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

A R T I C L E   I N F O

A B S T R A C T

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.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Assessing others' personality traits is a key adaptive problem that social cognition evolved to address. Understanding people's person-alities 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 evolu-tionary 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).

& 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.eh-bonline.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