social psychology, 113*(3), 430–452.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*(4), 587–610.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532.

Jonas, E., McGregor, I., Klackl, J., Agroskin, D., Fritsche, I., Holbrook, C., … Quirin, M. (2014). Threat and defense: From anxiety to approach. In *Advances in Experimental Social Psychology* (pp. 219–286). Academic Press.

Judd, C. M. (1978). Cognitive effects of attitude conflict resolution. *Journal of Conflict Resolution, 22*(3), 483–498.

Kalla, J. L., & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review, 114*(2), 410–425.

Liberman, V., Minson, J. A., Bryan, C. J., & Ross, L. (2012). Naïve realism and capturing the "wisdom of dyads". *Journal of Experimental Social Psychology, 48*(2), 507–512.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47*(6), 1231–1243.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098–2109.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253.

Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology, 88*(1), 22.

Minson, J. A., Chen, F. S., & Tinsley, C. H. (2019). Why won't you listen to me? Measuring receptiveness to opposing views. *Management Science*, in press.

Minson, J. A., & Dorison, C. A. (2021). Toward a psychology of attitude conflict. *Current opinion in psychology, 43*, 182–188.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin, 37*(10), 1325–1338.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502.

Morewedge, C. K., & Buechel, E. C. (2013). Motivated underpinnings of the impact bias in affective forecasts. *Emotion, 13*(6), 1023.

Mullen, E., & Skitka, L. J. (2006). Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, group differentiation, or anger? *Journal of Personality and Social Psychology, 90*(4), 629–643.

Nam, H. H., Jost, J. T., & Van Bavel, J. J. (2013). "not for all the tea in China!" political ideology and the avoidance of dissonance-arousing situations. *PLoS ONE, 8*(4).

Neale, M. A., & Bazerman, M. H. (1983). The role of perspective-taking ability in negotiating under different forms of Arbitration. *ILR Review, 36*(3), 378–388.

O'Brien, E., & Ellsworth, P. C. (2012). More than skin deep: Visceral states are not projected onto dissimilar others. *Psychological Science, 23*(4), 391–396.

Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science, 306*(5695), 462–466.

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.

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.

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal:" Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology, 68*(3), 404.

Ross, L. (2018). From the fundamental attribution error to the truly fundamental attribution error and beyond: My research journey. *Perspectives on Psychological Science, 13*(6), 750–769.

Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*(3), 279–301.

Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology, 27*, 255–304.

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

Ross, M. H. (1993). *Management of conflict: Interpretations and interests in comparative perspective*. Yale University Press.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Sherman, D. K., Brookfield, J., & Ortosky, L. (2017). Intergroup conflict and barriers to common ground: A self-affirmation perspective. *Social and Personality Psychology Compass, 11*(12), 1–13.

Sherman, D. K., Lokhande, M., Müller, T., & Cohen, G. L. (2021). Self-Affirmation Interventions. In G. M. Walton, & A. J. Crum (Eds.), *Handbook of Wise Interventions: How Social Psychology Can Help People Change* (pp. 63–91). essay, The Guilford Press.

Skitka, L. J. (2014). The psychological foundations of moral conviction. In J. Wright, & H. Sarkissian (Eds.), *Advances in Moral Psychology* (pp. 148–166). New York, NY: Bloomsbury Academic Press.

Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The psychology of moral conviction. *Annual Review of Psychology, 72*, 347–366.

Skitka, L. J., & Wisneski, D. C. (2011). Moral conviction and emotion. *Emotion Review, 3* (3), 328–330. https://doi.org/10.1177/1754073911402374

Steele, C. M., & Liu, T. J. (1981). Making the dissonant act unreflective of self: Dissonance avoidance and the expectancy of a value-affirming response. *Personality and Social Psychology Bulletin, 7*(3), 393–397.

Steele, C. M., & Liu, T. J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology, 45*(1), 5–19.

Thomas, K. W. (1992). Conflict and conflict management: Reflections and update. *Journal of Organizational Behavior, 13*(3), 265–274.

Tjosvold, D., Wong, A. S. H., & Chen, N. Y. F. (2014). Constructively managing conflicts in organizations. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 545–568.

Van Boven, L., & Loewenstein, G. (2005). Empathy gaps in emotional perspective taking. *Other minds: How humans bridge the divide between self and others*, 284-297.

Van Boven, L., Loewenstein, G., Dunning, D., & Nordgren, L. F. (2013). In *Changing places: A dual judgment model of empathy gaps in emotional perspective taking* (pp. 117–171). Academic Press.

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science, 18*(3), 184–188.

Van Kleef, G. A., & Côté, S. (2007). Expressing anger in conflict: When it helps and when it hurts. *Journal of Applied Psychology, 92*(6), 1557–1569.

Van Kleef, G. A., & Côté, S. (2018). Emotional dynamics in conflict and negotiation: Individual, dyadic, and group processes. *Annual Review of Organizational Psychology and Organizational Behavior, 5*, 437–464.

Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2004). The Interpersonal Effects of Emotions in Negotiations: A Motivated Information Processing Approach. *Journal of Personality and Social Psychology, 87*(4), 510–528.

Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2010). An interpersonal approach to emotion in social decision making: The Emotions as Social Information model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 45–96). Academic Press.

Webb, T. L., Chang, B. P. I., & Benn, Y. (2013). 'The ostrich problem': Motivated avoidance or rejection of information about goal progress. *Social and Personality Psychology Compass, 7*(11), 794–807.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology, 35*, 345–411.

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science, 14*(3), 131–134.

Wilson, T. D., Wheatley, T. P., Kurtz, J. L., Dunn, E. W., & Gilbert, D. T. (2004). When to fire: Anticipatory versus postevent reconstrual of uncontrollable events. *Personality and Social Psychology Bulletin, 30*(3), 340–351.

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78*(5), 821–836.

Wiltermuth, S. S., & Flynn, F. J. (2013). Power, moral clarity, and punishment in the workplace. *Academy of Management Journal, 56*(4), 1002–1023.

Wolf, E. B., Lee, J. J., Sah, S., & Brooks, A. W. (2016). Managing perceptions of distress at work: Reframing emotion as passion. *Organizational Behavior and Human Decision Processes, 137*, 1–12.

Yeomans, M., Brooks, A. W., Huang, K., Minson, J., & Gino, F. (2019). It helps to ask: The cumulative benefits of asking follow-up questions. *Journal of Personality and Social Psychology, 117*(6), 1139–1144.