AMBIGUITY UNDER SCRUTINY: MORAL JUDGMENT OF MICROAGGRESSIONS

Joanna Korman
Bentley University

Boyoung Kim
George Mason University

Bertram F. Malle and David M. Sobel
Brown University

Microaggressions are subtle, offensive comments that are directed at minority group members and are characteristically ambiguous in meaning. In two studies, we explored how observers interpreted such ambiguous statements by comparing microaggressions to faux pas, offenses caused by the speaker having an incidental false belief. In Experiment 1, we compared third-party observers' blame and intentionality judgments of microaggressions with those for social faux pas. Despite judging neither microaggressions nor social faux pas to be definitively intentional, participants judged microaggressions as more blameworthy. In Experiment 2, microaggressions without explicit mental state information elicited a similar profile of judgments to those accompanied by explicit prejudiced or ignorant beliefs. Although they were, like faux pas, judged not to cause harm intentionally, microaggressive comments appeared to be judged more blameworthy on account of enduring prejudice thought to be lurking behind a speaker's false beliefs. Our current research demonstrates a distinctive profile of moral judgment for microaggressions.

Keywords: blame, intentionality, microaggressions

Some prejudice is easily detected. If a person utters a racial or gender slur toward a member of a minority group, they are making an explicit, aggressive remark; it stands to reason that they intend to denigrate the person and group in question, and to cause harm. Such remarks are easily recognized as deserving of moral condemnation. But as a large literature on modern, aversive, and symbolic racism has demonstrated, prejudicial beliefs can also be expressed in subtler ways (Conrey et al., 2005; Gaertner & Dovidio, 1986; Greenwald et al., 1998; Tarman & Sears, 2005; Williams, 2019). For instance, instead of explicitly stating that Asian Americans cannot be "real" Americans, a White individual might, right when he first meets a person of East Asian ethnicity, ask: "Where are you from?" Such *microaggressions* may indirectly reveal prejudicial beliefs without overtly stating them (Sue, 2010; Sue et al., 2007).

While microaggressive comments often cause harm to their recipients (e.g., Huynh, 2012; O'Keefe et al., 2014; Torres et al., 2010), interpreting the meaning and moral significance of these expressions is complex for social perceivers. Microaggressions can intentionally communicate a veiled negative message about members of a particular group (e.g., suggesting that East Asians cannot be "real" Americans), but they might also be interpreted as a more benign social faux pas. A social faux pas results when a speaker makes a comment that is intended to be positive or neutral ("What a cute couple you are!"), but causes accidental offense due to a mistaken background belief (e.g., the "couple" in question is actually a father and daughter). While faux pas stories were originally developed to detect advanced theory of mind deficits in individuals with autism spectrum disorder, typically developing adult observers easily understand that the speaker's background belief is simply mistaken in such cases (e.g., Baron-Cohen et al., 1999; Wang & Su, 2006). If microaggressions are viewed as similarly resulting from confusion or forgetfulness, they may yield a similarly benign interpretation.

The discriminatory behaviors resulting from modern, aversive, and symbolic racism have been widely studied (e.g., Daumeyer et al., 2019; Holroyd et al., 2017; Simon et al., 2019; Swim et al., 2003; Washington & Kelly, 2016). However, such investigations usually present participants with unambiguously discriminatory—even if unintentional—actions, and as such do not address a key characteristic feature of microaggressions: their ambiguous meaning (Sue et al., 2007). Potentially microaggressive statements may be ambiguous on several dimensions. First, it is often unclear as to whether such statements cause offense intentionally (are the result of overt, conscious prejudice toward outgroup members) or unintentionally (meant as neutral or positive comments that only inadvertently reveal prejudicial beliefs about a particular group). For example, consider a comment made by a White individual to an American of East Asian descent after the latter individual states she is from Milwaukee: "Where are you really from?" This comment may be understood on two levels: (1) as a neutral, if skeptical, question intended only to find out the person's place of origin or (2) as intended to communicate the prejudiced belief that all people in the United States of East Asian descent must be "from" somewhere other than the United States. Because both interpretations are plausible, it can be difficult for hearers and recipients to interpret these utterances.

Even if the ambiguity regarding the intentionality of the offense is resolved, if the offense is determined to be unintentional, a crucial second ambiguity may remain: The harm caused may be the result of either implicit prejudice, or of a more innocent "faux pas." For example, it is possible that the speaker's question resulted not from a general ignorance of the fact that individuals of East Asian descent are often born in the United States, but out of a more specific incidental error. Consider a case in which the speaker and recipient were boss and employee, respectively, and the boss's comment reflected a mistaken memory from reviewing personnel files that the employee in question was from Chicago, rather than Milwaukee. That is, if the boss thought his employee was really from Chicago, and not Milwaukee, the comment, "Where are you really from?" takes on quite a different meaning: The boss is making a request based on incidentally inaccurate information.

This discussion leads to a distinction between the harm caused by the outcome of the utterance and the blameworthiness of the utterance. That is, upon being asked "Where are you really from?" an individual might take offense, but whether the speaker deserves blame for causing that offense is a separate judgment. For social perceivers to blame a person for their actions, they must either infer or already know several key pieces of information about that action, including the harm caused by the action, the intentionality of the action that caused the harm, and the content of the specific mental states driving the intentions.

Most research on blame judgments for moral harm presents actions that describe unambiguous intentions and mental states. For example, the degree to which a harm is judged to be caused intentionally positively relates to the degree to which the perpetrator deserves blame (e.g., Young et al., 2007). In one case, a character named Grace puts a white substance in her friend's coffee and, consequently, her friend dies. In one version of the vignette, readers clearly understand that Grace killed her friend intentionally (the vignette states that she knew the substance to be toxic), while in another version, readers understand that she killed her friend unintentionally (the vignette states that she believed that the substance was harmless). Grace's mental states were unambiguous, and so was the harm that resulted from her action.

Previous research on social judgments of prejudicial acts has focused on acts that are unambiguously harmful, and either clearly intentional or unintentional (e.g., Daumeyer et al., 2019; Simon et al., 2019; Swim et al., 2003). These studies often ask participants to judge the speaker's beliefs about discrimination, which is presumed to be morally bereft. What these studies do not address is whether the speaker deserves blame for the harm caused by their action. Against this backdrop of prior research, microaggressions present a thorny new social-cognitive dilemma: Instances of possible prejudice may be interpreted as instances of egregious prejudice on the one hand or as entirely innocent misunderstandings on the other. The challenge of grasping the relations among the harm caused by the statement, the intentionality of the statement, and the belief states of the speaker make microaggressions difficult to interpret both within the field of psychology (e.g., Lilienfeld, 2017; Williams, 2019, 2020) and in society at large (e.g., Friedersdorf, 2015).

   The goal of the present study is to examine how the assumed relation between the intentionality of the offense caused by a potentially microaggressive remark and blame for that remark is complicated by the ambiguous nature both of the offense itself, and of the beliefs held by the individual making the statement. As discussed above, even when people can detect the offense caused by microaggressions, there is often ambiguity as to whether the offense was intentional. When the offense is not intended, there is further ambiguity as to whether the mental state of the speaker was merely incidentally mistaken, or bore evidence of more morally problematic ignorance or overt prejudice. In addition, it is not obvious how to measure the harm caused by microaggressions. While the harm caused by many morally bad actions can be described in concrete physical terms (e.g., a person is physically injured or dies; money is stolen from a bank), the harm caused by a negative remark is psychological. Such psychological harms can be considered from multiple perspectives, including that of the recipient of the microaggressive remark and that of a third-party observer of the speaker's behavior.

   Because they track considerations of the intentionality of an action as well as the specific contents of the underlying mental states, we regard blame judgments for potential microaggressive behaviors as uniquely informative. Crucially, to capture and explain the pattern of third-party observers' blame judgments for microaggressions, we presented participants with scenarios that represent the ambiguity characteristic of microaggressive remarks, and invited participants to rate these remarks on the relevant social-cognitive variables, including the intentionality of the speech-act, the offense (a proxy for psychological harm) it caused, and the blame that the speaker should receive for causing that offense. Of importance is whether microaggressive remarks show a profile of blame distinct from that of social faux pas—remarks that also cause offense, but do so due to an incidental error on the part of the speaker given a different set of prior beliefs about the situation. Asking someone, "Where are you from?" might be judged as causing offense either because it is understood as a clear faux pas or because it is understood as a possible microaggression. Do the pattern of blame and associated social-cognitive variables meaningfully differ based on this distinction? This was the question we asked in Experiment 1. In Experiment 2, we addressed whether distinct patterns of blame for microaggressions could be identified based on whether the speaker's relevant beliefs or background knowledge—either a prejudiced belief or morally problematic ignorance—was explicitly (and unambiguously) revealed to social perceivers, versus left unrevealed.

EXPERIMENT 1

Participants read stories describing either social faux pas (in two distinct conditions) or microaggressions. The situations depicted in the two faux pas conditions were identical except for the identity of the recipient. Thus, participants judged that the context of the same comment changed when directed at a majority group member versus a minority group member. We captured participants' profile of moral judgments for these remarks by eliciting ratings on three questions: how

much blame the speaker deserved for the remark, how offended the recipient was by the remark, and whether the speaker intended to offend the recipient.

METHOD

Participants. An a priori power analysis estimated a sample size based on a power of .80 and an alpha level of .05. Assuming a large effect size for standard one-way ANOVA, $f = 0.40$, the 18 cells of the experiment required 21 participants per cell (Cohen, 1992). We aimed to recruit 24 participants per cell to ensure our counter-balancing strategy. Due to a technical issue, we recruited 463 participants, which exceeded a planned sample size of 432, from Amazon Mechanical Turk (MTurk). Participants were compensated $0.15 for their participation. Excluding participants who failed to provide consent or answer any questions, the final sample consisted of 457 participants (153 male, 294 female, 1 other, 9 unknown, M_{age} = 38.12 years, SD_{age} = 13.25 years). Results of significance testing were unchanged when keeping only the first 21 participants for each cell, so we report results from the full sample.

The sample's ethnic background was distributed as follows: 40 participants reported as identifying as Hispanic/Latinx, 408 identified as not Hispanic/Latinx, and 9 participants did not identify their ethnicity. The racial distribution of the sample was as follows: 4 participants were American Indian or Alaska Native, 39 participants were Black/African American, 344 were Caucasian, 34 were Asian or Pacific Islander, 20 identified as mixed race, 7 identified as other, and 9 did not identify their race.

Materials. Each scenario depicted one character making a comment to another character based on a false belief. The intended interpretation of this comment as prejudicial or nonprejudicial differed across the three conditions: clear *microaggression* (comment with prejudicial connotation directed at a member of a minority group), *faux pas-ambiguous* (a comment ambiguous as to whether prejudicial or merely an accidental insult), and *faux pas-benign* (an identical comment but directed at a nonminority group member). In keeping with Sue et al. (2007), which focused on racial microaggressions but acknowledged the significance of microaggressions for other groups as well, scenarios in the microaggression condition included comments directed at individuals across a range of racial identities (e.g., Black, Latinx, American Indian, East Asian), and one scenario that depicted a female doctor of unspecified race.

For example, one story in the microaggression condition depicted Samantha, a cardiac surgeon who works in a large university hospital. Samantha is called into the hospital for an emergency consult on a Saturday. She enters the patient's room wearing a white coat. When she walks in, the patient asks, "When will the doctor be here?" In contrast, in the faux pas-ambiguous condition, an additional feature of the story provided an alternative interpretation of the comment as a nonprejudicial social faux pas. The patient makes the same comment to Samantha as in the microaggression condition, but instead of Samantha coming into the patient room wearing a white coat, she has been called in right after her Saturday run, and enters

wearing exercise clothing, covered in sweat. Finally, in the faux pas-benign condition, the recipient of the offender's comment is no longer a member of a minority group: The doctor in this scenario is male. Thus, in the faux pas-benign condition, the comment is presented as a social faux pas with no possible prejudicial interpretation. A complete list of scenarios is provided in the supplemental materials.

Design and Procedure. The three conditions (microaggression, faux pas-ambiguous, faux pas-benign) were presented in a fully between-subjects design. There were six scenario contents each formulated in three variants to represent each of the three conditions, for a total of 18 distinct stories. Each participant saw a single, randomly selected story. After reading the scenario, participants answered three dependent measures, presented in random order. They were instructed to indicate (1) how offended the hearer (e.g., "Samantha") was by the comment made by the speaker (e.g., the patient) on a scale from 0 (Not at all offended) to 6 (Extremely offended); how much blame the speaker deserved for what they said on a scale from 0 (No blame at all) to 6 (Maximum blame possible); and to what extent the speaker intended to offend the hearer on a scale from 0 (Not at all) to 6 (Very much). To capture the occurrence of microaggressions in real life, neither the degree of harm (i.e., offense) caused by microaggressive comments nor the intentionality of the offense was explicitly stated in the stories. Rather than assuming that participants would exhibit uniform judgments of the offense caused by microaggressions, we separately measured the two key components of blame judgments—perceived harm (offense) and intentionality. Lastly, we asked participants an open-ended question, "Why do you think the speaker (e.g., the patient) said that?"

RESULTS

Figure 1 shows the mean ratings of offense, blame, and intentionality. We performed linear mixed-effects model analyses separately on each rating, treating condition (three levels: microaggression, faux pas-ambiguous, faux pas-benign) as a fixed effect and scenario (six distinct contents) as a random effect. The first contrast compared the microaggression condition and the average of the two different faux pas conditions. The second compared the faux pas-ambiguous and the faux pas-benign conditions. We describe the results of each of these analyses below. Deidentified data and analysis code for this experiment and Experiment 2 can be found at https://osf.io/s964v/?view_only=b4fd80ebff6d407f9e60bf6b4e83de8e.

Offense Ratings. Offense ratings served as an indicator of the harm people believe to be caused by the comment in question. Participants rated the degree of offense in the microaggression condition ($M = 4.36$, $SD = 1.41$) higher than in the other two conditions ($M = 3.09$, $SD = 1.83$), $\beta = 0.33$, $SE = 0.04$, 95% confidence interval (95% CI) = [0.25, 0.42], $t = 7.77$, $p < .001$. They also rated the degree of offense as higher in the faux pas-ambiguous condition ($M = 3.42$, $SD = 1.75$) than in the faux pas-benign condition ($M = 2.78$, $SD = 1.86$), $\beta = 0.15$, $SE = 0.04$, 95% CI [0.07, 0.24], $t = 3.51$, $p < .001$.
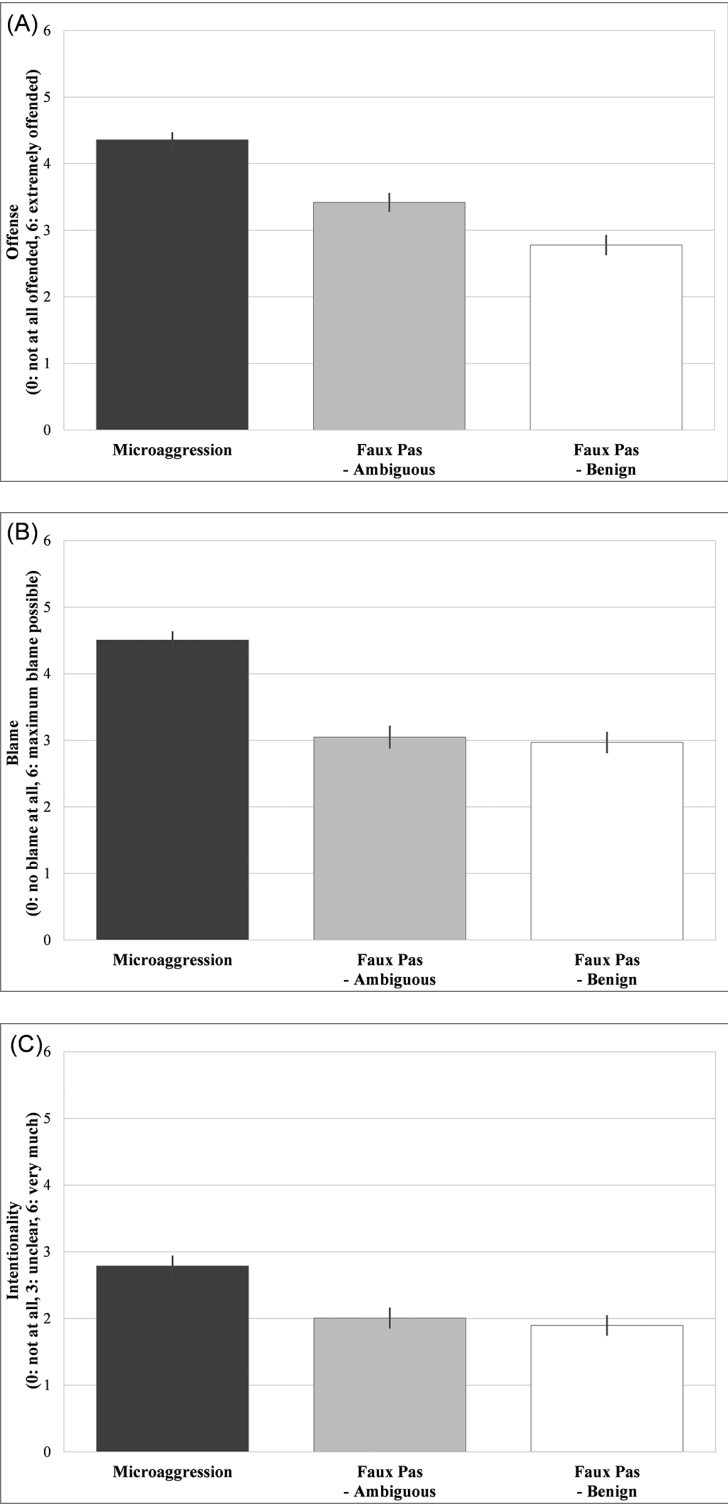
FIGURE 1. Experiment 1's mean judgment ratings: (A) Offense, (B) Blame, (C) Intentionality. Error bars represent standard error of the mean.

Intentionality Judgments. Participants rated the speaker's comment in the micro-aggression condition ($M = 2.79$, $SD = 1.93$) as more intentional than the comments in the other two faux pas conditions ($M = 1.95$, $SD = 1.91$), $\beta = 0.22$, $SE = 0.04$, 95% CI [0.13, 0.31], $t = 4.88$, $p < .001$. However, they still rated the intentionality of the comment in the microaggression below the midpoint (i.e., *unclear*), suggesting that they viewed the comment as more of an unintentional rather than intentional offense (see Figure 1C). There was no difference between the faux pas-ambiguous condition ($M = 2.01$, $SD = 1.92$) and the faux pas-benign condition ($M = 1.90$, $SD = 1.90$), $p = .23$.

Blame Judgments. As shown in Figure 1B, participants rated the speakers as deserving more blame in the microaggression condition ($M = 4.51$, $SD = 1.66$) compared to the average of the faux pas-ambiguous and faux pas-benign conditions ($M = 3.01$, $SD = 2.05$), $\beta = 0.36$, $SE = 0.04$, 95% CI [0.27, 0.44], $t = 8.40$, $p < .001$. There was no difference between the faux pas-ambiguous condition ($M = 3.05$, $SD = 2.09$) and the faux pas-benign condition ($M = 2.97$, $SD = 2.00$), $t = 0.89$ $p = .37$.

Intentionality as a Predictor of Blame. Numerous findings suggest that when an action leads to a negative outcome, people blame the action more harshly the more intentional it is (e.g., Monroe & Malle, 2019; Ohtsubo, 2007; Young et al., 2007). In the present experiment, it is possible that participants blamed speakers in the microaggression condition more than the speakers in the two faux pas conditions solely because they perceived the offense to be more intentional in the microaggression condition than in the faux pas conditions. By contrast, it is also possible that, while intentionality is one significant predictor of blame, it does not solely account for the observed differences in blame judgments between the microaggression and faux pas conditions. To investigate whether the differences in blame judgments among the conditions were solely related to the difference in intentionality judgments or whether the addition of condition (i. e., a comment was or was not a microaggression) uniquely affected blame judgments, we conducted a model comparison analysis. In the first mixed-effects linear regression model (Model$_{Baseline}$), we predicted blame ratings on the basis of intentionality only (included as a fixed effect) and scenario (included as a random effect). The second, full model (Model$_{Cond}$) included all predictors from the baseline model, but added condition as a fixed-effect parameter. We specified the contrasts for condition such that the first contrast compared the microaggression and the composite of the two faux pas conditions, and the second contrast compared the faux pas-ambiguous and faux pas-benign conditions.

The goodness of the model fit was evaluated based on Akaike information criterion (AIC; Akaike, 1974) values and log likelihood ratio (LR) test results. We also estimated the proportion of variance in the blame ratings explained by intentionality in Model$_{Baseline}$ and the proportion of variance explained by both intentionality and condition in Model$_{Cond}$ by computing the $R^2_{marginal}$, which provides an estimate of the variance explained by a fixed-effect parameter in a generalized mixed-effects model (Johnson, 2014; Nakagawa & Schielzeth, 2013; Nakagawa et al.,

2017). The difference between the $R^2_{marginal}$ value of $Model_{Baseline}$ and the $R^2_{marginal}$ value of $Model_{Cond}$ would show how much the explained variance in the blame judgments has increased by adding condition as an additional fixed-effect parameter to the model.

A comparison of these two models revealed that the second model explained participants' blame ratings significantly better than the first did, $\chi^2(4) = 50.41$, $p < .001$. When including only intentionality in the model, the AIC value was 1787.70 and the LR value was −889.87; but when condition was added to the model, the AIC value dropped to 1745.30 and the LR value was −864.66. Participants blamed speakers more harshly for microaggressive comments than for faux pas, even after accounting for their judgments of intentionality. Further, the $R^2_{marginal}$ value for $Model_{Baseline}$ was .28 and for $Model_{Cond}$ was .35, indicating an improvement in the proportion of the explained variance for blame ratings by adding condition into the model.[1] This suggests that blame judgments of microaggressions may be explained by factors apart from intentionality, and that these factors are unique to microaggressions. One possible explanation for this finding is that participants inferred that underlying microaggressive behaviors are prejudicial beliefs. We explored this possibility in Experiment 2.

EXPERIMENT 2

In Experiment 1, participants judged microaggressions to be more blameworthy than social faux pas, even though they were perceived as being unlikely to be motivated by an intent to offend. In Experiment 2, we explored the basis for these differences. We hypothesized that microaggressions were deemed more blameworthy than faux pas because participants inferred that the speaker held beliefs that either were overtly prejudicial or demonstrated a lack of awareness (ignorance) of key information in a way that implied prejudice when paired with a potentially microaggressive statement. For the purposes of the present analysis, we use the term *ignorant* to refer specifically to the speaker's lack of awareness of information that is relevant to the recipient's group membership. If people infer a prejudicial belief from microaggressive comments like those presented in Experiment 1, then their moral judgments for such comments should look similar to those for comments that are accompanied by an explicitly stated prejudicial or prejudice-implying ignorant belief.

We explored both explicitly prejudiced and prejudice-implying ignorant beliefs because an inference of either—or both—of these could underlie observers' inferences about the speaker's mental state when interpreting an utterance

---

1. Another factor that could have influenced judgments of blame is that the offense participants perceived was caused by the speaker's comment. Thus, we verified whether a unique contribution of condition in predicting blame ratings remained significant even after adding offense as an additional parameter to the baseline model. Condition remained a significant predictor even with these modifications. Adding condition to a baseline model including intentionality and offense ($Model_{Baseline'}$ vs. $Model_{Cond'}$) significantly better explained blame ratings, $\chi^2(8) = 43.16$, $p < .001$; $Model_{Baseline'}$: AIC = 1690, LR = −839.08, $R^2_{Marginal} = .43$; $Model_{Cond'}$: AIC =1663, LR = −817.50, $R^2_{Marginal} = .47$. Detailed analyses and results are reported on page 20 of the supplemental materials.

as a microaggression. On the one hand, the utterance may be blamed harshly by an observer only when the speaker is thought to harbor consciously prejudicial thoughts about which groups can or should be members of high-status professions (e.g., the speaker believes that doctors are, or should be, men). However, participants may make an alternate inference: that the speaker is ignorant about the minority group member's identity (e.g., a woman doctor) in the particular situation. Although ignorance of some specific details may be plausibly benign in the case of a faux pas (e.g., not realizing that a woman wearing exercise clothes is actually a doctor may seem an acceptable mistake to make), when the speaker's ignorance of a particular fact is due specifically to the recipient's status as the member of a minority group (e.g., not realizing that a woman wearing a white coat is actually a doctor), social perceivers may interpret it as implying prejudice, and therefore resulting in a similarly morally blameworthy microaggressive statement. Going forward, we will refer to the condition depicting prejudice-implying ignorant beliefs as the "ignorance" condition.

To test this hypothesis, Experiment 2 presented participants with a microaggression condition identical to that of Experiment 1, with no explicit mention of the speaker's belief, and compared it with two conditions in which the same comment was accompanied by an explicitly stated ignorant or prejudicial belief that participants may have inferred. A fourth condition was similar to the faux pas-ambiguous condition in Experiment 1, except the ambiguous belief participants may have inferred in Experiment 1 was explicitly stated. That is, the speaker's false belief about the identity of the recipient was stated (e.g., upon seeing Samantha, the patient assumes she can't be the doctor), but the reason for this belief remained ambiguous in the story. As in Experiment 1, there were two possible explanations: Either the mistake was due to the recipient's minority group status, or it was due to an alternative cause, stated earlier in the story, unrelated to group status (e.g., the fact that the recipient was atypically dressed for someone of that profession).

In addition, we included four additional "neutral comment" control conditions, which were parallel to each of the four microaggression conditions, except that the speaker's comment was unambiguously neutral (e.g., commenting on the weather). The neutral comment conditions where the speaker's either prejudicial or ignorant beliefs were explicitly stated would provide a baseline for people's judgments of objectionable beliefs alone without objectionable behaviors.

METHODS

Participants. Following Experiment 1, we recruited 24 participants per scenario within each condition based on a power analysis, assuming a power of .80 and an alpha level of .05 (Cohen, 1992). Three hundred and ninety-five participants were recruited from MTurk and each was randomly assigned to one of the 16 cells. Upon completion of the study, they were compensated with $0.15. One participant did not complete the consent form and 3 participants did not complete the study. After excluding responses from these 4 participants, we conducted analyses on data

from the remaining 391 participants (140 male, 242 female, 6 other, 3 unknown, $M_{age} = 35.55$, $SD_{age} = 12.91$).

The ethnicity of the sample was as follows: 23 participants reported as identifying as Hispanic/Latinx, 365 identified as not Hispanic/Latinx, and 3 participants did not identify their ethnicity. The racial distribution of the sample was as follows: 34 participants identified as Black/African American, 304 identified as Caucasian, 31 identified as Asian or Pacific Islander, 17 identified as mixed race, 2 identified as other, no participants identified as American Indian or Alaska Native, and 3 did not identify their race.

Materials. We created the vignettes used in the eight different conditions here by taking two of the scenarios used in Experiment 1 and pairing each with one of four possible mental states that we hypothesized people may infer upon hearing about a potentially microaggressive comment (prejudicial belief, ignorance, ambiguous false belief, and no explicit belief), with a comment that is either microaggressive (e.g., in the example about Samantha the doctor, "When will the doctor be here?") or neutral (e.g., in the same example, "Good morning, how are you today?").

The scenarios in the *microaggressive comment-no explicit belief* condition were the same as those in Experiment 1: The speaker made a microaggressive comment to the hearer but the speaker's mental state was not explicitly described. This condition thus represented a replication of data from the previous experiment and served as a contrast with the other conditions. In the *microaggressive comment-prejudicial belief* condition, the scenario explicitly described the speaker's belief, which was prejudicial. For example, in the doctor scenario, the participants read that "The patient has never seen Samantha before, but he generally assumes that doctors are men." In the *microaggressive comment-ignorant belief condition*, rather than holding an outright prejudicial belief, the speaker was described as being ignorant about the hearer's group membership. For instance, the participants read that "The patient has never seen Samantha before, and doesn't know who his doctor will be." The *microaggressive comment-ambiguous belief* condition was identical to the faux pas-ambiguous condition in Experiment 1 except that, after a description of a circumstance that could have led the patient to assume that Samantha was not the doctor (e.g., wearing exercise clothes), they read a sentence describing a false belief the patient had formed based on that circumstance, i.e., "Upon seeing Samantha, the patient assumes she can't be the doctor." Consistent with Experiment 1's *faux pas-ambiguous* condition, therefore, whether the speaker's comments had been derived from prejudicial/ignorant or nonprejudicial/nonignorant beliefs remained ambiguous in the *microaggression-ambiguous belief* condition.

The four neutral conditions—*neutral comment-no explicit belief, neutral comment-prejudicial belief, neutral comment-ignorant belief, neutral comment-ambiguous belief*— were the same as their counterparts in the microaggression conditions, except for the speaker's comment, which was benign (see the supplemental materials for all the stimuli).

Procedure. The procedure of Experiment 2 was identical to that of Experiment 1. Participants read the vignette based on the condition to which they were assigned and rated how offended the recipient was by the remark, how much blame the speaker deserved for the remark, and whether the speaker intended to offend the recipient, all on the same scales as in Experiment 1.

RESULTS

Figure 2 shows participants' mean ratings of offense, blame, and intentionality across the eight conditions. We again fitted separate linear mixed-effects models for offense, intentionality, and blame. Each model included condition as a fixed effect and scenario as a random effect. We set the contrast for the condition parameter so that the microaggressive comment-no explicit belief condition served as the reference level (the intercept of the model). This allowed us to contrast the microaggressive comment-no explicit belief condition with each level of the other seven conditions.

Offense Ratings. Participants rated the hearer in the microaggressive comment-no explicit belief condition to be much more offended by the comment compared to the hearer in each of the four conditions where neutral comments were made ($p < .001$). These results confirm the effectiveness of the type-of-comment (microaggressive vs. neutral) manipulation. Participants only judged the hearer to experience significant offense in the presence of a microaggressive comment.

More importantly, regardless of the type of belief with which a microaggressive comment was paired, participants expected the hearer to be offended to a similar degree (see Figure 2A). There was no statistical difference in ratings of the perceived offense between the microaggressive comment-prejudicial belief condition ($M = 3.76$, $SD = 1.74$) and the microaggressive comment-no explicit belief condition ($M = 4.04$, $SD = 1.54$, $p = .39$); nor was there a difference between the microaggressive comment-ignorant belief condition ($M = 3.78$, $SD = 1.75$) and the microaggressive comment-no explicit belief condition ($p = .43$).

Consistent with the findings from Experiment 1's faux pas-ambiguous condition, in the microaggressive comment-ambiguous belief condition ($M = 2.84$, $SD = 1.82$), participants thought the hearer would be less offended compared to in the microaggressive comment-no explicit belief condition ($M = 4.04$, $SD = 1.54$), $\beta = -0.19$, $SE = 0.05$, 95% CI $[-0.29, -0.09]$, $t = -3.66$, $p < .001$. Thus, when there was an idiosyncratic explanation for the microaggressive comment (e.g., a doctor wearing exercise clothes) and this information was consistent with a speaker's false belief, participants judged that knowledge of this information would mitigate the hearer's offense. To replicate the analysis from Experiment 1, we also conducted a separate analysis comparing the four conditions depicting a microaggressive comment. We specified the first-level contrast for condition as the ambiguous belief condition versus the average of the prejudicial, ignorant, and no explicit belief conditions; the second-level contrast for condition as the prejudicial belief
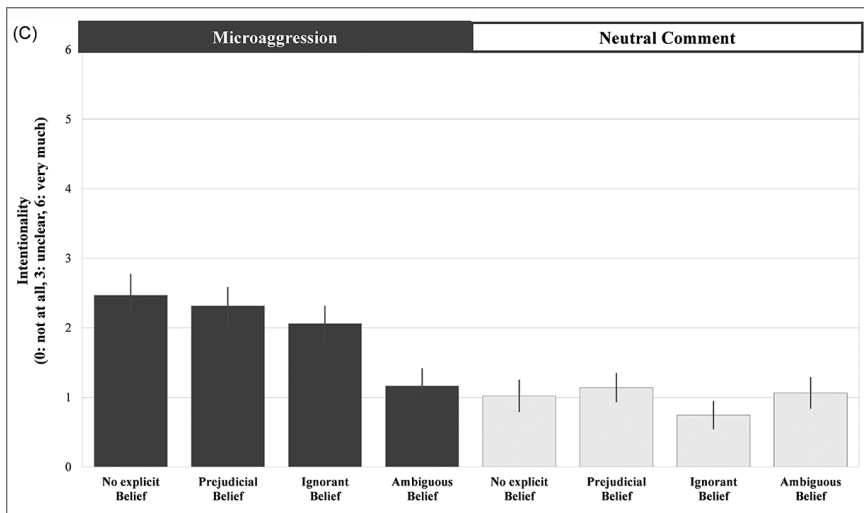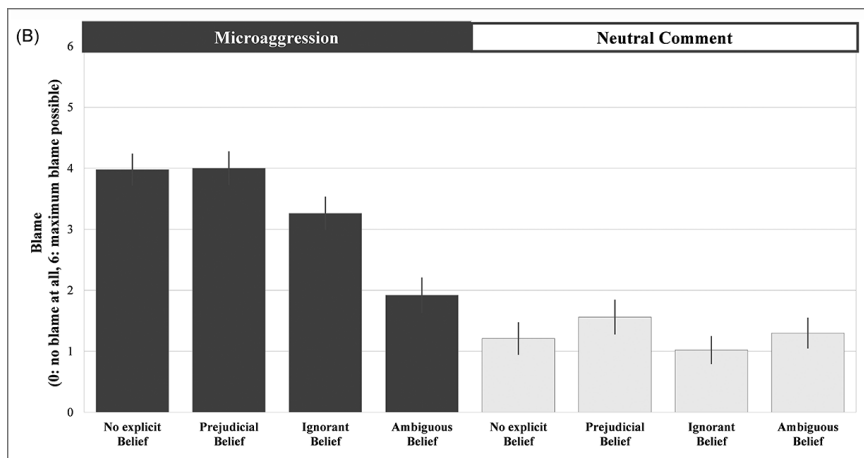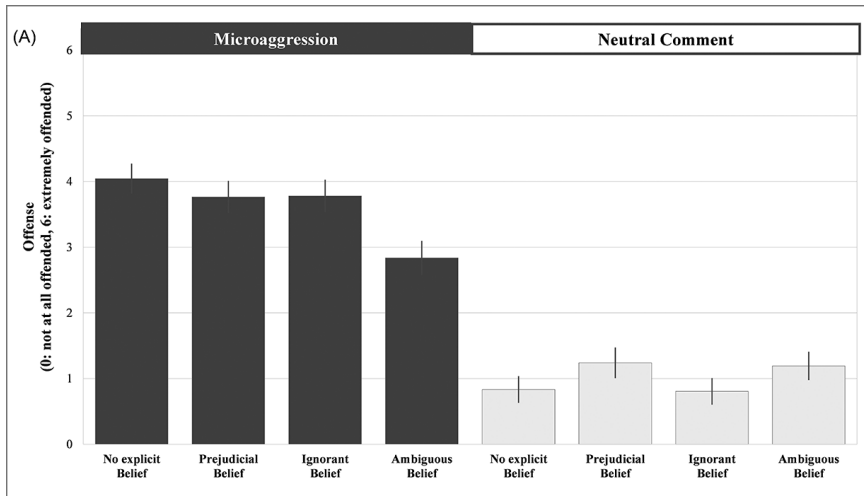
FIGURE 2. Experiment 2's mean judgment ratings: (A) Offense, (B) Blame, (C) Intentionality. Error bars represent standard errors of the means.

condition versus the average of the ignorant and no explicit belief conditions; and the third level contrast for condition as the ignorant belief condition versus the no explicit belief condition. The only effect we found was that participants judged the microaggressive comment-ambiguous belief condition to be less offensive than the composite of the other three conditions depicting a microaggressive comment, $\beta = -0.25$, $SE = 0.07$, 95% CI [-0.39, -0.12], $t = -3.66$, $p < .001$.

Finally, there was no difference among the four conditions in which the comment was neutral (neutral comment-no explicit belief $M = 0.83$, $SD = 1.40$; neutral comment-prejudicial belief $M = 1.24$, $SD = 1.66$; neutral comment-ignorant belief $M = 0.80$, $SD = 1.44$; neutral comment-ambiguous belief $M = 1.19$, $SD = 1.48$). In general, when the comment was neutral, participants rated the speaker's comment as less offensive than in the microaggression conditions. This was true even when the speaker's mental state was blatantly prejudiced. Participants recognized that when the speaker has a prejudicial belief that is concealed from the hearer, a neutral comment will not offend.

Intentionality Judgments. As in Experiment 1, average ratings of intentionality were below 3 on a scale ranging from 0 to 6 across all eight conditions (see Figure 2C). Thus, we replicated Experiment 1's findings that people perceive microaggressive comments as more likely to be unintentional offenses than intentional ones. However, as expected, people saw more intentionality in the microaggressive comment-no explicit belief condition ($M = 2.47$, $SD = 2.07$) than in any of the neutral comment conditions (all p values $< .001$; neutral comment-no explicit belief $M = 1.02$, $SD = 1.62$; neutral comment-prejudicial belief $M = 1.14$, $SD = 1.50$; neutral comment-ignorant belief $M = 0.75$, $SD = 1.44$; neutral comment-ambiguous belief $M = 1.06$, $SD = 1.57$). That is, even without receiving any explicit information about the beliefs that underlie microaggressive comments, participants judged them to cause offense more intentionally than benign comments.

Further, participants judged the microaggressive comment in the no explicit belief condition as similarly unintentional to the microaggressive comment in the prejudicial belief ($M = 2.31$, $SD = 1.95$, $p = .66$) and ignorant belief ($M = 2.06$, $SD = 1.81$, $p = .25$) conditions. In the microaggressive comment-ambiguous belief condition ($M = 1.16$, $SD = 1.78$), however, participants indicated that speakers intended to offend the hearers less than they did in the microaggressive comment-no explicit belief condition ($M = 2.47$, $SD = 2.07$), $\beta = -0.24$, $SE = 0.06$, 95% CI [-0.36, -0.11], $t = -3.69$, $p < .001$. Again, our separate analysis on the four microaggressive comment conditions revealed a statistically significant difference only between the microaggressive comment-ambiguous belief condition and the average of the other three microaggressive comment conditions, $\beta = -0.25$, $SE = 0.07$, 95% CI [-0.38, -0.11], $t = -3.59$, $p < .001$.

Blame Judgments and Relations Between Intentionality and Blame. We hypothesized that if participants inferred that the speaker had a prejudicial or ignorant belief from the microaggressive comment, they would still judge the comment to be blameworthy, even if it was not intended to offend. In support of this hypothesis,

participants' blame ratings in the microaggressive comment-no explicit belief condition ($M = 3.98$, $SD = 1.76$) were no different from those in the microaggressive-comment prejudicial belief ($M = 4.00$, $SD = 1.98$, $p = .95$) or the microaggressive comment-ignorant belief condition ($M = 3.26$, $SD = 1.95$, $p = .06$). In contrast, participants blamed the speaker less in the microaggressive comment-ambiguous belief condition ($M = 1.92$, $SD = 2.05$) than they did in the microaggressive comment-no explicit belief condition ($M = 3.98$, $SD = 1.76$), $\beta = -0.31$, $SE = 0.06$, 95% CI [−0.42, −0.20], $t = -5.43$, $p < .001$ (see Figure 2B).

As in Experiment 1, participants blamed the speaker in the microaggressive comment- ambiguous belief condition (similar to the faux pas-ambiguous condition in Experiment 1) less harshly than they did in the average of the other three microaggressive comment conditions, $\beta = -0.38$, $SE = 0.06$, 95% CI [−0.50, −0.25], $t = -5.84$, $p < .001$. When the explicit description of the speaker's belief did not offer a definitive answer to the question of whether the speaker had a prejudicial or nonprejudicial belief, then participants judged the comment as less blameworthy.

Finally, we found significant differences between the microaggressive comment-no explicit belief condition and each of the neutral comment conditions, including the neutral comment-prejudiced belief condition (all p values $< .001$; neutral comment-no explicit belief $M = 1.21$, $SD = 1.86$; neutral comment-prejudicial belief $M = 1.56$, $SD = 2.02$; neutral comment-ignorant belief $M = 1.02$, $SD = 1.64$; neutral comment-ambiguous belief $M = 1.30$, $SD = 1.73$). This indicates that in the absence of a comment that was overtly offensive, people made judgments of blame strictly based on the content of the comment (regardless of any mental states that were mentioned).

We also examined whether differences in blame judgments were better explained when considering only the perceived intentionality of the offense or whether the addition of condition had additional unique predictive power over and above intentionality alone. As in Experiment 1, we built two linear mixed-effect models: In the baseline model ($Model_{Baseline}$) we predicted blame ratings on the basis of intentionality only (included as a fixed effect) and scenario (included as a random effect). The second model ($Model_{Cond}$) included all predictors from the baseline model, but also added condition as a fixed effect. For this model comparison analysis, we only included the four conditions where a potentially microaggressive comment was made by the speaker. Further, because the blame ratings in the microaggressive comment-no explicit belief condition were no different from those in the microaggressive-comment prejudicial belief ($p = .95$) or the microaggressive comment-ignorant belief condition ($p = .06$), we collapsed the ratings from these conditions into a single level. Thus, condition in this model comparison analysis had two levels, the microaggression (no explicit belief, prejudicial, and ignorant combined) condition and the microaggressive comment-ambiguous belief condition. Consistent with the findings of Experiment 1, the second model was a better fit to the data than the first, $\chi^2(2) = 28.69$, $p < .001$. When including intentionality alone in the model, the AIC value was 764.72 and the LR value was −378.36; but when condition was added, the AIC dropped to 740.02 and the LR value was −364.01. When only intentionality was included as a fixed-effect parameter in the

model (Model$_{\text{Baseline}}$), the R^2_{marginal} value was .34. When condition was added to the model (Model$_{\text{Cond}}$), the R^2_{marginal} value increased to .42. This change in the R^2_{marginal} indicates that by adding condition to the baseline model, the proportion of variance explained by the fixed-effects parameters improved.[2] Blame judgments of microaggressions thus displayed a unique profile relative to those of (nonprejudicial) faux pas. As in Experiment 1, in Experiment 2, comparatively strong blame assigned to microaggressions was better explained when the model included the nature of the speaker's comment—as represented in the experiment's four conditions—in addition to the perceived intentionality with which the comment was made.

Exploration of Free Responses. In Experiment 2, we manipulated the content of the mental states that we expected participants to have inferred across the four main conditions (microaggression-no explicit belief, microaggression-prejudicial belief, microaggression-ignorant belief, microaggression-no ambiguous belief). If our manipulation in the microaggression-prejudicial and microaggression-ignorant belief conditions was successful, the participants in these two conditions would have interpreted the microaggressive comments in a manner similar to the interpretation they had derived for the microaggressive-explicit belief condition, but distinct from their interpretation of the microaggression-ambiguous belief condition. Our rating scales did not directly capture how participants interpreted the microaggressive comments in the four different conditions. Thus, we also analyzed the content of participants' open-ended responses to the question, "Why do you think the speaker said that?" for the scenarios about John the award winner and Samantha the doctor. To investigate whether these scenarios were interpreted as microaggressions, we verified whether participants recognized the targeted group membership (i.e., Black in John's story and female in Samantha's story) and made negative judgments of the microaggressive comment or the speaker.

Our free-response coding classified a response as falling under the *Group membership with negative judgments* code when the participants mentioned the targeted group membership (e.g., "Because John is Black," "Because she is a woman") and described either the speaker or their comment negatively, (e.g., "being prejudiced, biased, or a racist"). To compare the proportion of responses coded into the *Group membership with negative judgments* category in the microaggression-no explicit belief condition with that of each of the other three variants of the microaggression condition, we performed two-sample randomization tests (two-tailed, nsim = 300). The proportion of the responses coded into the *Group membership with negative judgments* category was significantly smaller in the microaggression-ambiguous belief condition (4.08%), compared to the microaggression-no explicit

2. We again verified whether a unique contribution of condition in predicting blame ratings remained significant after adding offense to the baseline model and subsequently to the comparison model including condition. Adding condition to a baseline model including intentionality and offense (Model$_{\text{Baseline'}}$ vs. Model$_{\text{Cond'}}$) significantly better explained blame ratings, $\chi^2(4) = 23.50$, $p < .001$; Model$_{\text{Baseline'}}$: AIC = 740.97, LR = −364.48, R^2_{Marginal} = .43; Model$_{\text{Cond'}}$: AIC = 725.47, LR = −352.73, R^2_{Marginal} = .48. Thus, the effect of adding condition in explaining blame ratings remained significant. Detailed analyses and results can be found on page 20 of the supplemental materials.

belief condition (40.00%, $p < .001$). Further, compared to the microaggression-no explicit belief condition (40.00%), there were a significantly larger proportion of responses falling into the *Group membership with negative judgments* category in the microaggression-prejudicial belief condition (76.47%, $p < .001$). We found no significant difference between the microaggression-no explicit belief condition (40.00%) and the microaggression-ignorant belief condition (42.00%, $p = .97$).

These results suggest that participants in the microaggression-no explicit belief, microaggression-prejudicial belief, and microaggression-ignorant belief conditions interpreted the speaker's utterance as an offensive comment directed at a member of a specific minority group. Notably this was not the case in the microaggression-ambiguous belief condition. In this latter condition, the presence of an alternate explanation—for example, the comment being made because the doctor was wearing exercise clothes, and not due to her gender—appears to have led participants to a more benign interpretation.

A similar analysis of Experiment 1 yielded a similar pattern of results. Compared to the faux pas-ambiguous condition (3.85%), the proportion of responses assigned to the *Group membership with negative judgments* category was significantly greater in the microaggression (42.31%) condition, $p < .001$. The frequency of this kind of justification in the microaggression condition was also significantly greater than in the faux pas-benign condition (0%, $p < .001$). Detailed analyses and results can be found on pages 21–22 in the supplemental materials.

GENERAL DISCUSSION

Prejudice and discrimination can be manifest overtly and covertly, and sometimes a statement or behavior has an offensive or discriminatory effect unintentionally. Yet while investigations of how such implicit prejudice is judged by social perceivers often take for granted a prejudicial or discriminatory interpretation, such clarity is less common in everyday life. Our studies highlighted this ambiguity by examining the contrast between microaggressions and social faux pas—comments similar to microaggressions in that they also cause offense and are the result of false beliefs, but distinct from them in that the beliefs involved are benign rather than prejudicial.

Two experiments profiled participants' judgments about the perceived offense, intention to offend, and blameworthiness of microaggressions. In Experiment 1, we contrasted clear microaggressions to social faux pas. We found a distinct profile for the blameworthiness of microaggressions based on their social-cognitive components. Participants judged microaggressive comments to be offenses that were not clearly motivated by an intent to offend, but that caused offense and were more blameworthy than social faux pas. Even though these latter utterances also cause offense, they were judged less blameworthy than microaggressions. But the speaker's intentions alone did not best account for these differences in blameworthiness. Rather, these differences were best explained when the nature of the utterance (microaggression or faux pas) was considered *in addition* to the speaker's intentions.

In Experiment 2, participants judged microaggressive comments that were paired with explicitly prejudiced beliefs as similarly blameworthy to those paired with ignorant beliefs, and more blameworthy than those paired with ambiguous beliefs (i.e., social faux pas). Moreover, participants judged the blameworthiness of a microaggressive statement caused by an explicit prejudicial belief as greater than that of a neutral statement, even if the speaker of the neutral statement was known to hold the same prejudicial belief.

These findings suggest that when no explanation other than the recipient's membership in a minority group is available, participants interpret a potentially microaggressive statement as a reflection of the speaker's prejudicial or ignorant (and prejudice-implying) belief. By contrast, when such an alternative explanation is available, the scenario is interpreted as a faux pas and is unlikely to be accompanied by the assumption of a prejudiced belief. For example, a patient may ask "Where's the doctor?" when he sees a sweaty man in jogging clothes walk into his exam room, only to realize a moment later that this person is his doctor. The faux pas, while unfortunate and perhaps momentarily embarrassing, is understandable given the situation: It was reasonable for him to expect the doctor not to be wearing sweaty jogging clothes. Furthermore, it would never occur to the speaker (or the observer of such a comment) that the individual he has encountered in this situation (the actual doctor) should not or would not be a doctor simply because he is male. But participants judge the same question ("Where's the doctor?") made to a female doctor in a white coat quite differently. In such a case, there is no available alternative explanation appealing to the doctor's attire; on the contrary, the female doctor is dressed just as one would expect a typical doctor to dress. Participants may thus blame the speaker more harshly because, while the speaker may not intend to express prejudiced or ignorant beliefs, participants may believe that such beliefs are below the surface. Even though the speaker's failure to recognize that the doctor was right in front of him may well have been a mistake, it is a mistake that (participants reason) may well have been caused by the speaker's prejudicial belief about whether women are or can be doctors.

In the present experiments, we often provided participants with privileged information and the speaker's intention or beliefs; in the real world, such information must be inferred from patterns of behavior. Our findings suggest how third-party observers might make such inferences upon witnessing microaggressions. When a member of a majority group says something to a member of a minority group that could be interpreted as a microaggression, the comment may be interpreted as arising from either the speaker's benign lack of knowledge (arising from an incidental or idiosyncratic false belief), or as the result of a prejudicial view of the characteristics of members of the minority group. As is evident in Experiment 2, participants blame speakers for uttering a microaggression whether they infer prejudice that is overt (believing that doctors cannot or should not be women) or more covert, tied to a morally problematic lack of awareness (e.g., a patient *should* easily recognize that the woman wearing a white coat who just walked into the room is actually his doctor, and the failure to do so suggests the patient also does not expect or believe women to be doctors).

These findings are also relevant to previous work on blame for negligent acts (Nuñez et al., 2014). That work shows that when an agent has knowledge of a set of circumstances that could lead to harm (e.g., a dinner guest has a peanut allergy), the agent is deemed negligent if they fail to exercise appropriate care to avoid that harm in their own actions (e.g., ensuring they do not use peanuts in food being given to the guest). Our findings extend this idea of blameworthy negligence to microaggressions. While a comment that reflects a failure of awareness may be considered by social perceivers as a benign faux pas if that failure is merely idiosyncratic (e.g., failing to realize that a woman in exercise clothes is actually a doctor), a microaggressive comment that reflects a more sustained belief about members of a particular group (seeing a woman in a white coat and failing to realize she could be the doctor) will be blamed more harshly. In other words, social perceivers appear to hold speakers to an accountability standard whereby they should not only exercise care in not causing physical harm, but also take care not to hold beliefs about that world that will lead them toward inaccurate conclusions (and actions and utterances that result from these conclusions) given particular evidence. Microaggressions, when they occur, may thus reveal a type of negligence in one's obligation not to hold prejudicial beliefs. The inferences participants appear to make about speakers' underlying beliefs when judging microaggressions also raise interesting possibilities for future work, including that participants will interpret subsequent statements made by speakers of microaggressions as more indicative of prejudicial beliefs than they will speakers whose statements were faux pas.

A limitation of the present experiments is that we did not explicitly describe the harm caused to the listener; this is also an avenue for subsequent investigation. Microaggressions have long been documented as an impactful phenomenon in the mental health of their recipients (e.g., Daumeyer et al., 2019). While previous work in moral psychology has shown that participants blame unintentional harms less harshly than intentional harms (e.g., Young et al., 2007), the unique pattern of blameworthiness related to microaggressions suggests that participants may also track distinct categories of unintentional harm. In particular, prejudicial mental states may drive relatively harsh blame judgments even in the absence of a negative outcome; for example, if a person makes a comment that a third-party observer views as a microaggression, but the intended recipient of the comment is not offended.

To conclude, the present experiments showed that participants judged microaggressive statements, like social faux pas, as offensive and unintentional. In contrast to faux pas, microaggressions were judged as more blameworthy, potentially because participants judged them as indicating ignorant or prejudicial beliefs on the part of the speaker, as opposed to the false belief on the part of one who commits a faux pas. Moreover, speakers were not blamed for neutral statements even if they held prejudicial beliefs. Although microaggressive statements may have appeared similar to mere faux pas on the surface, as both are driven by mistaken beliefs, social perceivers drew clear distinctions between these mistakes. While faux pas were interpreted as being driven by incidental, situation-specific mistakes, microaggressive mistakes were interpreted as being caused by erroneous

beliefs about minority groups as a whole—and were seen as grounded in prejudice. Our studies suggest that social perceivers believe that people should know better than to harbor such prejudiced beliefs, and they hold speakers morally accountable for microaggressive remarks that they take to result from such beliefs.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Baron-Cohen, S., O'Riordan, M., Jones, R., Stone, V., & Plaisted, K. (1999). A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *29*, 407–418.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi .org/10.1037//0033-2909.112.1.155

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: the quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*(4), 469–487. https://doi.org/10 .1037/0022-3514.89.4.469

Daumeyer, N., Onyeador, I., Brown, X., & Richeson, J. (2019). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, *84*, 103812. https:// doi.org/10.1016/j.jesp.2019.04.010

Friedersdorf, C. (2015, September). The rise of victimhood culture. *The Atlantic.* Retrieved February 2021, from https:// www.theatlantic.com/politics/archive/ 2015/09/the-rise-of-victimhood-culture /404794/

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Academic Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037//0022-3514 .74.6.1464

Holroyd, J., Scaife, R., & Stafford, T. (2017). What is implicit bias? *Philosophy Compass*, *12*(10), e12437. https://doi .org/10.1111/phc3.12437

Huynh, V. W. (2012). Ethnic microaggressions and the depressive and somatic symptoms of Latino and Asian American adolescents. *Journal of Youth and Adolescence*, *41*(7), 831–846.

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944–946.

Lilienfeld, S. O. (2017). Microaggressions: Strong claims, inadequate evidence. *Perspectives on Psychological Science*, *12*(1), 138–169. https://doi.org/10.1177/1745 691616659391

Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology*, *116*, 215–236. https://doi .org/10.1037/pspa0000137

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif .2017.0213

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Nuñez, N., Laurent, S. M., & Gray, J. (2014). Is negligence a first cousin to intentionality? Lay conceptions of negligence and its relation to intentionality. *Applied Cognitive Psychology*, *28*(1), 55–65. https:// doi.org/10.1002/acp.2957

Ohtsubo, Y. (2007). Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect 1. *Japanese Psychological Research*, *49*(2), 100–110.

O'Keefe, V. M., Wingate, L. R., Cole, A. B., Hollingsworth, D. W., & Tucker, R. P. (2014).

Seemingly harmless racial communications are not so harmless: Racial microaggressions lead to suicidal ideation by way of depression symptoms. *Suicide and Life-Threatening Behavior*, *45*, 567–576.

Simon, S., Moss, A., & O'Brien, L. (2019). Pick your perspective: Racial group membership and judgments of intent, harm, and discrimination. *Group Processes & Intergroup Relations*, *22*(2), 215–232. https://doi.org/10.1177/1368430217735576

Sue, D. W. (2010). *Microaggressions in everyday life: Race, gender, and sexual orientation.* John Wiley & Sons.

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, *62*(4), 271–286.

Swim, J. K., Scott, E. D., Sechrist, G. B., Campbell, B., & Stangor, C. (2003). The role of intent and harm in judgments of prejudice and discrimination. *Journal of Personality and Social Psychology*, *84*(5), 944–959. https://doi.org/10.1037/0022-3514.84.5.944

Tarman, C., & Sears, D. O. (2005). The conceptualization and measurement of symbolic racism. *Journal of Politics*, *67*, 731–761.

Torres, L., Driscoll, M., & Burrow, A. L. (2010). Racial microaggressions and psychological functioning among highly achieving African-Americans: A mixed-methods approach. *Journal of Social and Clinical Psychology*, *29*(10), 1074–1099.

Wang, Y., & Su, Y. (2006). Theory of mind in old adults: The performance on Happé's stories and faux pas stories. *Psychologia*, *49*(4), 228–237.

Washington, N., & Kelly, D. (2016). Who's responsible for this? Moral responsibility, externalism, and knowledge about implicit bias. In M. Brownstein & J. Saul (Eds.), *Implicit bias and philosophy, Vol. 2* (pp. 11–36). Oxford University Press.

Williams, M. T. (2019). Psychology cannot afford to ignore the many harms caused by microaggressions. *Perspectives on Psychological Science*, *15*(1), 38–43. https://doi.org/10.1177/1745691619893362

Williams, M. T. (2020). Microaggressions: Clarification, evidence, and impact. *Perspectives on Psychological Science*, *15*(1), 3–26. https://doi.org/10.1177/1745691619827499

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240.