Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: www.tandfonline.com/journals/pcem20

Biased judgments of emotion are resistant to changes in the prevalence of anger

Delaney McDonagh & Timothy Sweeny

To cite this article: Delaney McDonagh & Timothy Sweeny (26 May 2025): Biased judgments of emotion are resistant to changes in the prevalence of anger, Cognition and Emotion, DOI: 10.1080/02699931.2025.2506667

To link to this article: https://doi.org/10.1080/02699931.2025.2506667

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2025.2506667

Biased judgments of emotion are resistant to changes in the prevalence of anger

Delaney McDonagh and Timothy Sweeny

Department of Psychology, University of Denver, Denver, CO, USA

ABSTRACT

Emotional expressions are often nuanced and are thus evaluated with an unavoidable degree of uncertainty. We recently showed that perceivers are biased to interpret others' expressions negatively, especially when they are seen in crowds. We argued that the flexible nature of this bias is protective, and that it emerges because perceivers learn about the threatening nature of crowds over prolonged windows of time. Here we evaluated whether this negativity bias can recalibrate as perceivers learn about changes in the prevalence of threat in the short-term. Perceivers viewed single faces or crowds of four faces and indicated whether their expressions were happy or angry. We manipulated the prevalence of anger across three experiments, displaying angry faces on 75% or 25% of trials. We replicated our previous findings, again showing that observers were biased to evaluate others as angry, especially in crowds. Surprisingly, the strength of anger bias was not affected by the prevalence of anger, neither when perceivers learned about this implicitly nor when they were explicitly informed about it in advance. This suggests that anger bias is flexible, but more to contextual cues, like the number faces being evaluated, and less to statistical cues, like the prevalence of anger.

ARTICLE HISTORY
Received 7 August 2024
Revised 4 April 2025
Accepted 9 May 2025

KEYWORDS
Anger; crowd; emotion perception; threat; emotion bias

The human visual system is very sensitive to emotion, allowing perceivers to categorise prototypical expressions with high accuracy (Calvo & Nummenmaa, 2016; Elfenbein & Ambady, 2002). Yet the facial expressions people typically see are often nuanced or ambiguous, and the judgements people make about these expressions are thus made with an unavoidable degree of uncertainty. When making these sorts of low confidence evaluations, most perceivers are biased to interpret facial expressions as negative (Mihalache et al., 2021; Neta et al., 2009, 2013; Neta & Whalen, 2010). This negativity bias appears to be protective – a kind of tradeoff aimed to reduce the cost of categorisation errors when they inevitably occur, by prioritising detection of threat when it is present over misperception of threat when it is absent (Haselton et al., 2009). Recent work also indicates that negativity bias is not a fixed heuristic. Rather, it is flexible and can shift in strength based on contextual threat cues that perceivers presumably learn over long periods of time. For example, crowds are associated with greater potential for threat than individuals (Baumeister et al., 2015; Vilanova et al., 2017), and in turn, a bias to report that people are angry is amplified when their faces are seen in crowds (Mihalache et al., 2021). Yet to our knowledge, little work has examined whether this flexibility extends to dynamic cues about the prevalence or likelihood of threat that perceivers could only gather in the short term. Here, we examine whether perceivers adjust their negativity bias as they implicitly or explicitly learn about changes in the prevalence of angry or happy faces across trials in a binary emotion categorization task. In doing so, we hope to gain a more

CONTACT Delaney McDonagh delaney.mcdonagh@du.edu Department of Psychology, University of Denver, 2155 S Race St, Frontier Hall, Denver, CO 80210, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02699931.2025.2506667.

nuanced understanding of negativity bias as a dynamic process and evaluate its persistence over time.

Biases in the perception of emotional expressions

Negatively biased judgments about the emotional expressions of individuals appear to manifest under many different types of uncertainty. For example, when evaluating surprised expressions – which are considered ambiguous in terms of their emotional valence – people are biased to interpret the faces as negative (Neta et al., 2009, 2013; Neta & Whalen, 2010). Further, a bias to endorse anger (and not just negativity) manifests in judgements of expressions with weak intensity or degraded visibility (Mihalache et al., 2021). Contextual cues can further bias these perceptions. For example, observers are more likely to perceive individuals as angry when they are holding a household object that could be used as a weapon (e.g. gardening shears) compared to an object that is regarded as harmless (e.g. a watering can) (Holbrook et al., 2014). While these negativity biases may present as errors in judgement, according to Error Management Theory, they may in fact be adaptive (Johnson et al., 2013). This framework posits that when a person is uncertain about any sort of judgement, be it an abstract decision about risk versus reward or something more perceptual in nature, the most effective decision-making strategy is to minimise potential cost if that judgment turns out to be in error (Johnson et al., 2013). In the case of negativity bias, although there is some cost in misjudging that another person's facial expression looks slightly angry (when it is in fact neutral or happy), this bias at least protects a perceiver from the even greater cost of missing another person's subtle display of anger when it is, in fact, present (Adams et al., 2006). Note, however, that negativity bias is a heuristic, and as such it is engaged as a sort of backup system for when perceivers are unable to make veridical interpretations of individuals and their expressions.

Even stronger biases occur when people evaluate the emotions of crowds. When viewing groups of faces, perceivers are quite precise in representing the mean emotion of the crowd, even if they retain little information about the emotion of each individual face within the set (Haberman & Whitney, 2009). When faced with ambiguity or uncertainty about the emotion of a crowd, however, the precision of crowd perception is forfeited in favour of adaptive decision-making and toward an especially potent heuristic of negativity bias. That is, perceivers appear to shift from making judgments primarily based on what they see, and instead fall back on beliefs or lay theories about how people act in crowds, all of which seem to point to an increased potential for aggression. Indeed, individuals often behave differently when alone than when a part of a crowd. Crowds can enable a process of deindividuation and reduction in responsibility which can lead to decreased inhibition for individual crowd members and even increased aggression at the level of the entire crowd (Baumeister et al., 2015; Vilanova et al., 2017). Therefore, the cost associated with failing to identify anger may be amplified when a perceiver evaluates a crowd compared to an individual. Recent findings from Mihalache et al. (2021) are consistent with this premise. They found that perceivers' bias to report anger was amplified when judgments were made about crowds of faces compared to individual faces, especially when those faces had subtle expressions or were partially occluded.

The findings from Mihalache et al. (2021) provide new evidence that negativity bias is flexible, in this case shifting based on information about people and how they tend to behave in crowds, learned presumably over a prolonged window of time. To our knowledge, however, little work has been done to examine the effect of *short-term* changes in one's dynamic visual environment on negativity bias. Here, we re-examine the experimental paradigm developed by Mihalache et al. (2021) to better understand how negativity bias persists (or adapts) based on relevant and observable statistics about the changing potential for threat in an individual's immediate environment. For example, bias may increase in strength if an individual believes that angry faces are more prevalent than happy faces in an experimental context. Based on ample evidence outside the field of emotional decision making (as well as very limited work within this field), we expected that negativity bias may indeed exhibit this short-term flexibility.

Evidence for short-term changes in response bias

Perceivers are quite good at noticing when certain objects, patterns, or targets appear more frequently than others across trials in experimental paradigms.

This difference in frequency, or base rate, is robustly associated with a shift in bias toward the more prevalent item or response option (Bohil & Wismer, 2015; Godwin et al., 2010, 2015; Maddox, 1995; Maddox & Bohil, 2003; Peltier & Becker, 2016, 2017; Wolfe & Van Wert, 2010). For example, if a perceiver notices that rightward-tilted lines are more prevalent in an orientation discrimination task, they could be expected to adopt a bias to report that lines with vertical orientations are tilted to the right. The opposite is also true; when targets are infrequent, perceivers have been shown to adopt a more conservative bias to limit potential false alarms (Peltier & Becker, 2017). Both explicitly *and* implicitly learned information about prevalence have been shown to guide shifts in decision-making (Bohil & Wismer, 2015; Koehler, 1996). Implicit learning of target prevalence has even been shown to occur relatively quickly, in the previously mentioned experiments at the level of a few hundred trials (Bohil & Wismer, 2015; Godwin et al., 2010, 2015; Peltier & Becker, 2016). Effects of perceivers shifting their response biases based on target prevalence have been observed in experiments utilising implicitly learned base rates ranging from 90% to 10% (Godwin et al., 2010, 2015; Peltier & Becker, 2016, 2017), with more extreme base rate manipulations naturally leading to more extreme response biases. The aforementioned experiments employed a variety of stimuli and tasks in order to examine the effects of base rate, including medical diagnosis tasks (i.e. diagnose with disease A or B), weapon detection tasks in TSA bag search images (i.e. is a weapon present or not), shape discrimination tasks (i.e. is shape tall or short), and letter identification in visual search. The diversity of stimuli implemented across experiments and the consistent finding of changes in response bias related to target prevalence suggest that this mechanism is potentially domain general and may manifest similarly in the context of emotion discrimination. Exactly how base rate or target prevalence may influence negativity bias in the evaluation of emotional expressions, however, is less clear.

Previous work highlights how biases in decision-making respond to changes in base rate across a variety of contexts, yet very little work has been done within the domain of emotional expression perception. Additionally, to our knowledge, no work has examined how negativity bias in emotion perception changes during naturalistic, implicit learning, or how the flexibility of bias applies to the perception of individual faces and crowds. Some evidence suggests that explicitly learned base rate information can shift negativity bias in an emotion categorisation task (Lynn et al., 2012). Under low base rate conditions (i.e. 25% of trials included faces with more threatening expressions, and observers were made explicitly aware of this), observers' bias to endorse threat became more conservative. In other words, when faces with threatening expressions were shown less frequently, perceivers became less likely to report that faces were threatening when they were uncertain about the expression being shown. These results suggest that perceivers should be able to shift their bias to reflect short term changes in the prevalence of threat in their visual environment, but it is unclear whether these shifts can occur implicitly, without perceivers' explicit awareness of base rate. It is also unclear how implicit learning and explicit knowledge may interact when they are in conflict (e.g. when perceivers are informed that anger is prevalent, when in fact it is not, which they would have to learn implicitly). In sum, we aimed to examine these questions while replicating the original findings from Mihalache et al. (2021) with an expanded stimulus set.

The current study and hypotheses

Across three experiments, we examined the effects of explicitly and implicitly learned statistical information on an individual's decision making in a binary (angry versus happy) emotion categorisation task. In our first experiment, we examined the effect of implicit learning on negativity bias by changing the prevalence of angry faces in each experimental block. Observers were given no information regarding the prevalence of threat; therefore, any changes in response bias would presumably be due to implicit learning across trials. We hypothesised that observers would show the strongest anger bias in blocks of trials where angry faces were more prevalent than happy faces. In our second experiment, we examined whether changes in the base rate of angry faces in one context (e.g. in crowds) would influence bias in another context (e.g. judgments of individuals). Here, we were agnostic about whether an effect of prevalence would be specific to context, but if anything, we hypothesised that it would be strongest for judgments of crowds when anger was most prevalent in crowds. In our final experiment, we expanded upon our initial design to examine the impact of explicit information, implicit learning, and incorrect, explicit information about the prevalence of threat on negativity bias. Again, we hypothesised that anger bias would be strongest when observers were informed that angry faces would be more prevalent within a block of trials.

Experiment 1

Methods

Observers

Ninety undergraduate students (47 women, 35 men, eight who preferred not to say) from the University of Denver participated in Experiment 1. The number of observers in this sample was selected based on a statistical power analysis using an effect size from an experiment with comparable presentation times and design parameters to the current study (Mihalache et al., 2021). The estimate was based on the size of a main effect of anger bias for judgments of crowds against a null criterion value of zero, which was $d = 0.85$ in Experiment 1 of Mihalache et al. (2021). With an alpha of 0.05 and beta of 0.2, the projected sample size needed to achieve 80% power for this comparison in the current study was 13 observers. We increased our sample size substantially beyond this number, to 90 participants, in order to examine what we anticipated to be a subtle effect of implicitly learned threat prevalence on anger bias. Observers provided informed consent and received partial course credit for their participation. All observers had normal or corrected-to-normal visual acuity and completed the experiment in a dimly lit room. The study was approved by the Institutional Review Board at the University of Denver, and the research was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Stimuli

The stimuli used in this study consisted of the faces of four White men and four Black men from the NimStim Face Set (Tottenham et al., 2009). To create a range of facial expression intensities, we generated linear interpolations (i.e. morphs) between a neutral expression and two full-intensity emotional expressions (angry and happy) for each of the eight identities. We then selected two levels of emotion intensity (2% and 80%) for each of two expressions (angry and happy) for each identity, resulting in 32 unique faces. In this study, we elected to use angry and happy expressions to index negativity bias, as work by Mihalache et al. (2021) revealed that negativity bias is specific to categorizations of anger and not fear. For example, when people were discriminating between only fearful and angry expressions, they did not produce a fear bias, but instead continued to produce an anger bias. We included a low- and high-intensity examples of each expression so that we could measure bias both on trials where the visual signal was relatively strong or very weak, and because bias has been shown to be stronger in the case of the latter (Mihalache et al., 2021). Faces were edited to appear on a uniform, grey background (RGB = 170, 170, 170), and each face subtended a visual angle of 3.75° × 4.53°. We additionally generated a set of 32 masks that consisted of phase-scrambled versions of each of the 32 unique faces in our stimulus set. Pattern masks are often presented after target images to decrease overall visibility and prevent residual processing (Rolls et al., 1999). Phase-scrambling these masks ensured that they retained the exact same low-level image properties as the faces (e.g. overall luminance, contrast, etc.). Each mask was perfectly matched with one of the faces in the stimulus set, and the masks were always paired with those matched faces in the experiment. This approach gave us the best chance to reduce residual visual processing of the faces beyond their actual presentation time (Oyama et al., 1983) so that impressions and judgments of the faces would be based on perception of them during the trial and not decaying representations lingering in visual short-term memory after they had disappeared.

Procedure

The experiment was programmed using jsPsych, and observers completed the experiment online using their own personal computer. Each observer's screen resolution and physical distance from the screen was calculated with the Virtual Chinrest plugin (Li et al., 2020), which allowed us to scale the physical sizes of the stimuli to be identical on each observer's monitor and maintain consistent visual angles across observers' retinae.

In Experiment 1, observers completed a two-alternative-forced-choice (2AFC) task in which they categorised emotional expressions as happy or angry (see Figure 1). Each trial began with the presentation of a central fixation cross for one of four randomly determined durations (350, 400, 450, and 500 msec). Following the presentation of the fixation

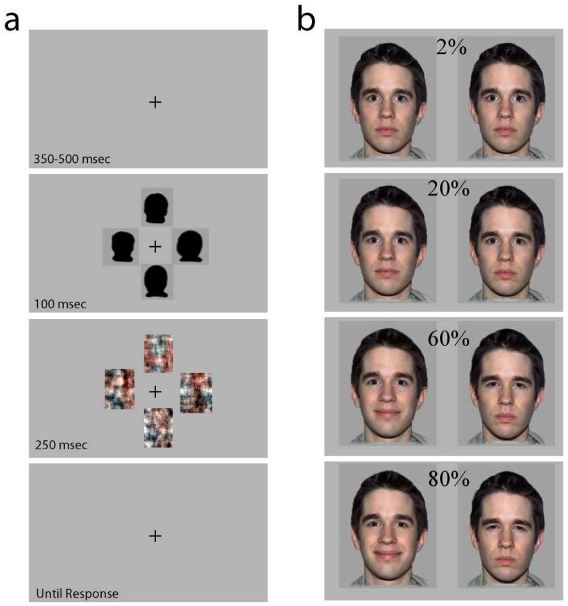

Figure 1. Stimuli and illustration of a typical trial sequence. (a) Observers viewed a crowd of four faces (demonstrated by silhouettes for copyright purposes) for 100 msec, followed by phase-scrambled masks for 250 msec. Then observers were presented with a blank screen until they made a response. (b) An example of one identity displaying 2%, 20%, 60%, and 80% anger and happiness. In Experiment 1, observers viewed expression intensities of 2% and 80%. In Experiments 2 and 3, observers viewed expression intensities of 20% and 60%.

cross, a single face (the single condition) or a crowd of four faces (the crowd condition) briefly appeared for 100 msec. We selected this duration to prevent observers from making deliberate saccades to individual faces (Findlay & Walker, 1999). In the crowd condition, each face had a unique identity but exhibited the same expression and expression intensity as the others in the crowd (e.g. four different men showing anger at 80% intensity). One face appeared in each quadrant of the screen, situated around a central fixation cross. The centres of adjacent faces were 5.84° away from each other in the horizontal and vertical directions. In the single condition, one face appeared in a randomly selected position chosen from these four locations. To prevent observers from predicting the exact locations of the faces prior to each trial and potentially focusing on one location in advance, we randomly (and independently) jittered the location of each face anywhere from 1 to 15 degrees in both the horizontal and vertical directions on each trial.

Immediately after the face(s) disappeared, they were replaced by phase scrambled versions of themselves (see the Stimuli section) for 250 msec. Once the phase scrambled images disappeared, observers were tasked with categorising the expression of the face(s) as angry or happy by pressing the "f" or the "j" key on the keyboard. The association between expressions and these keys was counterbalanced across observers. There was no time limit to provide a response; however, observers were advised to respond quickly, balancing both speed and accuracy. The experiment automatically advanced to the next trial 500 msec after an observer made their response.

The experiment consisted of two blocks of 640 trials, half from the single condition and the other half from the crowd condition. Trials within each block were presented in a randomised order. One block had a high prevalence of threat, in which 75% of the trials included faces with angry expressions. The other block had a low prevalence of threat, where 25% of the trials included faces with angry expressions. We selected these values of prevalence based on previous work that examined the impact of explicit learning about the prevalence of threat on emotion judgments (Lynn et al., 2012). The prevalence of threat was always the same for the single and crowd conditions within a block. Block presentation order was counterbalanced across observers.

Analyses

We elected to employ Signal Detection Theory (Macmillan & Creelman, 2005) to quantify both observer's sensitivity (d-prime or d') and bias (criterion or c) during the discrimination of angry and happy expressions. This analysis framework allowed us to understand an observer's decision-making under conditions of uncertainty, by separately indexing the ability to discriminate accurately between stimuli (sensitivity) and their tendency to favour a specific response (bias). By individually examining these facets of decision-making, we were able to gain a more nuanced understanding of each observer's performance. Observers' responses were categorised in terms of four different outcomes, with angry expressions as the target signal. A "hit" corresponded to when an angry face was correctly identified as angry, a "false alarm" was recorded when a happy face was incorrectly identified as angry, a "miss" occurred when an angry face was incorrectly identified as happy, and a "correct rejection" corresponded to when a happy face was correctly identified as happy. With these four possible response outcomes, we then derived d-prime and criterion values. D-prime is a measure of sensitivity that captures an observer's ability to correctly discriminate between two responses (happy and angry, in our case). Criterion measures an observer's bias to respond in a certain way (i.e. happy or angry), regardless of what emotion is present on a given trial. For each observer, d-prime and criterion were calculated for each condition of interest using z-transformed hit (h) and false alarm (fa) rates, where hit rate = (hit)/(hit + miss), false alarm rate = (f.a.)/(f.a. + correct rejection), $d' = Z\ \text{h} - Z\ \text{fa}$, and $c = -0.5(Z\ \text{h} + Z\ \text{fa})$. Hit and false alarm rates of 1.0 were replaced with a value determined by (1–1/2*n), with n reflecting the number of possible hits or false alarms. Hit and false alarm rates of 0.0 were replaced with a value determined by (1/2*n), with n again reflecting the number of possible hits or false alarms. D-prime values around zero indicate that the observer has little or no sensitivity for discriminating between angry and happy expressions, and positive values of d-prime reflect high sensitivity. Criterion values can range from negative to positive infinity. Negative criterion values indicate a liberal use of the angry response option, while positive criterion values indicate a conservative use of the angry response option. Notably, bias (criterion or *c*) is often strongest when sensitivity (d-prime or *d′*) is weakest (Mihalache et al., 2021). We therefore expected negative bias to be strongest for judgments of the faces with weak expressive intensity.

Results

Preliminary analyses

A repeated-measures ANOVA with d-prime as our dependent variable revealed main effects of Numerosity, $F(1, 89) = 75.84$, $p < 0.001$, $h_p^2 = 0.46$, Intensity, $F(1, 89) = 108.2$, $p < .001$, $h_p^2 = .55$, and an interaction between the two, $F(1, 89) = 15.1$, $p < .001$, $h_p^2 = .15$. These results indicate that observers were better at discriminating between happy and angry expressions when faces were presented within the context of a crowd compared to a single face, and when faces had high expressive intensities compared to low intensities. This analysis provides confirmation that observers understood the task and were able to discriminate between the happy and angry expressions, directly replicating the results of Mihalache et al., 2021. Next, a repeated-measures ANOVA with criterion as the dependent variable revealed main effects of Numerosity, $F(1, 89) = 75.84$, $p < .001$, $h_p^2 = .46$, and Intensity, $F(1, 89) = 108.2$, $p < .001$, $h_p^2 = .55$, and an interaction between the two, $F(1, 89) = 15.1$, $p < .01$, $h_p^2 = .15$. This analysis replicates the primary findings from Mihalache et al., 2021, in which observers exhibited a bias to report that faces had angry expressions. This bias was amplified when faces had low expressive intensities compared to high intensities and when faces were presented within the context of a crowd compared to a single face (Figure 2A). This replication is important – it confirms that our design and sample size were clearly robust enough to capture an effect of crowd context on anger bias, giving us confidence that we should also be able to capture an effect of prevalence if it is indeed present.

Primary analyses

In order to examine an effect of Prevalence (high-threat-prevalence versus low-threat-prevalence) on sensitivity and to determine if that effect was dependent on Block Order (e.g. whether the high-prevalence block was viewed first), we conducted a mixed ANOVA with d-prime as our dependent variable, a within-subjects factor of Prevalence, and a between-subjects factor of Block Order. There was no main effect of Prevalence, $F(1, 89) = 3.12$, $p = .08$, $h_p^2 = .03$, and no main effect of Block Order, $F(1, 89) = .08$, $p = .78$, $h_p^2 < .001$. There was, however, an interaction between the two, $F(1, 89) = 5.77$, $p < .05$, $h_p^2 = .06$. These results indicate that observers' sensitivity for discriminating between happy and angry expressions was greater for the high-threat-prevalence block when it was presented first. There was no significant difference in sensitivity between blocks, however, when the low-threat-prevalence block was presented first.

We then conducted a mixed ANOVA with criterion as the dependent variable, a within-subjects factor of Prevalence, and a between-subjects factor of Block Order. Again, there was no main effect of Block Type, $F(1, 89) = 3.121$, $p = .081$, $h_p^2 = .03$, and no main effect of Block Order, $F(1, 89) = .08$, $p = .78$, $h_p^2 < .001$. There was, however, an interaction between the two, $F(1, 89) = 5.78$, $p < .05$, $h_p^2 = .06$. We followed up on this interaction by conducting paired-samples t-tests between the high- and low-threat-prevalence blocks separately for observers who viewed the high-prevalence block first or second. We found that bias to endorse anger was significantly more liberal (more negative criterion values) in the high-prevalence block compared to the low-prevalence block,

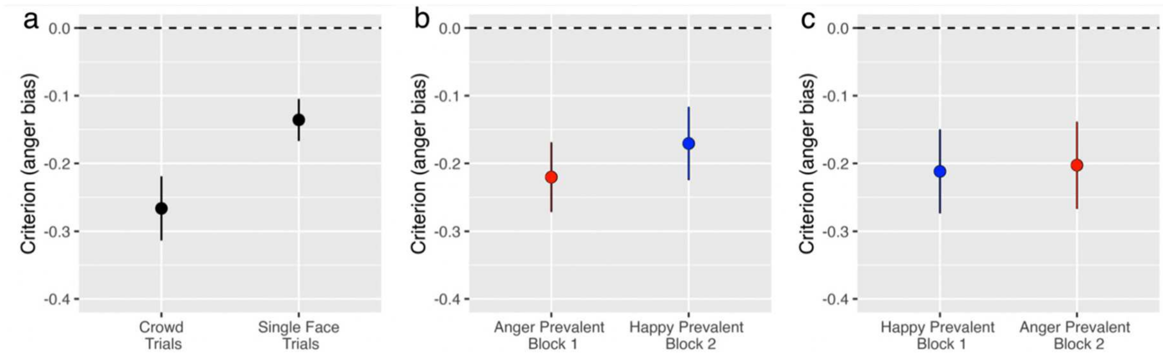

Figure 2. Criterion scores as an index of anger bias in Experiment 1. Negative criterion values indicate stronger bias to report that facial expressions are angry. (a) Bias as a function of numerosity across trials from all observers as a function of crowd or single-face context, which was manipulated within-observers. (b) Bias scores for observers who completed block-order 1, which featured an angry-prevalent block followed by a happy-prevalent block. (c) Bias scores for the other half of observers who completed block-order 2, which featured a happy-prevalent block followed by an angry-prevalent block.

when the high-threat-prevalence block was presented first, $t(46) = -1.31$, $p < .05$, $d = .76$ (Figure 2B), but not when the low-prevalence block was presented first $t(42) = .22$, $p = .84$, $d < .001$ (Figure 2C). In summary, in Experiment 1, we found *some* evidence that observers were more biased to endorse anger during a block of trials where angry expressions were more prevalent than happy expressions, and when this information was learned implicitly. But this effect of prevalence depended on whether the high-prevalence block was presented first.

Experiment 2

To better understand the results of Experiment 1, we aimed to examine how base rate might affect emotional decision-making when manipulated independently for trials with crowds and individual faces. Crowd perception employs a process known as ensemble coding, which allows perceivers to effectively extract summary statistical, or gist-level information from groups of redundant stimuli (Alvarez, 2011; Whitney & Yamanashi-Leib, 2018). This ensemble mechanism allows perceivers to make judgments about visual features across a wide variety of domains, including emotional expressions (Haberman & Whitney, 2007, 2009). Crowds and individual faces, therefore, are processed via different visual mechanisms. If the prevalence of threat differed for trials with individual faces and crowds, observers could conceivably be able to account for this and then differentially apply this learning to adjust their bias for trials with individual faces and crowds. It may even be the case that observers learn about prevalence faster with crowds since they contain more information, and thus provide a larger sample early in the experiment compared with individual faces, or because bias tends to be more potent with crowds in the first place. Alternatively, any unique learning about threat prevalence from crowds or single faces may instead be applied indiscriminately to any future judgments of emotion, operating more generally on trials with single faces or crowds. We intentionally examined this question in the presentation order that revealed significant results in Experiment 1 (i.e. high-threat-prevalence block followed by low-threat-prevalence block).

Methods

Observers

Ninety-one observers from the University of Denver (61 female, 26 male, four who preferred not to say) participated in Experiment 2. Observers provided informed consent and received partial course credit for their participation. All observers had normal or corrected-to-normal visual acuity and completed the experiment in a dimly lit room.

Stimuli and procedure

The stimuli and procedures were identical to those described in Experiment 1 with a few exceptions. Rather than presenting faces with expression intensities of 2% and 80%, we selected intensities of 20% and 60%. This change was made to allow perceivers to gather more information that could be used for learning about prevalence from all faces presented across trials (presumably, the faces with 2% intensities were not strong signals for learning prevalence in Experiment 1). Additionally, we manipulated the prevalence of threat independently for single and crowd trials within each block. In one block, there was a high prevalence of threat for the crowd condition, in which 75% of trials in the crowd condition included faces with angry expressions and 25% of trials in the single condition included faces with angry expressions. For the other block, there was a high prevalence of threat for the single condition, in which 75% of trials in the single condition included faces with angry expressions and 25% of trials in the crowd condition included faces with angry expressions. Block order was counterbalanced across observers.

Results

Confirmatory analyses

A repeated-measures ANOVA, with d-prime as a dependent variable, revealed main effects of Numerosity, $F(1, 90) = 53.41$, $p < .001$, $h^2_p = .37$, Intensity, $F(1, 90) = 90.12$, $p < .001$, $h^2_p = .50$, and an interaction between the two, $F(1, 90) = 52.25$, $p < .001$, $h^2_p = .37$. Additionally, a repeated-measures ANOVA, with a dependent variable of criterion, revealed main effects of Numerosity, $F(1, 90) = 53.41$, $p < .001$, $h^2_p = .37$, and Intensity, $F(1, 90) = 90.12$, $p < .001$, $h^2_p = .5$, and an interaction between the two, $F(1, 90) = 52.25$, $p < .001$, $h^2_p = .37$. These analyses demonstrated similar patterns to previous work (Mihalache et al., 2021) and our first experiment, where observers were able to discriminate between the happy and angry expressions in the experiment and overall exhibited a bias to report that faces had angry expressions. This bias was again amplified when faces had low expressive intensities compared to high intensities and when faces were presented within the context of a crowd compared to a single face.

Primary analyses

To answer our question of if observers learn and apply prevalence information separately based on Numerosity, we examined criterion values separately for trials from the single and crowd conditions. We conducted two mixed ANOVAs (one for crowd trials and one for single face trials), with criterion as the dependent variable, a within-subjects factor of Block Type, and a between-subjects factor of Block Order. The ANOVA for single face trials did not reveal any main effects of Block Type, $F(1, 89) = .28$, $p = .60$, $h^2_p < .001$, or Block Order, $F(1,89) = 0.019$, $p = .89$, $h^2_p < .001$. This ANOVA did, however, reveal an interaction between Block Type and Block Order, $F(1, 89) = .25$, $p < .01$, $h^2_p = .11$. The ANOVA on crowd trials revealed similar patterns with no main effects across Block Type, $F(1, 89) = .005$, $p = .94$, $h^2_p < .001$, Block Order, $F(1,89) = 0.012$, $p = .91$, $h^2_p < .001$, or an interaction between the two, $F(1,89) = 1.72$, $p = 0.19$, $h^2_p = .02$. These results indicate that there was not a significant change in observer bias based on Block Type or Order (Figure 3).

Experiment 3

To more comprehensively understand the somewhat surprising effects of short-term learning of the prevalence of threat on negativity bias, we revisited and expanded upon our design from Experiment 1. In our final experiment, we examined the effect of valid explicit instruction, invalid explicit instruction, and no instruction (i.e. a re-examination of our implicit learning condition from Experiment 1) on observers' bias to endorse seeing anger.

Methods

Observers

One-hundred and forty-seven observers from the University of Denver (91 female, 51 male, five who preferred not to say) participated in Experiment 3. Observers provided informed consent and received partial course credit for their participation. All observers had normal or corrected-to-normal visual acuity and completed the experiment in a dimly lit room.

Stimuli and procedure

The stimuli and procedures were identical to those described in Experiment 1 with a few exceptions. In this experiment, in the same way as Experiment 2, we opted to present faces with expression intensities of 20% and 60%. Additionally, prior to participating in each experimental block (i.e. a high threat prevalence block or low threat prevalence block), observers were given information about the prevalence of emotion they would see across the upcoming trials. Using a between subjects' design, observers were given either valid information, invalid information, or no information. For the valid information condition, observers were provided with the actual threat

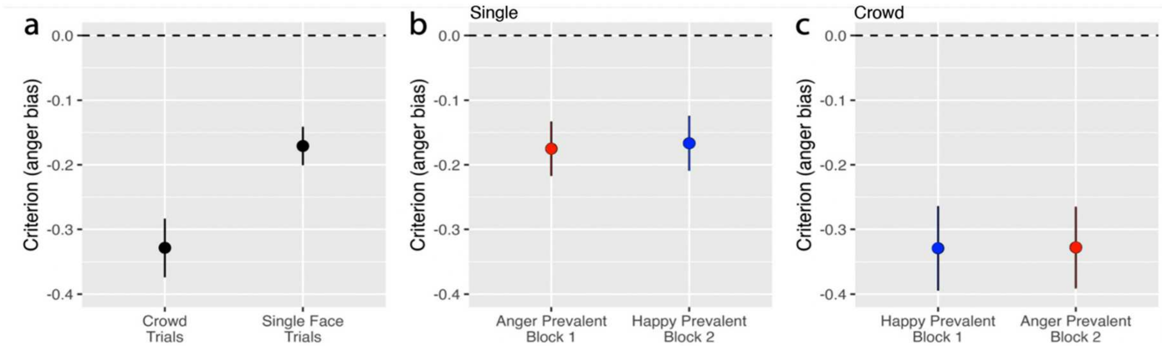

Figure 3. Criterion scores as an index of anger bias in Experiment 2. Negative criterion values indicate stronger bias to report that facial expressions are angry. (a) Bias as a function of numerosity across trials from all observers as a function of crowd or single-face context, which was manipulated within-observers. (b) Bias scores for judgments of single faces as a function of prevalence. (c) Bias scores for judgments of crowds as a function of prevalence.

prevalence of the upcoming experimental block. For example, prior to the high threat prevalence block, observers saw the following report: "On the following trials, 75% of trials will consist of faces with angry expressions. 25% of trials will consist of faces with happy expressions". For the invalid information condition, observers were given incorrect information about the threat prevalence of the upcoming experimental block. For example, prior to the high threat prevalence block, observers saw the following report: "On the following trials, 25% of trials will consist of faces with angry expressions. 75% of trials will consist of faces with happy expressions". For the no information condition, observers did not see an instruction screen and were automatically directed to the first experimental trial. Instructions were repeated halfway through the experiment before observers began the second block of trials. 50 observers completed the valid information condition, 51 observers completed the invalid information condition, and 46 observers completed the no information condition.

Results

Preliminary analyses

A repeated-measures ANOVA, with d-prime as the dependent variable, revealed main effects of Numerosity, $F(1,\ 146) = 32.46$, $p < .001$, $h_{p}^{2} = .18$, Intensity, $F(1,\ 146) =\ 46.94$, $p < .001$, $h_{p}^{2} = .24$, and an interaction between the two, $F(1,\ 146) = 27.86$, $p < .001$, $h_{p}^{2} = .16$. Next, a repeated-measures ANOVA, with criterion as the dependent variable, revealed main effects of Numerosity, $F(1,\ 146) = 32.46$, $p < .001$, $h_{p}^{2} = .18$, Intensity, $F(1,146) = 45.99$, $p < .001$, $h_{p}^{2} = .24$, and an interaction between the two, $F(1,\ 146) = 27.86$, $p < .001$, $h_{p}^{2} = .16$. This criterion analyses again demonstrated similar patterns to previous work (Mihalache et al., 2021), where observers overall exhibited a bias to report that faces had angry expressions. Again, this bias was amplified when faces had low expressive intensities compared to high intensities and when faces were presented within the context of a crowd compared to a single face.

Primary analyses

In order to examine an effect of Prevalence (high-threat-prevalence versus low-threat-prevalence) on criterion and to determine if that effect was dependent on Information Type (e.g. whether perceivers received accurate prevalence information, inaccurate prevalence information, or no prevalence information), we conducted a mixed ANOVA with criterion as our dependent variable, a within-subjects factor of Prevalence, and a between-subjects factor of Information Type. There was no main effect of Prevalence, $F(1,\ 144) = 0.014$, $p = .905$, $h_{p}^{2} = .01$, and no main effect of Instruction Type, $F(1,\ 144) = 0.812$, $p = .446$, $h_{p}^{2} < .001$. There was also no interaction between the two, $F(1,\ 144) = 0.514$, $p < .599$, $h_{p}^{2} < .001$. These results indicate that observers' bias to report anger regardless of expression was not significantly different for the high-threat-prevalence or low-threat-prevalence blocks (Figure 4). Further, there was no significant difference in anger bias based on the type of information observers received (e.g. Consistent, Inconsistent, None).

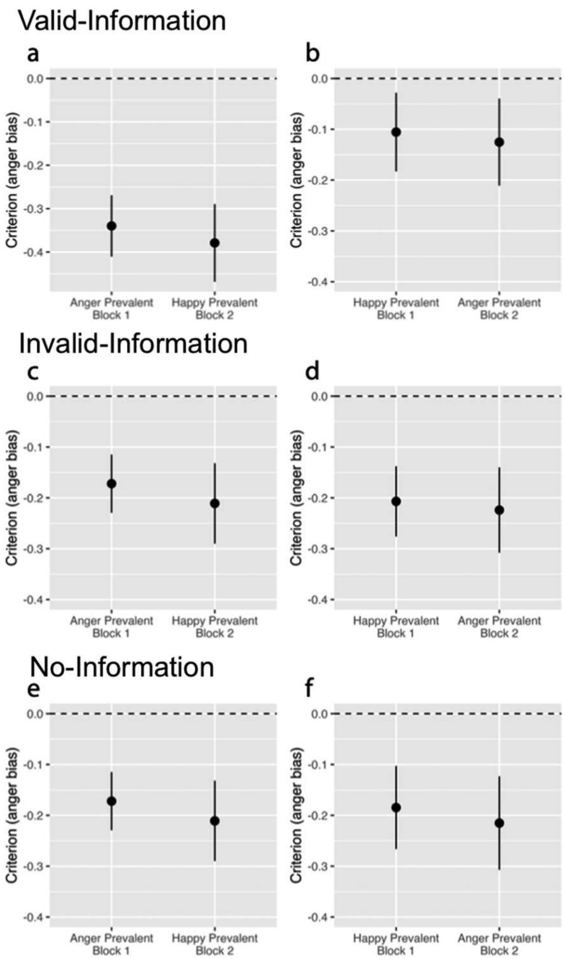

Figure 4. Criterion scores as an index of anger bias in Experiment 3. Negative criterion values indicate stronger bias to report that facial expressions are angry. The valid-information condition: (a) Bias scores for observers who completed block-order 1, which featured an angry-prevalent block followed by a happy-prevalent block. (b) Bias scores for the other half of observers who completed block-order 2, which featured a happy-prevalent block followed by an angry-prevalent block. Observers shown in panels a and b were explicitly given the correct prevalence information for each block prior to completing experimental trials. The invalid-information condition: (c) bias scores for observers who completed block-order 1, which featured an angry-prevalent block followed by a happy-prevalent block. (d) Bias scores for the other half of observers who completed block-order 2, which featured a happy-prevalent block followed by an angry-prevalent block. Observers shown in panels c and d were explicitly given the incorrect prevalence information for each block prior to completing experimental trials. The no-information condition (e) Bias scores for observers who completed block-order 1, which featured an angry-prevalent block followed by a happy-prevalent block. (f) Bias scores for the other half of observers who completed block-order 2, which featured a happy-prevalent block followed by an angry-prevalent block. Observers shown in panels e and f were not given any information about emotion prevalence for each block prior to completing experimental trials.

Discussion

The purpose of this investigation was to test the hypothesis that modifying the prevalence of angry faces in a perceiver's short-term visual environment would impact the strength of their bias to report that faces look angry. We predicted that experiencing a high (or low) prevalence of anger across a series of trials would lead perceivers to adopt a stronger (or weaker) anger bias. We tested this hypothesis in a binary angry/happy discrimination task, and we were particularly interested in whether this predicted effect of short-term statistical learning would occur when perceivers could only learn about the prevalence of angry and happy faces implicitly, over the course of a block of experimental trials. We based our hypothesis on evidence that this kind of shift in negative bias can occur when information about emotional prevalence is explicitly known (Lynn et al., 2012), as well as robust evidence that explicit and implicit learning about prevalence can influence bias in perceptual judgments outside the realm of emotion (Bohil & Wismer, 2015; Godwin et al., 2010, 2015; Maddox, 1995; Maddox & Bohil, 2003; Peltier & Becker, 2016, 2017; Wolfe & Van Wert, 2010). As we expected, observers were quite biased to report that faces were negative, and more specifically angry, replicating previous findings (Mihalache et al., 2021; Neta et al., 2009, 2013; Neta & Whalen, 2010). But surprisingly, we found no evidence that emotional prevalence impacted the strength of this anger bias. This was the case when observers learned about emotional prevalence either implicitly or explicitly. The fact that prevalence had no impact on the strength of anger bias demonstrates that anger bias is surprisingly resistant to short-term changes to one's statistical emotional visual environment. And yet we *did* find evidence that anger bias is malleable, in our case amplified by the context of a crowd or perceptual uncertainty. This again replicates and extends our previous work (Mihalache et al., 2021), illustrating that anger bias can yet be recalibrated based on visual cues in the moment, in this case in the face of dramatic and noticeable changes in the prevalence of negative and positive affect across trials.

A substantive base of evidence outside of emotion perception (e.g. shape and letter discrimination, weapon detection, and medical diagnoses) indicates that response biases should be flexible in the short-term and able to adaptively shift based on statistical regularities learned implicitly through experience or explicitly through instruction (Bohil & Wismer, 2015; Godwin et al., 2010, 2015; Maddox, 1995; Maddox & Bohil, 2003; Peltier & Becker, 2016; Peltier & Becker, 2017; Wolfe & Van Wert, 2010). And yet, we found that anger bias is surprisingly persistent and appears to remain unaffected by changes in prevalence, even when perceivers are explicitly told that anger will be a rare and infrequent occurrence. Our findings thus conflict with these reports and at least require us to revisit our predictions and assumptions about the nature of anger bias. It may be the case that the anger bias heuristic is so potent that its persistence and stability is more important than adaptability to short-term changes in the prevalence of emotion. Keeping one's bias fixed would at least allow a perceiver to manage the cost of errors in affective decision making consistently, albeit at some cost in terms of flexibility in the short term. For example, perceivers may be willing to persist with a strong angry bias even during a short run of trials with an abundance of happy faces because they would be prepared to catch more angry faces should the high prevalence shift from happy to angry faces at some point in the future.

Interpretation of null results

Interpretation of null results should be performed with caution. Nevertheless, we believe we can still be confident in the conclusions we have drawn for several reasons. First, the observers in our experiments completed many trials (e.g. 640 trials in both the happy and angry prevalence blocks for a total of 1280 trials in Experiment 1). We believe this number of trials is sufficient for a perceiver to implicitly learn that happy or angry faces were more prevalent and shift their bias accordingly. We can assume that if perceivers learned about prevalence, they did so primarily from the trials in which the expressions had high intensity (on which their sensitivity was highest), and that shifts in bias would have been more likely to occur on the trials where expressions had low intensities. There were 320 trials of each intensity within each block in Experiment 1 – still a substantial amount. Second, perceivers were clearly discriminating between the happy and angry faces, indexed by their high *d′* values across all three experiments. Our assumption that perceivers had an adequate opportunity to learn about prevalence would be on shaky ground if they were unable to identify the expressions, at least on most trials (a perceiver with a d′ of zero would have no perceptual "signal" from which to track the frequency of the expressions across a block). Third, our design was sensitive enough to easily capture other shifts in anger bias based on contextual cues *within* a trial. Across our three experiments we replicated the primary results of Mihalache et al. (2021), demonstrating that anger bias was amplified when observers viewed low-intensity emotional expressions compared to high intensity expressions and when faces were presented in the context of a crowd compared to a single face. Had we failed to produce this baseline effect of bias, we would again be on shaky ground for interpreting the lack of effects from prevalence. Replicating this previous work has its own intrinsic value, strengthening the evidence that anger bias is especially strong during judgments of multiple people compared to judgments of individuals.

We were especially surprised not to have found an effect of explicit knowledge of prevalence on the strength of anger bias in Experiment 3. We even manipulated the consistency of this explicit knowledge with the actual prevalence of emotional expressions, aligning the two sources of information or pitting them against each other. Neither appeared to matter as anger bias remained stable and strong. This conflicts with findings from Lynn et al. (2012) who found that perceivers adopted a weaker negative bias when discriminating between more and less threatening expressions after they were explicitly informed that more threatening faces would be less prevalent. We used the same base-rate values to manipulate prevalence in our design (75% versus 25%), but unlike their design, we displayed faces with both positive and negative valence. Perceivers in Lynn et al. also received feedback and were rewarded for correctly categorising expressions, whereas no such manipulations were included in the present investigation. Additionally, we included a more extreme manipulation in our experiment, examining the effect of both a high prevalence (75% angry trials) and low prevalence condition (25% angry trials) on anger bias. Our total number of trials differed for each condition, with Lynn et al. including 178 trials, while we included 640 trials. The aim of our experiment was not to replicate this prior work, but instead to take their theoretical motivations and extend it to our approach for measuring bias. We do not doubt the veracity of this previous work, nor do we question the authenticity of our own null results.

Rather, we suspect that effects of prevalence may be more sensitive to contextual cues and subtle variations in experimental design than we had anticipated.

Limitations and future directions

The current studies also have some important limitations. While we did include faces in our stimulus set belonging to two different racial groups, all the faces were of males. Our primary focus was to examine the effects of prevalence manipulations on bias, so we employed a highly controlled stimulus set, which allowed us to test our effects under restricted conditions (Mook, 1983). While this approach would not be lauded for its generalizability, it allowed us to examine how these perceptual mechanisms work, without the added noise and variability introduced with a more complex stimulus set that is more heterogenous in race and gender. New research examining of the effect of various demographic features (e.g. race, age, gender) on the effect of anger bias will be important to consider. And in fact, we include a supplemental analysis of anger bias and its interaction with the race of the faces in our stimulus set in the Supplemental Materials. Additionally, as we noted above, we believe that observers had ample opportunity to learn about the prevalence of angry and happy faces in each block of trials. It is possible, however, that a different design could allow perceivers to learn more rapidly. Future studies should examine factors that could improve the strength and rate of learning including a greater number of trials, or more trials with higher intensity expressions. It is also possible that some participants may have taken part in more than one of the experiments. This would have been a rare occurrence, however, and since the experiments were run over the course of multiple years, we do not believe that participation in more than one experiment would have been likely to influence a perceiver's performance on the task.

In conclusion, we replicated a recent finding showing that perceivers were biased to evaluate others as being angry, especially when they were seen in crowds. And yet the strength of anger bias was not affected by the prevalence of anger in our three experiments, neither when perceivers learned about the likelihood of encountering angry faces implicitly while they completed the experiments, or even when they were explicitly informed about it in advance. These results suggest that anger bias is indeed flexible, but more to contextual cues like the number faces being evaluated in a moment, and less to statistical cues about prevalence that are learned across short periods of time.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Adams, R. B., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and Emotion*, *30*(2), 177–186. https://doi.org/10.1007/s11031-006-9020-2

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Baumeister, R. F., Ainsworth, S. E., & Vohs, K. D. (2015). Are groups more or less than the sum of their members? The moderating role of individual identification. *Behavioral and Brain Sciences*, *39*(137), 1–56. https://doi.org/10.1017/S0140525X15000618

Bohil, C. J., & Wismer, A. J. (2015). Implicit learning mediates base rate acquisition in perceptual categorization. *Psychonomic Bulletin & Review*, *22*(2), 586–593. https://doi.org/10.3758/s13423-014-0694-2

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–674. https://doi.org/10.1017/S0140525X99002150

Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition*, *18*(10), 1439–1463. https://doi.org/10.1080/135-6285.2010.500605

Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M., & Donnelly, N. (2015). The effects of increasing target prevalence on information processing during visual search. *Psychonomic Bulletin & Review*, *22*(2), 469–475. https://doi.org/10.3758/s13423-014-0686-2

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(7), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

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

Cognition, *27*(5), 733–763. https://doi.org/10.1521/soco.2009.27.5.733

Holbrook, C., Galperin, A., Fessler, D. M. T., Johnson, K. L., Bryant, G. A., & Haselton, M. G. (2014). If looks could kill: Anger attributions are intensified by affordances for doing harm. *Emotion*, *14*(3), 455–461. https://doi.org/10.1037/a0035826

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 & Evolution*, *28*(8), 474–481. https://doi.org/10.1016/j.tree.2013.05.014

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(1), 1–17. https://doi.org/10.1017/S0140525X00041157

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-019-57204-1

Lynn, S. K., Zhang, X., & Feldmann-Barrett, L. (2012). Affective state influences perception by affecting decision parameters underlying bias and sensitivity. *Emotion*, *12*(4), 726–736. https://doi.org/10.1037/a0026765

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 288–301. https://doi.org/10.1037/0278-7393.21.2.288

Maddox, T. W., & Bohil, C. (2003). A theoretical framework for understanding the effects of simultaneous base-rate and payoff manipulations of decision criterion learning in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 307–320. https://doi.org/10.1037/0278-7393.29.2.307

Mihalache, D., Lamer, S. A., Allen, J., Maher, M., & Sweeny, T. D. (2021). Anger bias in the evaluation of crowds. *Journal of Experimental Psychology: General*, *150*(9), 1870–1889. https://doi.org/10.1037/xge0001025

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*(4), 379–387. https://doi.org/10.1037/0003-066X.38.4.379

Neta, M., Kelley, W. M., & Whalen, P. J. (2013). Neural responses to ambiguity involve domain-general and domain-specific emotion processing systems. *Journal of Cognitive Neuroscience*, *25*(4), 547–557. https://doi.org/10.1162/jocn_a_00363

Neta, M., Norris, C. J., & Whalen, P. J. (2009). Corrugator muscle responses are associated with individual differences in positivity-negativity bias. *Emotion*, *9*(5), 640–648. https://doi.org/10.1037/a0016819

Neta, M., & Whalen, P. J. (2010). The primacy of negative interpretations when resolving the valence of ambiguous facial expressions. *Psychological Science*, *21*(7), 901–907. https://doi.org/10.1177/0956797610373934

Oyama, T., Wantanabe, T., & Funakawa, M. (1983). Effects of test-mask similarity on forward and backward masking of patterns by patterns. *Psychological Research*, *45*(3), 303–313. https://doi.org/10.1007/BF00308709

Peltier, C., & Becker, M. W. (2016). Decision processes in visual search as a function of target prevalence. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1466–1476. https://doi.org/10.1037/xhp0000248

Peltier, C., & Becker, M. W. (2017). Target-present guessing as a function of target prevalence and accumulated information in visual search. *Attention, Perception, & Psychophysics*, *79*(4), 1064–1069. https://doi.org/10.3758/s13414-017-1297-6

Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, *11*(3), 300–311. https://doi.org/10.1162/089892999563409

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Vilanova, F., Beria, F. M., Costa, A. B., & Koller, S. H. (2017). Deindividuation: From Le Bon to the social identity model of deindividuation effects. *Cogent Psychology*, *4*(1), 1308104. https://doi.org/10.1080/23311908.2017.1308104

Whitney, D., & Yamanashi-Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-0044232

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121–124. https://doi.org/10.1016/j.cub.2009.11.066