Evolution and Human Behavior 34 (2013) 358–365

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

Seeing storms behind the clouds: Biases in the attribution of anger

Andrew Galperin [a], Daniel M.T. Fessler [c,e,*], Kerri L. Johnson [b,d,e], Martie G. Haselton [b,d,e]

[a] Oracle Corporation
[b] Department of Psychology, University of California at Los Angeles
[c] Department of Anthropology, University of California at Los Angeles
[d] Department of Communication Studies, University of California at Los Angeles
[e] Center for Behavior, Evolution, and Culture, University of California at Los Angeles

ARTICLE INFO

Article history:
Initial receipt 3 November 2012
Final revision received 21 June 2013

Keywords:
Anger
Emotions
Attribution
Cognitive bias
Error management

ABSTRACT

Anger-prone individuals are volatile and frequently dangerous. Accordingly, inferring the presence of this personality trait in others was important in ancestral human populations. This inference, made under uncertainty, can result in two types of errors: underestimation or overestimation of trait anger. Averaged over evolutionary time, underestimation will have been the more costly error, as the fitness decrements resulting from physical harm or death due to insufficient vigilance are greater than those resulting from lost social opportunities due to excessive caution. We therefore hypothesized that selection has favored an upwards bias in the estimation of others' trait anger relative to estimations of other traits not characterized by such an error asymmetry. Moreover, we hypothesized that additional attributes that i) make the actor more dangerous, or ii) make the observer more vulnerable increase the error asymmetry with regard to inferring anger-proneness, and should therefore correspondingly increase this overestimation bias. In Study 1 ($N = 161$), a fictitious individual portrayed in a vignette was judged to have higher trait anger than trait disgust, and trait anger ratings were more responsive than trait disgust ratings to behavioral cues of emotionality. In Study 2 ($N = 335$), participants viewed images of angry or fearful faces. The interaction of factors indicating target's formidability (male sex), target's intent to harm (direct gaze), and perceiver's vulnerability (female sex or high belief in a dangerous world) increased ratings of the target's trait anger but not trait fear.

1. Introduction

Assessing others' personality traits is a key adaptive problem that social cognition evolved to address. Understanding people's personalities allows us to predict others' future behavior and facilitates navigating complex social interactions (Ross, 1977). However, because personality is invisible, it is difficult to assess. Past behavior may reveal underlying traits, but inferences about them (especially from a single observation) are highly uncertain, for two reasons. First, behaviors are produced not only by enduring dispositions, but also by fleeting situations. Proper discounting of situational influences requires repeated observations of an individual across multiple situations (Kelley, 1972), and this cannot always be achieved. Second, people strategically manage their behaviors, at times actively inhibiting the expression of negative traits and compromising observers' ability to discern personal characteristics.

Here, we explore the hypothesis that assessments of an individual's propensity to become angry are adaptively biased. Given that i) conspecifics were a primary source of danger for ancestral humans (Keeley, 1996), and ii) anger motivates violence (Fessler, 2010; Frank, 1988; Sell, 2009), an important adaptive challenge was predicting an individual's enduring inclination to become angry (i.e., trait anger), a process we term "anger attribution". Importantly, anger attribution is inherently imperfect, making complete accuracy unlikely, if not impossible.

1.1. Adaptive rationality and error management

The "adaptive rationality" approach contends that the mind was shaped by selection to enhance fitness in ancestral environments rather than to yield accurate judgments (Haselton et al., 2009; see also Funder, 1995, and Krueger & Funder, 2004). Therefore, human cognition can manifest seemingly irrational biases that are, in fact, "adaptively rational." Anger attribution is one domain in which this might occur. Perceivers can commit one of two errors: underestimate an individual's trait anger (false negative) or overestimate it (false positive). On average, underestimations will have been costlier than overestimations in ancestral populations: assuming that an anger-prone individual was temperate placed the perceiver at risk of assault, whereas assuming that a temperate individual was anger-prone merely led to foregoing potentially profitable interactions. Thus, overall accuracy (i.e., committing false negative and false positive errors with equal frequency) did not maximize fitness over evolutionary time. Rather, in line with error management theory (Haselton

* Corresponding author. Department of Anthropology, University of California, Los Angeles, 341 Haines Hall, Los Angeles, CA 90095-1553.
E-mail address: dfessler@anthro.ucla.edu (D.M.T. Fessler).

1090-5138/$ – see front matter
http://dx.doi.org/10.1016/j.evolhumbehav.2013.06.003

& Buss, 2000; Haselton & Nettle, 2006), we hypothesize that selection favored a biased tendency to commit the less costly false positive — overestimating trait anger. Although the same logic applies to the estimations of *state* anger, our predictions focus squarely on *trait* anger because traits predict future behavior, and it is costly to underestimate an individual's anger not only in the moment, but also in future interactions.

Absent objective baselines, investigating a hypothesized bias in judgment requires points of comparison; we employed other negative emotional dispositions, for which we predicted either no biases, or reverse biases (trait *underestimation*). For instance, in the case of fear directed toward the perceiver, there is no clear asymmetry in the costs of underestimating or overestimating another's propensity to experience fear. Therefore, we do not expect an evolved bias for perceptions of trait fear. If a target displays fear or disgust toward something or someone other than the perceiver, it was likely to have been adaptive to over-attribute their emotions to the situation (and underestimate the corresponding trait), since this enhances alertness to potential hazards. More formally:

Hypothesis 1. Behaviors indicative of anger will be attributed to personality to a greater degree than behaviors indicative of other negative emotions.

Ancestral error cost asymmetries were not static, but instead varied by context (Haselton & Galperin, 2013; Johnson, Blumstein, Fowler, & Haselton, 2013). Psychological adaptations formed by these variable asymmetries should therefore be influenced by contextual cues. Specifically, cues that a person is able or likely to aggress against the perceiver increase the costs of underestimating trait anger. In turn, this exaggerated error asymmetry would have made erring on the side of caution (i.e., overestimating trait anger) even more beneficial, leading to an exaggerated dispositional bias. Cues that someone poses a threat include attributes of the target individual (e.g., formidability; gaze direction), attributes of the perceiver (e.g., self-perceived vulnerability), or a combination thereof. These factors should not affect assessments of other emotion traits because they do not affect the relevant error cost asymmetries. More formally:

Hypothesis 2. Increasing the danger that the target poses to the perceiver will increase dispositional attributions of angry behaviors but will not increase dispositional attributions of behaviors associated with other negative emotions.

2. Study 1

We tested the possibility that, ceteris paribus, an unfamiliar individual would be viewed as more dispositionally prone to anger than to another negative emotion (disgust). Participants read vignettes about a fictitious man who reacted with anger and disgust to situations commonly eliciting each emotion, then rated the protagonist's trait anger and disgust. We predicted that the man's trait anger would be rated higher than his trait disgust. In testing this prediction, we sought to address an alternative explanation: compared to a single display of disgust, a single display of anger may indeed be more informative about an individual's personality, such that the predicted pattern of results is potentially explicable in terms of the accuracy of folk psychology. This is plausible because, being more proscribed than disgust displays, anger displays must overcome a higher inhibitory threshold, hence someone who is angry enough to show it might be anger-prone. However, this logic no longer holds when the observer views the eliciting situation as meriting an angry response. We therefore measured and controlled for the protagonist's perceived "overreaction," thus leveling the playing field for anger and disgust.

Hypothesis 1 thus translates as Prediction 1. The target's trait anger will be rated higher than his trait disgust, and will remain so even after controlling for any systematic discrepancy between the perceived appropriateness of his anger and disgust reactions.

We predicted that perceived trait anger would positively scale with perceived state anger in a seemingly irrational manner. If someone overreacts to a situation and becomes enraged, this is objectively informative about their underlying trait anger. However, if an angry response is merited, the event is not dispositionally informative: there is no rational reason to attribute the anger to disposition because any normal person would have acted thusly. We predicted that, because of the greater cost of underestimating anger, observers would nevertheless produce overly dispositional attributions, as it is safer to assume that the anger, though justified, is dispositional. We therefore predicted that even justified anger would lead to dispositional attribution, whereas disgust would lead to dispositional attribution only to the extent that it was seen as an unjustified overreaction.

Hypothesis 2 therefore translates as Prediction 2. Ratings of "overreaction" will fully mediate the positive association between state and trait ratings for disgust, but will not fully mediate this association for anger (i.e., there will be residual bias in attributions of anger but not disgust).

We predicted full, rather than merely partial-but-stronger mediation for disgust because anything less than full mediation indicates a bias. If judgments are normatively rational, and the target is perceived to be reacting appropriately to the stimulus, there should be zero correlation between states and corresponding traits. Since we proposed that disgust should follow this normative rule, we expected any positive correlation between perceived state and trait disgust to be entirely indirect (i.e., fully mediated by the overreaction factor).

2.1. Methods

2.1.1. Participants and procedure

To prevent trait and state ratings from being artificially similar, participation occurred in two sessions held on different days. In exchange for course credit, 441 UCLA undergraduates from two Introductory Psychology classes completed the first session and were provided with a unique identifier. They were subsequently invited to participate in the second session online. Over the next two months 161 of the participants completed the online survey; these individuals constitute the sample. Participation in the second session ranged from 15 to 66 days after the first session ($M = 24.8$, $SD = 14.5$); the time elapsed between sessions was not associated with any variables of interest ($ps > .11$). Participant sex and other demographics were not assessed (a limitation addressed in Study 2).

2.1.2. Materials

In Session 1, participants read two of four vignettes describing a fictitious male college student. A male target was chosen to provide a strong initial test of the trait attribution bias hypothesis. Men are disproportionately responsible for violence (Daly & Wilson, 1988), hence error management effects in judging trait anger should be most pronounced for male targets.

Vignettes described the protagonist in situations that would provoke reactions of both anger and contamination disgust in most people (see supplementary material, available on the journal's website at www.ehbonline.org). Each participant read one "weak" vignette, in which the protagonist reacted to a mildly anger- and disgust-provoking situation with mild anger and disgust. Each participant also read one "strong" vignette, in which the protagonist reacted to more serious provocations of anger and disgust with appropriately intense anger and disgust. Thus, the individual was implicitly portrayed as an average, reasonable person in terms of how easily he becomes angered or disgusted in a range of situations. No vignette contained the words "anger," "disgust," or

synonyms thereof. Half of the participants read one pair of weak and strong vignettes (in randomized order); the other half read the other pair of weak and strong vignettes. Participants then rated the target's trait anger and disgust (in randomized order) relative to the average person on 1–9 scales, anchored by "much less angry (disgusted) than the average person" and "much more angry (disgusted) than the average person." Instructions specified rating contamination disgust and not moral outrage (Rozin, Haidt, & McCauley, 2000; Tybur, Lieberman, & Griskevicius, 2009).

In Session 2, which occurred between two and eight weeks after Session 1, participants read the same vignettes as before. They rated the absolute degree of the target's state anger and disgust on 1 to 9 scales, ranging from "not at all" to "extremely." They also rated how justified his reaction was, given the situation, on a −3 to 3 scale, ranging from "extreme underreaction" to "extreme overreaction." The latter measure allowed us to assess the degree to which participants viewed the target's reaction as justified, as well as to control for any unintended bias in the vignettes (e.g., having inadvertently portrayed the individual as easily disgusted rather than average).

2.2. Results

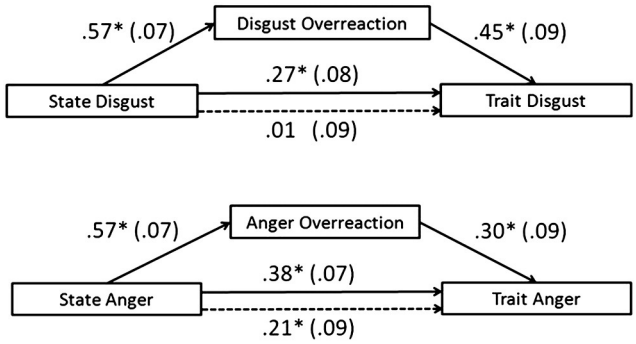
Participants judged the target to have displayed state anger and disgust at just above the scale midpoint (anger, $M = 6.08, SD = 1.35$; disgust, $M = 5.95$, $SD = 1.32$); these means did not statistically differ, $t(159) = 1.85$, $p = .07$. Participants also rated the target as mildly overreacting in terms of both anger ($M = .55, SD = 1.00$; one-sample against 0 $t(160) = 7.01$, $p < .001$) and disgust ($M = .25$, $SD = .90$; one-sample against 0 $t(160) = 3.47$, $p < .001$). The anger overreaction was stronger than the disgust overreaction, paired-samples $t(160) = 5.36$, $p < .001$.

Prediction 1. The target's trait anger will be rated higher than his trait disgust, and will remain so even after controlling for any systematic discrepancy between the perceived appropriateness of his anger and disgust reactions.

Before controlling for overreaction, ratings of trait anger ($M = 5.94$, $SD = 1.24$) were higher than those of trait disgust ($M = 5.57$, $SD = 1.16$), $t(160) = 3.88$, $p < .001$. Because measures were nested within participants, we used multilevel regression (HLM 7.0) to examine whether this difference remained significant after controlling for perceived overreaction. We regressed Trait Emotion Ratings onto Level 1 predictors that included Emotion Type (anger or disgust; dummy coded) and perceptions of the protagonist's Behavioral Overreaction. Unsurprisingly, the more that participants perceived the target as overreacting in terms of either emotion, the more they rated him as dispositionally inclined to experience that emotion ($B = 0.48$, $SE(B) = .09$, $t(160) = 5.51$, $p < .001$). Nevertheless, supporting Prediction 1.1, even with this variable controlled, the type of emotion was still significantly associated with the magnitude of the trait rating ($B = 0.21$, $SE(B) = .10$, $t(160) = 2.01$, $p = .046$), such that ratings were higher for marginal trait anger than for marginal trait disgust.

Prediction 2. Ratings of "overreaction" will fully mediate the positive association between state and trait ratings for disgust, but will not fully mediate this association for anger (i.e., there will be residual bias in attributions of anger but not disgust).

Two mediational models were run per standard techniques (Kenny, Kashy, & Bolger, 1998). Supporting Prediction 2.1, overreaction only partially mediated the total effect of state on trait ratings for anger ($c' = .21$, Sobel $z = 3.20$, $p = .001$), but fully mediated this effect for disgust ($c' = .01$, Sobel $z = 4.44$, $p < .001$); see Fig. 1. Thus, even after accounting for overreaction, participants continued to scale their trait anger ratings with their state anger ratings (which constitutes a bias), but did not do so for disgust.



Note. The standard errors of the regression coefficients are in parentheses.
* $p < .05$

Fig. 1. Standardized regression coefficients for the relationship between ratings of state and trait emotion as mediated by perceived overreaction in Study 1.

2.3. Discussion

Supporting Hypothesis 1 – that displays of anger will be viewed as more revealing of disposition than displays of other emotions – participants attributed more enduring anger than enduring disgust to a male protagonist, even after we accounted for systematic differences between perceptions of his state anger and disgust. In Study 2, to examine how the target's gender interacts with this main effect, we used female as well as male targets.

Supporting Hypothesis 2 – that the bias toward attributing anger to disposition will increase with the danger posed by the given individual – participants made increasingly dispositional attributions as the perceived level of anger displayed by the individual increased, regardless of how justified his emotional reaction was seen as being; the same was not true of disgust. These patterns are consonant with an evolved error management bias.

As noted earlier, absent objective baselines, tests of error management hypotheses rely on points of comparison in testing for predicted biases. Disgust, a negative emotion that resembles anger in multiple respects (Smith & Ellsworth, 1985), performed this role in Study 1. To demonstrate that the supportive evidence obtained in Study 1 was not an artifact of one particular comparison emotion, in Study 2 we used fear – which differs greatly from both anger and disgust (Smith & Ellsworth, 1985) – as the negative emotion control.

A main effect comparison of scale ratings of trait anger and any other negative emotion can be difficult to interpret. Although we controlled for perceived overreaction in Study 1, this may be imperfect, since participants might have difficulty translating the relevant cognitions into propositional statements regarding the degree of overreaction. This underscores the importance of introducing additional manipulations hypothesized to affect the ratings of trait anger but not of other negative emotions, a key piece of our framework explored in Study 2.

3. Study 2

In Study 2, we tested Hypothesis 1 using a new comparison emotion (fear), and tested Hypothesis 2 by manipulating the danger posed by the target to the perceiver. Participants viewed photographs of faces that varied by sex, eye gaze direction (direct/averted), and emotion (anger/fear). Participants rated the trait and state levels for each emotion. This allowed us to test multiple subsidiary predictions. Per Hypothesis 1, we expected that, collapsed across manipulations, dispositional anger ratings would be higher than dispositional fear ratings. Moreover, as in Study 1, we expected this difference to be significant even after accounting for the perceived strength of the anger and fear expressions. Controlling for this source of normatively logical inferences about the targets' emotional traits ensures that any

remaining difference between the ratings of trait anger and trait fear constitutes a bias.

Hypothesis 1 thus translates as Prediction 1. Across conditions, dispositional anger ratings will be higher than dispositional fear ratings even after controlling for any systematic differences in the perceived state intensity of the anger and fear expressions.

Hypothesis 2 specifies that the degree of bias in anger attribution will be contingent on the danger posed by the target. Men generally pose a greater threat of violence than do women (Daly & Wilson, 1988) and are treated accordingly by hazard-avoidance mechanisms: for instance, fear learned in conjunction with an outgroup face is less easily extinguished when the face is male (Navarrete, Olsson, Ho, Mendes, Thomsen, & Sidanius, 2009). On average, underestimating a man's propensity to experience anger will be especially costly; the same is not true of fear.

Hypothesis 2 thus translates as Prediction 2a. The difference between dispositional anger and dispositional fear ratings will be higher for male than for female targets even after controlling for any systematic differences in the perceived state intensity of the anger and fear expressions.

Although, empirically, men do not become angry more frequently or more intensely than women, folk models nevertheless depict this, along with corresponding dispositional differences (Fischer & Evers, 2010). A positive result for Prediction 2a could therefore reflect the influence of gender stereotypes, hence it is important to augment tests of Hypothesis 2. An emotional expression coupled with direct gaze usually signals that the emotion is directed *toward* the perceiver (Adams & Kleck, 2003). In the case of anger, direct gaze indicates that the target likely harbors harmful intentions toward the perceiver — a possibility that is hazardous for the perceiver to ignore both in the moment and in future interactions. In such circumstances, it is especially costly for the perceiver to underestimate the target's anger-proneness. The same is not true, however, for fearful expressions. Per Hypothesis 2, we therefore expected that direct gaze would enhance the bias toward a dispositional interpretation when paired with anger expressions, but not when paired with fear expressions. (Note that a shift in gaze is a transient behavior and provides no normative information about the target's enduring traits. Thus, if anger attribution were affected by gaze as predicted, this would constitute evidence for a bias.)

Target's sex and eye gaze should interact to influence judgments of dispositional anger, as a potentially dangerous man indicating via direct gaze that he is angry at the observer presents an especially potent combination of danger cues. Furthermore, the impact of these factors should vary with the perceiver's vulnerability to assault. Because women are less physically formidable than men, they should be especially sensitive to interpersonal cues of danger.

Hypothesis 2 thus translates as Prediction 2b. There will be a four-way interaction between emotion condition (anger or fear), the participant's sex, the target's sex, and the target's eye gaze, such that, to a greater extent than male participants, female participants will rate male targets expressing anger with direct gaze as more predisposed toward anger than male targets expressing anger with averted gaze. This contrast will not be significant in the fear condition.

More generally, because natural selection weighs the benefits of precaution against its costs, psychological adaptations that serve to protect against violence can be expected to calibrate to individual differences in the susceptibility to aggression (cf. Snyder, Fessler, Tiokhin, Frederick, Lee, & Navarrete, 2011). *Self-perceived* vulnerability in particular is crucial. This is because the costs of encountering an antagonist depend in part on the individual and social resources that the actor brings to bear in coping with the hazard. Because individuals differ in these regards, the asymmetry in the costs of errors in anger attribution will vary as a function of both the objective baseline risk of assault in the individual's environment and the individual's capacity for coping with that risk. Subjective perceptions of the level of danger in the world plausibly reflect the combination of past encounters with danger and self-assessed capabilities for addressing it (Johns, 2011; Snyder et al., 2011). Accordingly, if the bias at issue is adjusted as a function of its utility for the individual, then this trait should be positively correlated with the extent to which the individual perceives the world to be dangerous.

This generates Prediction 2c. There will be a four-way interaction between emotion condition (anger or fear), the participant's self-perceived vulnerability, the target's sex, and the target's eye gaze such that, to a greater extent than less vulnerable individuals, more vulnerable individuals will rate male targets expressing anger with direct gaze as more predisposed toward anger than male targets expressing anger with averted gaze. This contrast will not be significant in the fear condition.

3.1. Methods

3.1.1. Participants

Via Amazon.com's Mechanical Turk, 372 U.S. participants (200 women, 147 men, 25 who did not specify their sex) were recruited for a 10-min online study of "perceptions of individuals" in exchange for $0.20. Software prevented repeat participation from any given computer. The anger condition ($N = 161$) was run in its entirety prior to the fear condition ($N = 211$), with identical recruitment procedures. The average age was 34.8 ($SD = 12.8$); 73% of participants were White.

3.1.2. Stimuli

Images were selected from the NimStim face set (Tottenham et al., 2009), which contains angry, fearful, and neutral faces posed by the same individuals. We selected four female and four male targets from faces identified by Tottenham et al. as having the most readily identifiable anger expressions. The same targets were later used in the fear condition.

Using the website www.faceresearch.org, we manipulated the extremity of the facial expressions by blending varying doses of the target's angry or fearful expression and the target's neutral expression; participants viewed these blended images, not the original images.

To create averted gaze, angry, fearful, and neutral images were digitally altered by moving the irises and pupils to the right side of each eye. These images and the unaltered images were then duplicated and flipped along the Y-axis for counterbalancing. Participants saw one of four image types: direct-gaze original, direct-gaze flipped, averted-gaze right, and averted-gaze left (i.e., averted-gaze right flipped). In all analyses, the two direct-gaze conditions were collapsed into one condition, as were the two averted-gaze conditions.

3.1.3. Design and Measures

The design of the study was 2 (angry or fearful faces: between-subjects) × 2 (direct or averted gaze: between-subjects) × 2 (target sex: within subjects). To avoid arousing suspicion regarding the nature of our manipulations, emotion and gaze varied between subjects. Each participant thus viewed and rated each of the eight target individuals' images in randomized order, all of which were either angry or fearful, and all of which displayed either direct or averted gaze. All measures and tasks were completed for each target individual before the participant saw an image of the next target; see supplementary material (available on the journal's website at www.ehbonline.org) for a sample image set and trial.

3.1.3.1. Image ratings. Each of the eight images was presented individually and appeared on screen for the duration of the participant's ratings of the respective target individual. The degree of anger or fear in the image was randomized among 70%, 80%, or 90% of the original angry

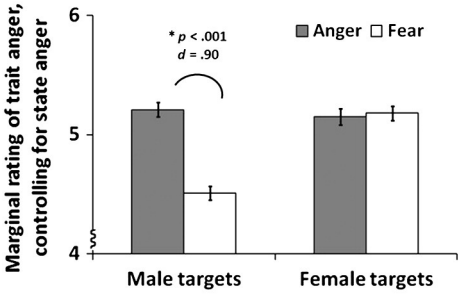


Fig. 2. The effects of targets' sex on participants' dispositional anger and fear ratings, controlling for participants' explicit ratings of state emotional intensity in the images in Study 2.

or fearful expression. Using 9-point scales anchored by "not at all" and "extremely," participants first provided an explicit assessment of each target's current emotional state ("How angry/scared does the person look in this picture?"). Then, on 9-point scales anchored by "much less than average" and "much more than average", participants inferred each target's enduring emotional trait ("Compared to the average person, how *often* do you think this person becomes angry/scared in real life?; Compared to the average person, how *easily* do you think this person becomes angry/scared in real life?"; $\alpha = 0.91$).

3.1.3.2. Frame-matching task. Next, participants completed an exploratory perceptual matching task tangential to the current topic (see supplementary material, available on the journal's website at www.ehbonline.org).

3.1.3.3. Demographics. Participants next reported their sex, age, and ethnicity. To assess self-perceived vulnerability to threat, participants then completed the Belief in a Dangerous World scale (BDW; Altemeyer, 1998), which contains 12 items ($\alpha = 0.89$) probing the extent to which the respondent thinks others are violent and life is full of hazards, on 5-point disagree–agree scales.

3.2. Results

Prediction 1. Across conditions, dispositional anger ratings will be higher than dispositional fear ratings even after controlling for any systematic differences in the state intensity of the anger and fear expressions.

Collapsing across conditions, we conducted a one-way ANCOVA predicting the trait rating (averaged across all eight targets) from the emotion condition (anger or fear) while controlling for averaged state emotion rating as a continuous covariate. Controlling for the state rating was necessary because it was higher for the anger images ($M = 5.16$, $SD = 1.15$) than for the fear images ($M = 4.54$, $SD = 1.06$), $t(370) = 5.45$, $p < .001$, and, as expected, state ratings were positively associated with the trait ratings in the ANCOVA, $F(1, 331) = 158.82$, $p < .001$. After controlling for the state ratings, the difference between the marginal means for ratings of trait anger ($M = 5.17$) and trait fear ($M = 4.85$) remained robust, $F(1, 368) = 17.53$, $p < .001$, supporting Prediction 1.2.

Prediction 2a. The difference between dispositional anger and fear ratings will be higher for male than for female targets even after controlling for any systematic differences in the state intensity of the anger and fear expressions.

Each participant's trait ratings were averaged across the four female targets and the four male targets. To test whether the differences between the ratings of trait anger and trait fear differed in magnitude for female and male targets, we ran a multilevel analysis. Trait Rating was regressed on Emotion Type (Level 2: fear = 0, anger = 1), Target Sex (Level 1: female = 0, male = 1), State Rating (Level 1, grand-mean centered), and the cross-level interaction of Emotion Type × Target Sex. This cross-level interaction was significant ($B = .81$, $p < .001$). Simple slopes for the association between Emotion Type and Trait Rating differed for female and male targets. The association between Emotion Type and Trait Rating was not significant for female targets ($B = -.08$, $p = .30$) but was positive and significant for male targets ($B = .73$, $p < .001$). This indicates that ratings of trait anger were higher than ratings of trait fear for male targets but not for female targets (see Fig. 2). Hence, these analyses qualified the results under Prediction 2.2a as not only being stronger for male targets as predicted, but, moreover, as being true *only* for male targets.

Prediction 2b. There will be a four-way interaction between emotion condition (anger or fear), the participant's sex, the target's sex, and the target's eye gaze, such that, to a greater extent than male participants, female participants will rate male targets expressing anger with direct gaze as more predisposed toward anger than male targets expressing anger with averted gaze. This contrast will not be significant in the fear condition.

We conducted a $2 \times 2 \times 2 \times 2$ repeated-measures ANOVA to examine the effects of the manipulations. The dependent measure

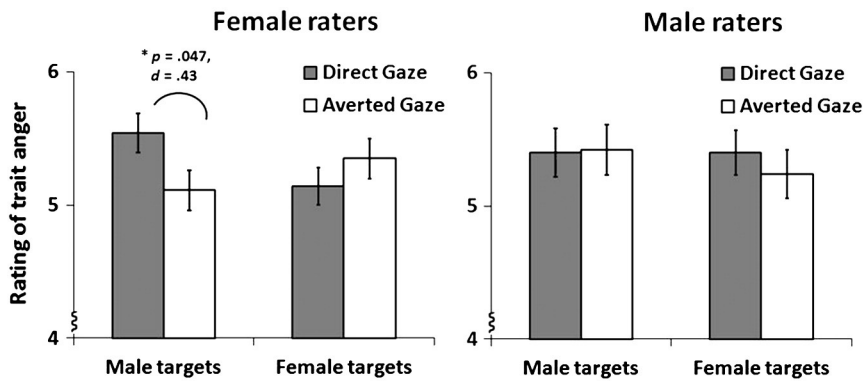


Fig. 3. The effects of gaze, target's sex, and participant's sex on participants' ratings of targets' predisposition toward becoming angry in Study 2.

again consisted of trait ratings averaged across the four same-sex targets. Emotion condition (anger or fear), gaze condition (direct or averted) and participant's sex were between-subjects variables, and target's sex was the repeated measure within participants.

The predicted 4-way interaction was not significant, $F(1, 338) = .49$, $p = .48$. However, to examine whether lower-order patterns were nonetheless consistent with the prediction, we followed this analysis with a 2 (gaze: direct or averted) × 2 (participant's sex) × 2 (target's sex) repeated-measures ANOVA run separately for the anger and fear conditions. Importantly for Prediction 2.2b, within the anger condition, the 3-way interaction between gaze, target's sex, and participant's sex was significant, $F(1, 147) = 5.23$, $p = .024$. Pairwise contrasts revealed that female participants judged male targets to be more dispositionally angry with direct gaze than with averted gaze ($F(1,147) = 3.91$, $p = .05$). No other contrasts within this 3-way interaction approached significance (all $ps > .35$). The 3-way interaction was not significant in the fear condition, $F(1, 191) = 1.81$, $p = .18$, and no contrast pairings within it were significant (all $ps > .10$; see Fig. 3). Thus, although this finding needs to be interpreted with caution, the pattern of results was consistent with Prediction 2b: the 3-way interaction emerged for anger but not for fear.

Prediction 2c. There will be a four-way interaction between emotion condition (anger or fear), the participant's self-perceived vulnerability, the target's sex, and the target's eye gaze such that, to a greater extent than less vulnerable individuals, more vulnerable individuals will rate male targets expressing anger with direct gaze as more predisposed toward anger than male targets expressing anger with averted gaze. This contrast will not be significant in the fear condition.

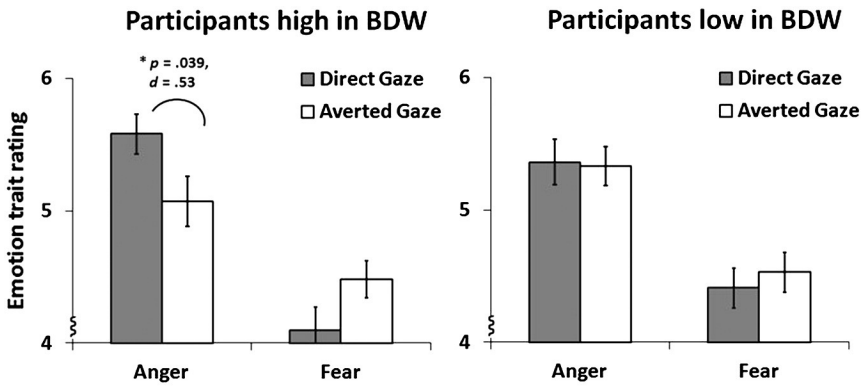
To test this prediction, BDW was dichotomized at the median and substituted for participant sex into the earlier repeated-measures ANOVA. As before, the other three factors were Emotion Type, Gaze, and Target Sex. The 4-way interaction was significant, $F(1, 338) = 4.29$, $p = .039$. A pairwise contrast revealed that participants who were high in BDW and rated angry male faces provided higher ratings for trait anger with direct gaze than with averted gaze, $F(1, 338) = 4.00$, $p = .046$, $d = .22$. However, this was not the case for participants who were low in BDW, $F(1, 338) = .01$, $p = .98$. This was also not the case for any judgments involving fear expressions — indeed, a pairwise contrast showed that there was a marginal opposite trend wherein participants high in BDW rated direct-gaze male fear faces as *less* dispositionally fearful than averted-gaze faces, $F(1, 338) = 3.23$, $p = .073$. Besides these, no other pairwise contrasts in the model approached significance ($ps > .13$). Therefore, Prediction 2c was supported (see Fig. 4).

Prediction 2b concerns participant sex, whereas Prediction 2c concerns self-perceived vulnerability. Tests of these predictions are distinct only if sex is not determinative of self-perceived vulnerability. Critically, the respective representation of the sexes in the high-BDW group did not differ significantly (51.3% of women, 41.8% of men, $\chi^2[1, N = 345] = 3.03$, $p = .08$), indicating that tests of Predictions 2b and 2c are independent of one another.

3.3. Discussion

Study 2 accomplished two goals. First, it replicated and qualified our earlier results, showing that trait anger is judged to be higher than another negative emotional trait (fear) when all else is equal. As in Study 1, across manipulations, targets were judged to be more prone to becoming angry than to feeling another negative emotion even when the images' emotional state intensity was held constant. This replication was qualified by showing that it is only true for male targets: men, but not women, were judged to be more predisposed to anger than to fear above and beyond any rational indications from the images that this was the case. This reveals an attribution process that is irrational in the classic sense (Kelley, 1972) but adaptively rational in its bias toward the error that has likely been consistently less costly over evolutionary time.

Fig. 3 shows the significant interaction indicating that this result was driven by lower ratings of women's marginal trait fear, relative to men. The most direct support for our prediction concerning dangerous individuals would have been to find that this difference was driven by higher ratings of men's marginal trait anger, relative to women. Although we did not find this pattern, the results of these studies still provide important insights. Indeed, direct comparisons between judgments made for male and female targets can be difficult to interpret because people might have different standards for each sex (Biernat, 2009). For instance, men are stereotyped as easily angered (Fischer & Evers, 2010) and women as easily frightened (Hess, Blairy, & Kleck, 2000). Likewise, independent of actual emotional state, by virtue of dimorphic features, male faces appear angrier than female faces (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007). Any or all of these factors might inform how men's and women's respective emotional expressions are interpreted. In contrast, direct comparisons of dispositional anger and fear *within* target sex are relatively unproblematic, because men's and women's fearful images are natural controls for their own angry images in terms of morphology and skill in posing emotions. Such comparisons indeed support Prediction 2a, that the difference between dispositional anger and dispositional fear ratings will be higher for male than for female targets (see Fig. 2).



Note. The 4-way interaction between emotion condition, gaze, target sex, and BDW was significant ($p = .022$). The leftmost contrast between high-BDW participants rating angry male targets with direct vs. averted gaze was significant ($p = .039$). No other simple contrast in this figure was significant ($ps > .086$), and no simple contrasts were significant for participants rating female targets ($ps > .19$; not pictured here).

Fig. 4. The joint effects of participants' Belief in a Dangerous World and gaze direction on ratings of male targets' dispositional anger and fear in Study 2.

Second, because these findings are also potentially explicable in terms of gender stereotypes or morphological influences on perceived expressions, additional features of Study 2 provide critical evidence supporting the notion that the danger posed by the target shapes the degree of bias in anger attribution. Even if the *manifestation* of certain personality traits might be increased by the characteristics of other people in the environment (e.g., individuals prone to violence are more likely to express this trait with victims who appear vulnerable, Buss & Duntley, 2008), in reality an individual's enduring personality does not change with shifting gaze or when examined by a more vulnerable observer. Nevertheless, participants' ratings of male targets' anger-proneness *did* appear to change based on these factors. Results showed that the dispositional attribution of angry expressions appears to be increased by a combination of the target's danger cues (direct gaze, male target) and the participant's elevated vulnerability (if the participant is female or believes that the world is dangerous). These findings echo prior findings that fear of sexual coercion motivates women's fear of, and bias against, outgroup male targets in particular (Navarrete, McDonald, Molina, & Sidanius, 2010). These nuanced results, inconsistent with an account based solely on gender stereotypes, provide additional support for the notion that the estimation of trait anger involves a true bias rooted in adaptive error management.

4. General discussion

These studies provide the first evidence that the estimation of trait anger is biased in an adaptively rational way. In Study 1, perceivers interpreted angry behaviors as a reflection of an actor's personality regardless of how justified these behaviors were, especially when the behaviors were intense. This pattern was not obtained for another negative emotion, disgust. Study 2 replicated and extended this general finding with a different comparison emotion, fear. In Study 2, perceivers' overestimation of trait anger was enhanced by combinations of factors associated with the target's capability and likelihood of aggressing against the observer and the observer's vulnerability to such aggression. Specifically, female participants and participants who considered the world dangerous saw more anger in the personalities of targets who were male and looking directly at them. These nuanced findings provide support for the core hypothesis and are difficult to explain under alternative accounts.

4.1. Theoretical implications

4.1.1. Cognitive versus behavioral biases

The current research adds to the growing list of documented cognitive biases rooted in error management (Haselton & Galperin, 2013; Johnson et al., 2013). Some researchers have argued that such biases are unnecessary (and therefore unlikely to exist) because adaptive behavior, not cognition, is what ultimately affects fitness; therefore, people can theoretically "decide" to behave in adaptively biased ways without having to make systematically biased judgments (McKay & Dennett, 2009; McKay & Efferson, 2010). For instance, a woman could decide to avoid a man who has expressed anger toward her in the past without overestimating his trait anger. Indeed, there might be downstream costs to psychological biases, if, for example, a mechanism's biased output is used by other mechanisms to which the same cost asymmetry does not apply.

While behavior is the ultimate determinant of fitness, the extent to which biased behavior is produced by biased cognition remains an empirical question. The corpus to which our results contribute reveals cognitive biases in a variety of judgment domains (Haselton et al., 2009; Haselton & Buss, 2009), suggesting that biased behavior frequently does flow from biased cognition (see Johnson et al., 2013, for discussion).

4.1.2. Ingroups and outgroups

For ancestral humans, the consequences of dealing with an anger-prone individual were not always negative, but rather depended on whether the individual was an assailant or an ally. A propensity for aggression would often have been a valued quality in allies, as long as it was directed toward outgroups and facilitated successful intergroup competition. The tests conducted in the current study were not designed to apply to allies in situations of intergroup conflict, and indeed, our findings suggest that participants implicitly treated unfamiliar individuals as non-allies by default. In the absence of readily observed cues of shared group membership (Boyd & Richerson, 2009; Henrich, 2004; Kurzban, Tooby, & Cosmides, 2001), it might generally have enhanced fitness to evaluate strangers with caution, as our participants did.

4.1.3. The correspondence bias and negativity bias

The correspondence bias (Gilbert & Malone, 1995; Ross & Nisbett, 1991) occurs whenever, to a logically unwarranted extent, people attribute others' behaviors to the target's enduring traits rather than to the situation. This bias has been documented across many judgment domains, including attitudes, moral character, competence, and emotionality. Researchers have typically focused on examining the mechanisms through which this bias operates across domains, rather than examining its ultimate cause (but see Andrews, 2001) or testing theoretically-driven hypotheses about how it might differ between domains. While our results could be classified as an instance of the correspondence bias, our research speaks directly to the latter issues, as domain-general or purely proximate explanations of the correspondence bias do not predict that angry behaviors will be attributed to enduring traits to a greater extent than disgusted or fearful behaviors.

An overarching pattern characterizing both our results and a majority of findings regarding the correspondence bias is that, when people evaluate others, bad looms larger than good (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001; Ybarra, 2002). This "negativity bias" facilitates adaptively attending to and addressing threats (Rozin & Royzman, 2001), and is manifested in people's tendency to attribute negative or socially undesirable behaviors especially strongly to enduring traits (e.g., Reeder & Spores, 1983; Ybarra, 2002). While the current results for anger (a generally socially undesirable trait) are consistent with this phenomenon, they also move beyond it by illustrating the adaptively rational ways in which context affects the degree of the bias for anger but not for other negative emotions.

4.2. Practical implications

Because people tend to see the bad in others, they are likely to avoid interacting or forming relationships with individuals who made a bad first impression even if they were situationally induced to behave this way. The specific case of the overestimation of trait anger suggests that people may avoid new acquaintances after a single instance of angry behavior, even if it was justified in the eyes of the perceiver. Moreover, this is especially likely when the target is formidable (e.g., a muscular man) and when the observer is either chronically vulnerable or feels temporarily unsafe. Although these patterns were adaptive in the social environments of our ancestors, modern humans live in a much safer world (Pinker, 2011). Hence, the biased overestimation of trait anger may lead people to mistakenly form negative impressions, eschewing relationships with others who might otherwise have become valued social partners. More broadly, our results potentially speak to the origins of stereotypes, particularly those linking gender and emotion. As noted earlier, folk models attribute greater trait anger to men. That such stereotypes arise and persist despite ready opportunities to observe that they are inaccurate is potentially explained by adaptively biased attributions,

given that angry men pose a much greater threat of violence than do angry women.

Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2013.06.003.

References

Adams, R. B., Jr., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, *14*, 644–647.

Altemeyer, B. (1998). The other "authoritarian personality". *Advances in Experimental Social Psychology*, *30*, 47–92.

Andrews, P. W. (2001). The psychology of social chess and the evolution of attribution mechanisms: Explaining the fundamental attribution error. *Evolution and Human Behavior*, *22*, 11–29.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*, 179–190.

Biernat, M. (2009). Stereotypes and shifting standards. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 137–152). New York: Psychology Press.

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3281–3288.

Buss, D. M., & Duntley, J. D. (2008). Adaptations for exploitation. *Group dynamics: Theory, research, and practice*, *12*, 53–62.

Daly, M., & Wilson, M. I. (1988). *Homicide.* New York: Aldine de Gruyter.

Fessler, D. M. T. (2010). Madmen: An evolutionary perspective on anger and men's violent responses to transgression. In M. Potegal, G. Stemmler, & C. D. Spielberger (Eds.), *Handbook of anger: Constituent and concomitant biological, psychological, and social processes* (pp. 361–381). Springer.

Fischer, A. H., & Evers, C. (2010). Anger in the context of gender. In M. Potegal, G. Stemmler, & C. D. Spielberger (Eds.), *Handbook of anger* (pp. 349–360). New York: Springer.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions.* New York: Norton.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. *Social Cognition*, *27*, 733–763.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*, 81–91.

Haselton, M. G., & Buss, D. M. (2009). Error management theory and the evolution of misbeliefs. *The Behavioral and Brain Sciences*, *32*, 522–523.

Haselton, M. G., & Galperin, A. (2013). Error management in relationships. In J. A. Simpson, & L. Campbell (Eds.), *Handbook of Close Relationships* (pp. 234-254). Oxford University Press.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, *10*, 47–66.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, *53*, 3–35.

Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of expression intensity, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior*, *24*, 265–283.

Johns, S. E. (2011). Perceived environmental risk as a predictor of teenage motherhood in a British population. *Health & Place*, *17*, 122–131.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology and Evolution*, *28*(8), 474–481.

Keeley, L. H. (1996). *War before civilization: The myth of the peaceful savage.* New York: Oxford University Press.

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology (4th ed)*, *1*. (pp. 233–265). Boston, MA: McGraw-Hill.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *The Behavioral and Brain Sciences*, *27*, 313–327.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*, 15387–15392.

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *The Behavioral and Brain Sciences*, *32*, 493–561.

McKay, R. T., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior*, *31*, 309–319.

Navarrete, C. D., McDonald, M. M., Molina, L. E., & Sidanius, J. (2010). Prejudice at the nexus of race and gender: An outgroup male target hypothesis. *Journal of Personality and Social Psychology*, *6*, 933–945.

Navarrete, C. D., Olsson, A., Ho, A. K., Mendes, W. B., Thomsen, L., & Sidanius, J. (2009). Fear extinction to an out-group face: The role of target gender. *Psychological Science*, *20*, 155–158.

Pinker, S. (2011). *The better angels of our nature: Why violence has declined.* New York: Viking.

Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, *44*, 736–745.

Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, *10*. (pp. 173–220). San Diego, CA: Academic Press.

Ross, L., & Nisbett, R. (1991). *The person and the situation: Perspectives of social psychology.* New York: McGraw-Hill.

Rozin, P., Haidt, J., & McCauley, C. R. (2000). Disgust. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 637–653) (2nd Edition). New York: Guilford Press.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.

Sell, A. (2009). Applying adaptationism to human anger: The recalibrational theory. In P. R. Shaver, & M. Mikulincer (Eds.), *Understanding and reducing aggression, violence, and their consequences.* Washington, DC: American Psychological Association.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*, 813–838.

Snyder, J. K., Fessler, D. M. T., Tiokhin, L., Frederick, D. A., Lee, S. W., & Navarrete, C. D. (2011). Trade-offs in a dangerous world: Women's fear of crime predicts preferences for aggressive and formidable mates. *Evolution and Human Behavior*, *32*, 127–137.

Tottenham, N., Tanaka, J., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249.

Tybur, J. M., Lieberman, D. L., & Griskevicius, V. (2009). Microbes, mating, and morality: Individual differences in three functional domains of disgust. *Journal of Personality and Social Psychology*, *29*, 103–122.

Ybarra, O. (2002). Naive causal understanding of valenced behaviors and its implications for social information processing. *Psychological Bulletin*, *128*, 421–441.