Journal of Behavior Therapy and Experimental Psychiatry 65 (2019) 101498

Contents lists available at ScienceDirect

Journal of Behavior Therapy and Experimental Psychiatry

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

Targeting separate specific learning parameters underlying cognitive behavioral therapy can improve perceptual judgments of anger

Spencer K. Lynn[a,1], Eric Bui[b,c], Susanne S. Hoeppner[b,c], Emily B. O'Day[c,2], Sophie A. Palitz[c,2], Lisa F. Barrett[a,d,e], Naomi M. Simon[f,c,*]

[a] Department of Psychology, Northeastern University, 360 Huntington Ave, Boston, MA, USA, 02115
[b] Harvard Medical School, 25 Shattuck St., Boston, MA 02115, Boston, MA, USA, 02114
[c] Center for Anxiety and Traumatic Stress Disorders, Massachusetts General Hospital, 1 Bowdoin St., Boston, MA, 02114, USA
[d] Psychiatric Neuroimaging Division, Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, USA
[e] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, USA
[f] New York University School of Medicine, One Park Avenue 8th Floor, New York, NY, 10016, USA

ARTICLE INFO

Keywords:
Social anxiety disorder
Generalized anxiety disorder
Emotion perception
Decision making
Signal detection theory
Cognitive behavioral therapy
Anger perception

ABSTRACT

Background and objectives: Anxiety disorders are characterized by biased perceptual judgment. An experimental model using simple verbal instruction to target specific decision parameters that influence perceptual judgment was developed to test if it could influence anger perception, and to examine differences between individuals with social anxiety disorder (SAD) relative to generalized anxiety disorder (GAD) or non-psychiatric controls.
Methods: Anger perception was decomposed into three decision parameters (perceptual similarity of angry *vs.* not-angry facial expressions, base rate of encountering angry *vs.* not-angry expressions, payoff for correct *vs.* incorrect categorization of face stimuli) using a signal detection framework. Participants with SAD (n = 97), GAD (n = 90), and controls (n = 98) were assigned an instruction condition emphasizing one of the three decision parameters. Anger perception pre-*vs.* post-instruction and its interaction with diagnosis were examined.
Results: For all participants, base rate instructions impacted response bias over and above practice effects, supporting the validity of this instructional task-based approach to altering response bias. We failed to find a similarity or payoff instruction effect, nor a diagnosis interaction.
Limitations: Future instructional tasks may need to more closely target core cognitive and perceptual biases in anxiety disorders to identify specific deficits and how to optimally influence them.
Conclusions: This study demonstrates that specific decision parameters underlying perceptual judgment can be experimentally manipulated. Although our study failed to show diagnosis specific effects, it suggests that individual parameter "estimation" deficits may be experimentally isolated and potentially targeted, with the ultimate goal of developing an objective approach to personalized intervention targeting biased perceptual judgments in anxiety disorders.

1. Introduction

During social interactions, we look into the face of another person and in the blink of an eye infer that person's mental state, including their intentions and emotions. These perceptions inform our expectations about what the other person will do or say, as well as our decisions about how to best respond. The accuracy of mental state inferences can be understood as a decision–a process that guides behavior pursuant to knowledge about the environment. Each decision relies on our "estimates" of socio-environmental parameters. We developed a mathematical model of perceptual decision making to describe these parameters, aspects of their subjective estimation, and subsequent effects on emotion perception. Our model incorporates key insights from behavioral economics—expected value and optimality—into a signal detection theory (SDT) framework (Fig. 1; Lynn and Barrett, 2014; Lynn et al., 2018).

A large body of evidence shows that signal detection response bias is, in part, a function of a perceiver's evaluation of: (1) the base rate of

* Corresponding author. New York University School of Medicine, One Park Avenue 8th Floor, New York, NY, 10016, USA.
E-mail address: Naomi.simon@nyulangone.org (N.M. Simon).
[1] Present address: Charles River Analytics, 625 Mount Auburn St., Cambridge, MA, USA 02138.
[2] Present address: Temple University, 1801 N Broad St., Philadelphia, PA, USA 19122.

https://doi.org/10.1016/j.jbtep.2019.101498
Received 28 December 2018; Received in revised form 9 July 2019; Accepted 11 July 2019
Available online 12 July 2019
0005-7916/

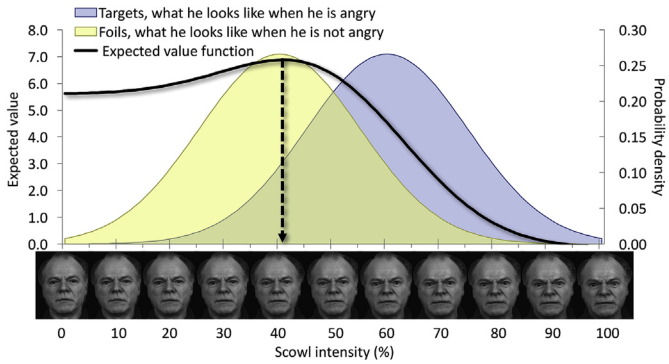

Fig. 1. The Signal Utility Estimator (SUE) model of emotion perception. As input, the model uses three experimenter-supplied parameters that define the decision environment: (1) the *similarity parameter*, modeled by the distributions of *target* and *foil* categories (here, *angry* and *not-angry* mental states, respectively) over some perceptual domain (e.g., a continuum of facial scowl intensity), (2) the *base rate parameter*, or relative frequency of encountering targets *vs.* foils, and (3) the decision *payoffs parameter*, or costs & benefits associated with correct or incorrect identification of targets and foils. The SUE output is the expected value (curve) of categorizing any given stimulus as a target *(i.e.,* as angry*)*. The point of maximum expected value is the optimal threshold location (dashed drop line) for a given set of parameter values. Judging all faces to the right of threshold as "*angry*" will maximize accrued benefit over a series of judgments. Signal detection theory characterizes the perceiver's ability to discriminate targets (e.g., facial and body actions indicating anger) *vs.* foils (e.g., non-threatening actions), called *perceptual sensitivity*, and the perceiver's tendency to identify signals (a given scowl intensity) as target *vs.* foil, called *response bias*. Our framework is aimed at assessing mental state inference with respect to the accuracy of a perceiver's subjective representations of the underlying parameters (e.g., Lynn et al., 2016).

occurrence of targets relative to foils (e.g., threats vs. non-threats), and (2) the payoff (benefits and costs) associated with correct *vs.* incorrect categorizations (e.g., Green & Swets, 1966). For example, a high base rate of encountering targets (angry people) demands a relatively low threshold for identifying even mildly scowling facial movements (i.e., the signal) as indicative of targets (e.g., anger). This is called *liberal* threat detection response bias. Similarly, a low cost for missed detections of targets calls for a relatively high threshold for identifying the signal as a target, called *conservative* threat detection response bias.

Strongly biased perceptual decision making is a salient feature of many psychiatric disorders. Social anxiety disorder (SAD) specifically is characterized by exaggerated concerns about negative evaluation and rejection in social situations with a bias to interpret or recall social interactions and events negatively (e.g., Alden, Taylor et al., 2008; Machado-de-Sousa, Arrais et al., 2010; Staugaard, 2010; Stopa & Clark, 2000). Cognitive behavioral theories suggest that in SAD, perceived threat and associated anxiety are enhanced by negative cognitions about the self and others in social situations that lead to hypervigilance toward social threat cues (e.g., Clark, Wells et al., 1995; Rapee, Heimberg et al., 1997). Wong and Rapee's (2016) conceptual causal model of the development and maintenance factors for SAD proposes that threat valuation is a key factor contributing to the development of "primary cognitive and behavioral processes that detect and eliminate social threat." This *social evaluative threat principle* proposes that primary cognitive (e.g., attention to self and threat in the environment) and behavioral processes (e.g., escape and avoidance) drive changes in other cognitive (e.g., negative self-evaluation cognitions and anticipatory processing) and neurobiological processes (e.g., hyperactivation of the amygdala in response to negative facial expressions). Elevations in assigned threat values contribute to elevated anxiety in social and performance situations; they conclude that many studies–including of scowling faces–support that elevated social anxiety is associated with enhanced detection and/or attention to threat (Wong & Rapee, 2016). Variation in the perception of anger in facial expressions is just one example of social evaluative threat that derives from perceptions of the external environment (i.e., people with whom an individual interacts). Although not fully generalizable to all social and evaluative situations and not the only possible negative expression type, scowling facial expressions have the advantage that they can be systematically manipulated and tested empirically when presented as stimuli in the laboratory setting, as utilized in different ways in many studies of SAD (e.g., Carré, Gierski et al., 2013; Evans, Wright et al., 2008; Grafton & MacLeod, 2016; Klumpp, Fitzgerald et al., 2015).

A broad, though not fully consistent, experimental literature has also supported an attentional bias towards emotional faces among those with social anxiety (e.g., Anderson, Dryman et al., 2013; Boal, Christensen et al., 2018; Liang, Tsai et al., 2017; McTeague, Laplante et al., 2018; Singh, Capozzoli et al., 2015). Similarly, in a set of emotion perception experiments in a signal detection framework with the same sample as the current study, anxiety severity was connected to differences in response bias (Lynn et al., 2018). Of relevance, prior research with self-reported responses to vignettes has further suggested that people with SAD overestimate both the base rate and the costs of aversive social interactions (e.g., Foa, Franklin et al., 1996; Uren, Szabó et al., 2004; Voncken, Bögels et al., 2003). Such altered perceptions can result in fear and avoidance of a wide array of social and performance activities and settings, leading to significant distress and impairment.

Cognitive behavioral therapy (CBT) for SAD seeks to address biases toward threat through a combination of cognitive restructuring techniques as well as behavioral techniques that enable gathering new evidence to support changes in cognitions and reduce avoidance (e.g., Clark, 2013, pp. 1–22). Cognitive interventions teach patients, via verbal instruction, to increase awareness of and change pathological misperceptions of the meaning and emotional significance of social and performance-related events through cognitive restructuring strategies. In CBT, patients explicitly learn to identify bias and gather evidence to "recalibrate" their perceptions to better reflect true environmental circumstances. For example, patients may be instructed to consciously evaluate the objective evidence that someone was truly judgmental of them and the consequences of this judgment if present. Patients do so by asking themselves: "*How often has the person really been judgmental in the past?*" (emphasis on the base rate parameter), "*What was the actual outcome in this instance?*" (emphasis on payoff), and "*What does the person do or say when they truly are or are not being judgmental?*" (emphasis on similarity). Each parameter corresponds to a mechanism by which an "affective recalibration" to environmental context may occur during CBT, which leverage principles of associative learning. CBT approaches do not, however, differentiate among aspects of stimulus representation that theories of learning and decision-making show to be independent, such as the three parameters highlighted above.

One question that has remained unaddressed is whether focusing on one specific decision parameter at a time might be a more precise and effective strategy for addressing overly biased perceptions. Targeting specific decision parameters could potentially lead to more personalized treatment of an individual's specific underlying emotion perception abnormalities. Another question is whether a strategy of using explicit knowledge tailored to a specific parameter estimate could reduce misperceptions of social threat in individuals with SAD, here experimentally modeled as judgments about anger perception in faces. Recent advances indicate that explicit instructions to examine salient cues (e.g., see Mogg, Bradley et al., 2016 for review of anxiety and attention to threat) may play a role in improving patient outcomes in people with anxiety disorders. For example, one study found that adding explicit directions to attention bias modification (ABM) that engage conscious attention such as adaptive positive visual search goals may result in better anxiety outcomes (Waters, Zimmer-Gembeck et al., 2015). This framework of threat and anxiety reflects more complex understanding that both bottom up, more automatic cognitive biases towards or away from threat, as well as top down systems involved in attentional control and threat evaluation likely interact in anxiety disorders, which may in part reflect imbalance of these systems (Mogg, Bradley et al., 2016). This study is a first step to examine how specific decisional parameters based on signal detection theory may be experimentally measured and influenced by a set of brief but specific simple standardized instructions, and whether such an experimental approach focused on anger perception is differentially effective in the setting of SAD compared to generalized anxiety disorder (GAD) and to the absence of psychiatric disorder.

1.1. The current study

Here, we experimentally explored the efficacy of parameter-specific instruction on response bias in emotion perception in individuals with SAD compared to individuals with GAD as well as a psychiatrically healthy control group. As reviewed above, prior studies indicate that SAD is associated with elevated social threat valuation (e.g., Wong & Rapee, 2016) and processing differences for depictions of negative-valence facial expressions, including anger, disgust, sadness, and fear, over positive-valence expressions (e.g., Machado-de-Sousa et al., 2010). We thus selected depictions of anger on faces for our paradigm based on this prior research and theories, and because of its congruence with symptoms of social anxiety, such as inferring a critical social evaluation.

Because GAD is a disorder that involves anxious worry not specifically focused on social perception, it was selected as an active control comparator for SAD. Specifically, while research with emotion perception and GAD has found mixed results with both greater and reduced emotion perception (e.g., Bui, Anderson et al., 2017; Rutter, Scheuer et al., 2019), this has tended to be present across emotions and not specific to social or evaluative threat cues. Thus, even though there is some suggestion that neural correlates of reactivity to scowling faces may be transdiagnostic (MacNamara, Klumpp et al., 2017), comparing SAD to GAD enables a control for these transdiagnostic factors potentially related to heightened non-specific anxiety and nervousness that might influence the experimental anger perception manipulation results. Thus, if participants with SAD behave differently from non-anxious controls in the parameter-specific manipulations of perceptual judgments about anger in faces, having a GAD control group would enable assessment of whether or not this might be a non-specific anxiety effect versus a specific effect fund only in those with social threat concerns.

We compared the effect of different sets of brief verbally-delivered instructions about three distinct aspects of anger perception. One instruction set drew attention to the base rate of occurrence of targets relative to foils. A second instruction set drew attention to the payoffs. A third instruction set drew attention to the perceptual similarity between targets and foils.

In a pre-*vs.* post-instruction comparison, our aims were to investigate if simply drawing attention to salient aspects of these judgments would affect response bias, and to compare the effect among participants with SAD, GAD, and psychiatrically healthy control participants. Specifically, we hypothesized that drawing people's attention to one of three specific environmental parameters theorized to influence response bias in signal detection theory would affect how people learn to categorize facial expressions.

Concerning *response bias*, we made two predictions:

(1) Because our experimental design implemented a liberal-bias base rate value, we predicted that, by drawing participants' attention to the base rate, our **base rate instruction** would result in a more liberal response bias (tendency to judge faces as angry) in the post-instruction task than in the pre-instruction task.
(2) Because the our experimental design implemented a conservative-bias payoff matrix, we predicted that our **payoff instruction** (emphasizing parameter values that affect response bias in the opposite direction of the base rate), by drawing participants' attention to the payoffs, would produce a more conservative response bias (tendency to judge faces as not angry) in the post-instruction task than in the pre-instruction task.

Concerning *perceptual sensitivity*, we predicted that our **similarity instruction,** by drawing participants' attention to features that distinguished our emotion categories, would result in higher perceptual sensitivity in the post-instruction task than in the pre-instruction task. If change in sensitivity following similarity instruction was no greater than in other conditions, this would argue for a practice effect on perceptual sensitivity.

Concerning diagnosis groups, we have previously speculated that a “misestimate” of base rate and/or payoffs may be computationally responsible for the interpretive bias exhibited by individuals with anxiety disorders (Lynn et al., 2018). Although SAD is a condition more clearly linked to abnormalities in *social* threat perception than are the GAD or control conditions, it could be the case that SAD is associated with either hyper- or hypo-sensitivity to information about base rate and/or about payoff. We predicted that SAD would show a stronger influence of one or more instruction conditions than the control group (e.g., consistent with theories that abnormalities in social threat valuation can be influenced by cognitive interventions in SAD: Wong & Rapee, 2016). We did not have predictions about specific instruction conditions vis-à-vis diagnosis group.

2. Methods

2.1. Participants & assessments

A total of 293 participants, men and women aged 18–65, were recruited by local hospital and media advertising. Experienced and trained clinician interviewers assessed participants for psychiatric diagnoses with the Structured Clinical Interview for DSM-IV (SCID; First, Spitzer et al., 2002). Participants included those with a primary diagnosis of current SAD (n = 101), GAD (n = 91), and psychiatrically healthy controls (n = 101). Other current psychiatric diagnoses were exclusionary (apart from specific phobia for participants with SAD or GAD), including GAD for SAD participants and SAD for GAD participants, respectively. Alcohol or substance abuse or dependence within the past 6 months, lifetime history of psychotic disorders, mental disorder due to a medical condition or substance use, or dementia were also exclusionary. Unstable severe medical illness, history of seizure disorder, current use of psychiatric medications, positive urine toxicology screen test, and uncorrected visual impairment were also exclusionary. Psychiatrically healthy control participants had no current or lifetime history of a DSM-IV Axis I diagnosis and were required to meet the same inclusion/exclusion criteria as the SAD and GAD participants.

Participants were consented, screened, and tested at the Center for Anxiety and Traumatic Stress Disorders at the Massachusetts General Hospital between December 21, 2011 and September 21, 2016. The present research was part of a larger study that involved completing questionnaires and participating in a series of perception experiments (R01MH093394). Following the clinical assessments, participants completed several versions of the perception tasks described here as part of the larger study (see Lynn et al., 2018). Of the participants enrolled in the larger study, 285 (97.3%) returned within one week to complete additional versions of the perception tasks, concluding with the task described here. Participants were compensated $150 after study completion. Thus, following attrition, 97 participants with SAD, 90 participants with GAD, and 98 psychiatrically healthy control participants participated in the task described here. Informed consent was obtained prior to participation. Study procedures were approved by the Partners Healthcare Institutional Review Board.

2.2. Perception task

We created stimulus sets, each depicting a range of facial scowl intensity, from 0 to 100% scowl in 10% increments (Fig. 1 depicts one set). Faces were shown one at a time on a computer screen. Participants categorized the faces as angry or not angry by answering the on-screen prompt "*Was that person angry?*", using their index fingers to press USB keyboard buttons labeled "*Yes*" and "*No*". Additional details of the general stimulus design procedure are described elsewhere (e.g., Lynn et al., 2014, 2016, 2018).

The optimal decision criterion for "angry" *vs*. "not-angry" faces that participants should seek to adopt in the tasks was determined by the values of the three SDT parameters as programmed into the experiment software: they should match their behavior to the contingencies of the task. The perceptual similarity parameter was implemented by drawing target and foil trials from Gaussian signal distributions imposed over the 11-item stimulus set. Mean target and foil stimuli were 60% and 40% scowl intensity, respectively, both with standard deviations of 15% scowl intensity (depicted in Fig. 1). The base rate was implemented as the proportion of target to foil trials. The base rate was 0.78: 78% of the trials were drawn randomly from the target distribution and had the correct answer "*Yes*" (i.e., angry) regardless of their signal strength. The payoff parameter was implemented as points earned or lost for correct and incorrect categorization of a stimulus as target or foil. Correct detections and correct rejections earned 10 points, missed detections lost 3 points, and false alarms lost 10 points.

The high base rate calls for a low threshold (liberal response bias). The low cost of missed detections relative to false alarms demands a high threshold (conservative response bias). The liberal-going effect of the base rate on response bias mathematically overwhelms the conservative-going effect proscribed by the payoffs, so that the observed bias of participants is expected to be liberal. The standard deviations set a perceptual sensitivity cap of d' = 1.3. At that sensitivity, optimal response bias was $c = -0.6$ (depicted in Fig. 1).

Each participant was presented with 230 trials. Each trial began with a white fixation cross (300 ms duration) centered on a black screen, followed by a face stimulus (500 ms duration). The response prompt then followed the face and remained on-screen until the participant responded. Participants received immediate on-screen feedback (including "*Yes - that was right*" or "*No - that was wrong*", the points just earned, and their cumulative points). Participants were instructed to earn as many points as they could (thereby optimizing their categorization of the stimuli). A 300 ± 100 ms inter-trial interval (black screen) followed the feedback. Participants finished the task in approximately 10 min. Stimulus set and response button location ("z" or "/" key) were counterbalanced across participants. The task was programmed in Matlab (The Mathworks, Inc.) with Psychophysics Toolbox (version 3; (Brainard, 1997).

2.3. Parameter-specific instruction

Participants completed the perception task twice, once before receiving instruction about one of the three signal detection parameters, and then again following that instruction. Participants were assigned to one of the instruction conditions via counter-balancing blocked by diagnosis, independent of other participant characteristics. We created three brief presentations (in Microsoft PowerPoint), one focusing on each of the three signal detection parameters (see Fig. 2). Each of the three instruction types briefly highlighted the connection between angry faces and social threat in the environment. The base rate instruction (panel B) informed participants of the likelihood of angry faces (comparable to cognitive interventions targeting correction of errors about likelihood of threat). The payoff instruction (panel C) included a higher point value ("cost" of 10 points lost) for false alarms than for missed detections (3 points lost), while rewarding correct categorizations of angry or non-angry faces equally. This task thus essentially incentivized a correction to the tendency to overvalue risk of missed detections of anger by including a greater penalty for over-interpretation of ambiguous faces as angry. These three conditions were intended to experimentally isolate simple instructions consistent with cognitive work done in cognitive behavioral therapy (CBT) to correct maladaptive biases resulting in hypervigilance to any suggestion of negative responses (e.g., to not miss any critical judgment or anger) during social or performance situations, as occurs in SAD. The similarity instruction (panel D) guided participants to attend to cues ("evidence") for their conclusions about the presence or absence of anger.

After completing the task the first time, a researcher administered the parameter-specific instruction and read each screen aloud to the participant. After answering any questions, the researcher started the post-instruction task. Following trials 20 and 40 of the task, the computer paused the task and informed the participant of his or her overall performance with the statement, "*You have earned ##% of the points we would expect.*" The researcher then said to the participant, "*In the next section, remember to keep thinking about [how likely someone is to be angry, benefits and costs of being correct or mistaken, the similarity between how a person looks when he or she is angry vs. not angry]*" and resumed the task. Following trial 41, the researcher left the room until the end of the task.[3]

2.4. Data analytic approach

Differences between diagnostic groups in basic demographics, consisting of age, gender (men *vs.* women), and race (coded into white *vs.* non-white), were examined with a one-way ANOVA for age and chi-square tests for gender and race. From the perception task data, we used the observed frequencies of correct detections (CDs), missed detections (MDs), false alarms (FAs) and correct rejections (CRs) to calculate *response bias* (*c*) and *perceptual sensitivity* (*d'*) (Macmillan & Creelman, 1991; Stanislaw & Todorov, 1999) as dependent measures. We calculated response bias as $c = -0.5*(z(CDs/(CDs + MDs)) + z(FAs/(FAs + CRs)))$ and perceptual sensitivity as $d' = z(CDs/(CDs + MDs))$-$z(FAs/(FAs + CRs))$, where z is the inverse of the normal cumulative distribution function. Trials with a response time < 250 ms were excluded from the calculations due to the likelihood of containing motor errors (2–4% of trials were excluded for approximately 20% of participants). We conducted repeated measures ANOVAs to test the effects of primary diagnosis (healthy control *vs.* SAD *vs.* GAD), time (pre-*vs.* post-instruction), and instruction type (base rate *vs.* payoff *vs.* similarity), as well as their interactions, on response bias and sensitivity. The key effects of interest were the interactions involving time, which indicated whether outcome changed from pre-to post-instructions. We also added gender, race (white *vs.* not white), and age as covariates to the models. We accounted for the correlation of within-subject observations by modeling time (pre-*vs.* post-instruction) as repeated measures per person with an unstructured covariance structure, and modeled stimulus sets (i.e., different photographic models) as a random effect. We used pairwise comparisons with false discovery rate (FDR) adjustment to investigate pre-*vs.* post-instruction differences for each type of instruction across all diagnoses, as well as within each diagnostic group. To test differences between diagnostic groups, we used specific contrasts that compared the pre-to post-instruction changes in each anxiety group (SAD or GAD) to healthy controls for each type of instruction, again adjusting for multiple comparisons with FDR adjustments. Data

[3] Approximately half of the participants received an earlier version of these instruction sets that contained additional explanatory text and did not include feedback at trials 20 and 40. We revised the instructions to improve clarity. The revision did not statistically significantly affect the pattern of results, and so we have not included the revision as a factor in the statistical analyses presented here.

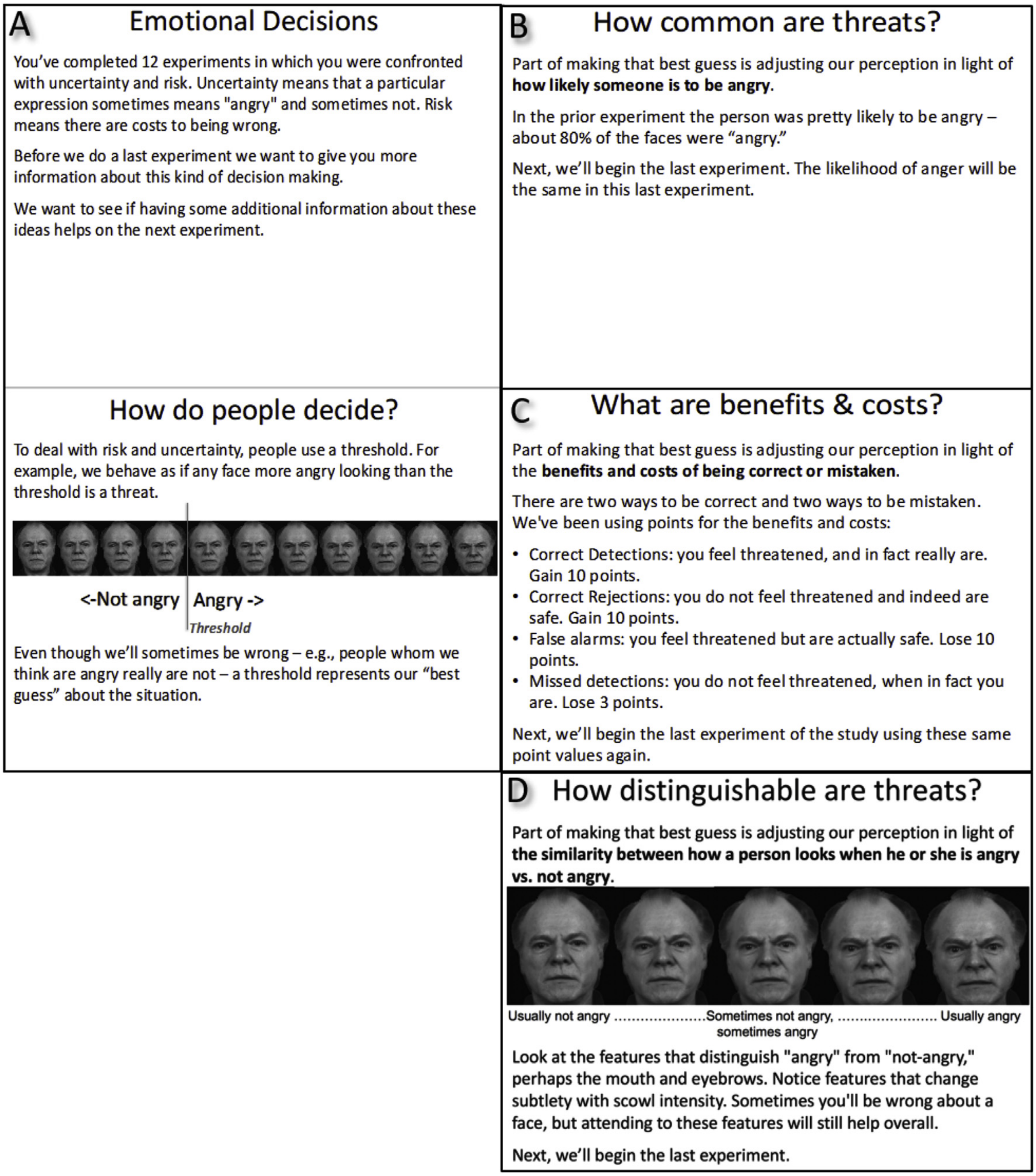

Fig. 2. Parameter-specific instruction. (A) All three instruction sets began with an introduction to the concepts of perceptual uncertainty and behavioral economic risk. (B) The base rate condition concluded with a description of the likelihood of encountering anger. (C) The payoff condition concluded with a description of benefits and costs. (D) The similarity condition concluded with a description of the association between facial expression intensity and the angry *vs.* not angry categories.

are expressed as estimated population marginal means (i.e., least-square means, LSM) with standard errors (SE), which were adjusted for covariates and other effects in the model. Within-group effect sizes of pre-instruction task to post-instruction task changes were calculated based on raw means and standard deviations and adjusted for pre-to post-instruction correlations (i.e., Hedges g_{av}; Lakens, 2013). Between group effect sizes are Cohen's d. All analyses were performed using SAS (Version 9.4 of the SAS System for Windows). Significance was evaluated at $\alpha = 0.05$ (two-tailed) unless otherwise noted.

3. Results

3.1. Sample characteristics

The 285 participants included more women (62.5%) and European Americans (69.4%), with a mean age of 31.5 ± 12.9 years. Diagnostic groups did not differ significantly in age, $F(2, 282) = 1.25$, $p > 0.28$. The SAD group trended towards a higher proportion of women (72.2% compared to 59.2% for controls and 56.7% for participants with GAD; $X^2(\mathrm{df} = 2) = 5.48$, $p = 0.06$), and included more white participants (84.4% compared to 61.9% for controls and 62.9% for participants with GAD; $X^2(\mathrm{df} = 2) = 14.12$, $p = 0.0009$).

Table 1
Type-3 tests of fixed effects of the three instruction types on response bias and perceptual sensitivity in participants with social anxiety (n = 97), generalized anxiety (n = 90), and in non-psychiatric controls (n = 98).

Effect	df	F	p
Bias (c)			
Diagnosis	2, 272	1.33	0.2651
Time (pre- *vs.* post-instruction)	1, 275	19.34	< 0.0001
Diagnosis x Time	2, 275	2.48	0.0853
Instruction type	2, 272	2.40	0.0924
Diagnosis x Instruction type	4, 272	0.65	0.6253
Time x Instruction type	2, 275	8.40	0.0003
Diagnosis x Time x Instruction type	4, 275	0.78	0.5391
Gender	1, 272	0.00	0.9672
Race (white *vs.* non-white)	1, 272	1.92	0.1670
Age	1, 272	3.77	0.0531
Sensitivity (d')			
Diagnosis	2, 266	0.40	0.6704
Time	1, 275	23.12	< 0.0001
Diagnosis x Time	2, 275	0.46	0.6295
Instruction type	2, 266	0.45	0.6389
Diagnosis x Instruction type	4, 266	0.46	0.7670
Time x Instruction type	2, 275	0.04	0.9590
Diagnosis x Time x Instruction type	4, 275	0.41	0.8001
Gender	1, 271	3.39	0.0669
Race (white *vs.* non-white)	1, 268	0.36	0.5506
Age	1, 269	15.78	< 0.0001

3.2. Response bias and parameter specific instructions

Response bias became more negative overall (i.e., more liberal-going, a tendency to judge faces as angry) from pre-to post-instruction (across all instruction types and diagnoses LSM ± SE: -0.60 ± 0.03 *vs.* -0.72 ± 0.03, $p < 0.0001$; Table 1). This effect was largely driven by a moderate effect of instruction on response bias in the group that received the base rate instruction (pre-*vs.* post-base rate instruction, across diagnoses, LSM ± SE: -0.60 ± 0.06 *vs.* -0.86 ± 0.05, adj. $p < 0.0001$, $g_{av} = -0.48$), as predicted. Contrary to prediction of causing a more conservative response bias, the payoff instruction did not result in significant change in response bias (across diagnoses, LSM ± SE: -0.59 ± 0.06 *vs.* -0.60 ± 0.05, adj. $p = 0.93$, $g_{av} = -0.00$). The similarity instructions were not expected to affect response bias, and, as anticipated, did not result in any overall change in response bias (across diagnoses, LSM ± SE: -0.62 ± 0.06 *vs.* -0.70 ± 0.05, adj. $p = 0.07$, $g_{av} = -0.20$).

3.3. Response bias and diagnosis

Of note, we did not find any difference in response bias between the 3 diagnostic groups prior to instruction (M ± SD SAD *vs.* control: -0.65 ± 0.53 *vs.* -0.53 ± 0.56, $d = -0.23$; GAD *vs.* control: -0.59 ± 0.50 vs. -0.53 ± 0.56, $d = -0.11$; main effect of diagnosis, Table 1). We also found no evidence of diagnosis-specific responses to instructions overall, or to specific instructions, after adjusting for multiple tests (diagnosis interactions, Table 1). The pre-to post-instruction effect (i.e., within-group effect) of the base rate instruction was small to medium for non-psychiatric controls ($g_{av} = -0.41$) and medium for both groups of patients with anxiety diagnoses (SAD: $g_{av} = -0.49$, comparison to control: adj. $p = 0.76$; GAD: $g_{av} = -0.60$, comparison to control: adj. $p = 0.93$; Fig. 3). The effect of the payoff instruction was very small for all diagnostic groups (control: $g_{av} = -0.17$; SAD: $g_{av} = 0.01$, comparison to control: adj. $p = 0.54$; GAD: $g_{av} = 0.21$, comparison to control: adj. $p = 0.20$). While the effect of similarity instruction was medium to large for non-psychiatric controls ($g_{av} = -0.68$; within-group adj. $p = 0.0346$) compared to small to negligible for participants with anxiety diagnoses (SAD: $g_{av} = -0.19$; GAD: $g_{av} = 0.08$), we were unable to detect a difference in this change between diagnostic groups (SAD comparison to control: adj. $p = 0.54$; GAD comparison to control: adj. $p = 0.20$). None of the covariates included significantly predicted bias (Table 1).

3.4. Perceptual sensitivity

Prior to receiving the instructions, we were not able to detect a difference between either the SAD group or the GAD group and non-psychiatric controls in perceptual sensitivity (M ± SD SAD vs. control: 0.69 ± 0.47 vs. 0.76 ± 0.38, $d = -0.17$; GAD vs. control: 0.73 ± 0.37 vs. 0.76 ± 0.38, $d = -0.08$). Perceptual sensitivity improved slightly from pre-to post-instruction (pre-*vs.* post for all instruction conditions, all diagnoses, LSM ± SE: 0.71 ± 0.03 *vs.* 0.82 ± 0.03, $p < 0.0001$, Table 1). Contrary to predictions that the similarity instruction would result in greater perceptual sensitivity and that this effect might vary by diagnosis, we were not able to detect a difference in the pre-to post-instruction effect between instruction types or diagnoses (Table 1). The pre-to post-instruction (*i.e.*, within group) effect sizes ranged from 0.15 to 0.40 for all instruction types for each diagnosis (base-rate instruction type: control $g_{av} = 0.18$, SAD $g_{av} = 0.27$, GAD $g_{av} = 0.36$; payoff instruction type: control $g_{av} = 0.40$, SAD $g_{av} = 0.31$, GAD $g_{av} = 0.15$; similarity instruction type: control $g_{av} = 0.19$, SAD $g_{av} = 0.41$, GAD $g_{av} = 0.36$). Age had a significant effect on sensitivity, indicating that sensitivity increased with age ($b = 0.0061 \pm 0.0015$, $p < 0.0001$).

4. Discussion

We examined the effects of receiving three types of brief, simple instructions on two independent aspects of anger perception in an experimental task administered to a large sample of well-characterized adults with a clinical diagnosis of SAD or GAD, and healthy controls. Instructions separately emphasized the relative base rate of encountering depictions of angry *vs.* not angry facial expressions in the task, the payoffs (costs and benefits) of correctly *vs.* incorrectly categorizing when a face depicted anger or not, or the similarity between what angry and not angry facial depictions looked like.

Base rate instruction was effective in modulating response bias; as predicted, participants who were instructed that 80% of the faces presented were likely to be angry consistently showed a more liberal response bias (i.e., identified more faces as angry) following the instruction. In this experiment, a practice effect on response bias could be expected to manifest as more liberal response bias in the post-instruction task than in the pre-instruction task because, in biased signal detection tasks, perceivers tend *not* to achieve enough bias to actually maximize their net benefit (Green & Swets, 1966; Maddox & Bohil, 2005); including in our prior work using designs similar to the present task, e.g., (Lynn et al., 2012, 2014). A lack of effect on response bias following either similarity or payoff instructions, indicated that this change was not merely due to a practice effect; of note, similarity instruction was not anticipated to impact response bias but rather designed to influence perceptual sensitivity.

Contrary to hypotheses, response bias did not change following payoff instruction, which had been designed to result in a more conservative response bias. It is possible that this failure to result in the anticipated effect may be explained by some aspects of the experimental design. Notably, the point values themselves were structured, with the cost of missed detections less than the cost of false alarms, such that the payoff parameter should impact response bias in the opposite direction of the base rate parameter. However, response bias during the task was dominated by the high base rate; optimal (points-maximizing) response bias on the task was relatively liberal overall (illustrated in Fig. 1). Thus, the strong base rate (i.e., 80% targets) may have interfered with an ability to observe an effect of drawing people's attention to benefits and costs in the payoff instruction condition. Future studies should consider utilizing tasks with a lower base rate, which may also better recapitulate emotion perception decision making in situations of

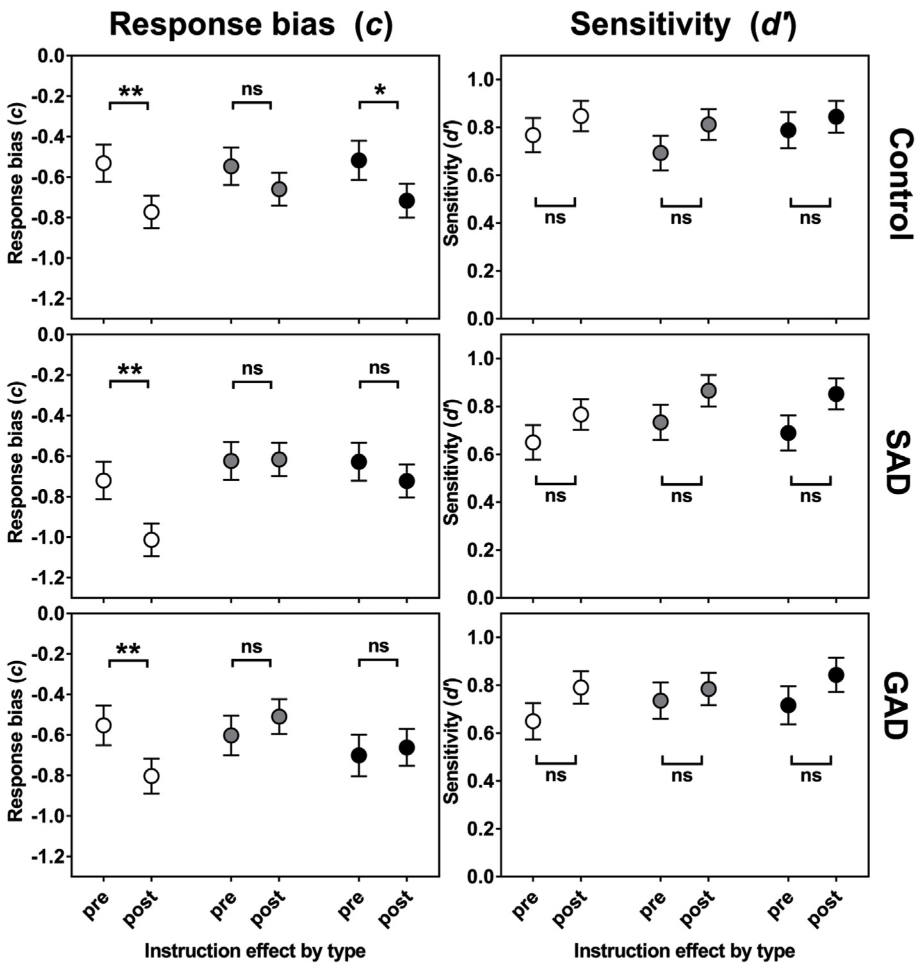

Fig. 3. Estimated population marginal means (least-square mean, LSM ± SE, adjusted for covariates and other effects in the model) of response bias and sensitivity from pre-to post-instruction in each diagnostic group for each instruction condition (white circles for *base rate* instruction, grey circles for *payoff* instruction, and black circles for *similarity* instruction). There were 98 participants in the non-psychiatric control group (base-rate n = 34; payoff n = 34; similarity n = 30), 97 in the SAD diagnostic group (base-rate n = 33; payoff n = 32; similarity n = 32), and 90 in the GAD diagnostic group (base-rate n = 31; payoff n = 31; similarity n = 28). Significant changes from pre-to post-instruction are indicated with * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

uncertainty.

The similarity instructions in this experiment were also ineffective at increasing perceptual sensitivity. There were numerical, though not statistically significant, increases in sensitivity across all participant groups in all instruction conditions. Therefore, similarity instruction did not produce strong evidence of an effect on perceptual sensitivity over and above any practice effect (i.e., improved performance due merely to increased learning afforded by a second exposure to the task). The failure of our similarity instruction to produce a notable improvement in perceptual sensitivity suggests that the instructions and approach can be improved. Difficulty impacting perceptual sensitivity via simple verbal instruction is consistent with the idea that simply pointing out facial features of emotion or directing attention to facial detail may be insufficient to change perception of emotion, and more complex instructions or behavioral tasks (e.g., guided practice) may be needed.

Limitations of this study include: (1) the potential confounding effect of testing fatigue as the instructions tasks were administered at the end of a fairly long signal detection session (i.e., 1–2 h); (2) the use of point values to represent payoff that might not be completely ecologically valid, such that correct identification of a facial expression may not have the same relevance to people as the negative personal consequences they might fear; (3) the limited power to detect interactions as the sample sizes (n ~ 30) were fairly small at the instruction by diagnosis level of the experiment; (4) the effect of instruction being examined under only one of many possible decision environments (combinations of parameter values). We created an experimental environment that was liberally biased by a high base rate of threat, and implemented unequal decision costs. While biased environments can more realistically model the exigencies of judgments made outside the laboratory than do the neutral-bias environments implemented in nearly all research in perceptual decision making, our results (both positive and null) may be somewhat dependent on the specific parameter values under which they were obtained (a caveat that also applies to results obtained from the neutral-bias environments). For example, a different decision environment, such as one demonstrated a priori to have a clear difference in performance between those with social anxiety or other diagnosis groups and healthy controls, might serve as a better vehicle for examining the effect of targeted parameter-specific instruction for individuals with one diagnosis versus another, or more generally with impairments in a specific parameter. It is also likely that our simple set of instructions delivered only once was an insufficient "dose" for a patient sample meeting diagnostic criteria for SAD. Future studies could either increase the frequency of the targeted instructions or examine a less severe population. Alternately, the parameters of the task and instructions could be matched to individual participant's baseline biases, and adapted over repetitions of the task. Finally, in our task, we gave brief feedback about points earned, but did not provide information about the specific types of errors made, which could potentially result in greater success learning to change response bias or correct perceptual sensitivity errors. Despite these limitations, our results nonetheless suggest that brief verbal instructions clarifying the base rate of encountering threat improves emotion perception performance (i.e., response bias closer to the real base rate of threat in the environment). This study supports the speculation that different signal detection parameters influencing emotion perception in individuals can be separated, measured, and at least in the case of response bias, be influenced by simple instructions (about base rate in this case). While other aspects of our experimental design likely would benefit from additional optimization, this initial result supports that additional research is indicated to optimize measurement of the impact of specific parameters on perceptual biases, and how best to influence change in these parameters with instruction. Elsewhere, we have speculated that individual impairments in social perception may be characterized by cognitive “misestimates” of one or more of the three signal detection environmental parameters investigated here (Lynn and Barrett, 2014; Lynn et al., 2018). If a patient's misinterpretation of environmental cues can be traced to an impaired estimate of a particular parameter (e.g., an over-estimate of the base rate of encountering threats, consistent with cognitive behavioral theories about over-estimation of and hypervigilance to perceived social threat in SAD), then the present results suggest that interventions specifically addressing at least the base rate parameter may be helpful to enhance correction of this problematic parameter (i.e., base rate assumption errors). While this initial experiment utilized single, very brief instructions substantially less intense or complex than a course of psychotherapy, this experimental approach if optimized could potentially develop a model that has implications for future more personalized psychosocial treatment targets.

CBT, for example, also successfully targets misestimation of the base rate of threat in the environment through verbal cognitive interventions as well as behavioral learning. While our experimental instructions were designed as a pilot study to examine the potential effects of brief uniform sets of parameter-specific instructions delivered by a non-clinician, and in this experiment, were designed to increase the correct categorization of anger from face stimuli depicting variably scowling expressions, CBT approaches for fear-based anxiety disorders would usually aim to decrease an over-estimation of threat in the environment. Thus, while our results confirm that verbal instructions regarding the frequency of the occurrence of threat in the environment may improve threat perception, future studies may wish to alter the experimental conditions to use base rate instruction to reduce threat perception. This experimental finding, however, may have implications for behaviors (and psychiatric conditions that may include high-risk behaviors) in which threat is under-estimated. Although social threat perception abnormalities are particularly clinically relevant for individuals with SAD, we did not have specific hypotheses about the expected magnitude of change by diagnosis and were not able to detect any differences between people with SAD or GAD and/or non-psychiatric control participants in response to the instruction sets.

5. Conclusions

We used a signal detection framework to break down an anger perception task into component decision parameters. In a scenario with a high base rate of threat signal, we found that base rate instructions affected response bias change over and above practice effects across individuals with SAD and GAD as well as a non-anxious control group, while similarity and payoff instructions had no discernible effects on either response bias or perceptual sensitivity. Although our study is not without limitations and would benefit from replication targeting specific parameters demonstrated to differ across diagnoses or individuals, the ability to target specific learning parameters that underlie perceptual judgment biases suggests the downstream potential to further develop this approach with an aim to ultimately target psychosocial approaches to a participant's individual parameter “estimation” deficit to achieve more efficient personalized interventions. However, this conclusion remains highly speculative. Future research should examine different approaches to specific parameter instruction, particularly for the payoff and similarity parameters, and under a variety of base rates, payoffs, and similarity values.

Funding

This work was supported by the National Institutes of Health [grant R01MH093394 to Naomi Simon and Spencer Lynn]. The views, opinions, and/or findings contained in this paper are those of the authors and shall not be construed as an official National Institutes of Health position, policy, or decision, unless so designated by other documents. The sponsor had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. We also thank our study participants and the study staff at the Massachusetts General Hospital Center for Anxiety and Traumatic Stress Disorders.

Conflicts of interest

The authors declare that they have no conflicts of interest with this study topic and received no funding that would affect the objectivity of this investigation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbtep.2019.101498.

References

Alden, L. E., et al. (2008). Social anxiety and the interpretation of positive social events. *Journal of Anxiety Disorders, 22*(4), 577–590.
Anderson, E. C., et al. (2013). Smiles may go unseen in generalized social anxiety disorder: Evidence from binocular rivalry for reduced visual consciousness of positive facial expressions. *Journal of Anxiety Disorders, 27*(7), 619–626.
Boal, H. L., et al. (2018). Social anxiety and attentional biases: A top-down contribution?" attention, perception. *& Psychophysics, 80*(1), 42–53.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.
Bui, E., et al. (2017). Heightened sensitivity to emotional expressions in generalised anxiety disorder, compared to social anxiety disorder, and controls. *Cognition & Emotion, 31*(1), 119–126.
Carré, A., et al. (2013). *Linear association between social anxiety symptoms and neural activations to angry faces: From subclinical to clinical levels*, *Vol. 9*, 880–886 6.
Clark, D. A. (2013). *Cognitive restructuring. The Wiley handbook of cognitive behavioral therapy.* 1–22.
Clark, D. M., et al. (1995). A cognitive model of social phobia. *41*(68), 00022–00023.
Evans, K. C., et al. (2008). *A functional MRI study of amygdala responses to angry schematic faces in social anxiety disorder*, *Vol. 25*, 496–505 6.
First, M. B., et al. (2002). *Structured clinical Interview for DSM-IV-TR Axis I disorders.* New York: Biometrics Research, New York State Psychaitric Institute Research Version, Patient Edition (SCID-I/P).
Foa, E. B., et al. (1996). Cognitive biases in generalized social phobia. *Journal of Abnormal Psychology, 105*(3), 433–439.
Grafton, B., & MacLeod, C. J. C. P. S. (2016). *Engaging with the wrong people: The basis of selective attention to negative faces in social anxiety*, *Vol. 4*, 793–804 5.
Green, D. M., & Swets, J. A. (1966). *Signal detection theory and Psychophysics.* New York: Wiley.
Klumpp, H., et al. (2015). *Prefrontal control and predictors of cognitive behavioral therapy response in social anxiety disorder*, *Vol. 11*, 630–640 4.
Lakens, D. (2013). *Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs*, *Vol. 4*, 1–12.
Liang, C.-W., et al. (2017). Sustained visual attention for competing emotional stimuli in social anxiety: An eye tracking study. *Journal of Behavior Therapy and Experimental Psychiatry, 54*, 178–185.
Lynn, S. K., & Barrett, L. F. (2014). Utilizing" signal detection theory. *Psychological Science, 25*(9), 1663–1673.
Lynn, S. K., et al. (2014). Gender differences in oxytocin-associated disruption of decision bias during emotion perception. *Psychiatry Research, 219*(1), 198–203.
Lynn, S. K., et al. (2016). Working memory capacity is associated with optimal adaptation of response bias to perceptual sensitivity in emotion perception. *Emotion, 16*(2), 155–163.
Lynn, S. K., et al. (2018). Associations between feelings of social anxiety and emotion perception. *Journal of Behavior Therapy and Experimental Psychiatry, 59*(Supplement C), 40–47.
Lynn, S. K., et al. (2012). Affective state influences perception by affecting decision parameters underlying bias and sensitivity. *Emotion, 12*(4), 726.
Machado-de-Sousa, J. P., et al. (2010). Facial affect processing in social anxiety: Tasks and stimuli. *Journal of Neuroscience Methods, 193*(1), 1–6.
Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.
MacNamara, A., et al. (2017). *Transdiagnostic neural correlates of affective face processing in anxiety and depression*, *Vol. 34*, 621–631 7.

Maddox, W. T., & Bohil, C. (2005). Optimal classifier feedback improves cost-benefit but not base-rate decision criterion learning in perceptual categorization. *Memory & Cognition, 33*(2), 303–319.

McTeague, L. M., et al. (2018). Face perception in social anxiety: Visuocortical dynamics reveal propensities for hypervigilance or avoidance. *Biological Psychiatry, 83*(7), 618–628.

Mogg, K., et al. (2016). *Anxiety and attention to threat: Cognitive mechanisms and treatment with attention bias modification*, *Vol. 87*, 76–108.

Rapee, R. M., et al. (1997). *A cognitive-behavioral model of anxiety in social phobia*, *Vol. 35*, 741–756 8.

Rutter, L. A., et al. (2019). *Emotion sensitivity and self-reported symptoms of generalized anxiety disorder across the lifespan: A population-based sample approach.* e01282.

Singh, J. S., et al. (2015). The effects of social anxiety and state anxiety on visual attention: Testing the vigilance–avoidance hypothesis. *Cognitive Behaviour Therapy, 44*(5), 377–388.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149.

Staugaard, S. R. (2010). Threatening faces and social anxiety: A literature review. *Clinical Psychology Review, 30*(6), 669–690.

Stopa, L., & Clark, D. M. (2000). Social phobia and interpretation of social events. *Behaviour Research and Therapy, 38*(3), 273–283.

Uren, T. H., et al. (2004). Probability and cost estimates for social and physical outcomes in Social Phobia and Panic Disorder. *Journal of Anxiety Disorders, 18*(4), 481–498.

Voncken, M. J., et al. (2003). Interpretation and judgmental biases in social phobia. *Behaviour Research and Therapy, 41*(12), 1481–1488.

Waters, A. M., et al. (2015). *Look for good and never give up: A novel attention training treatment for childhood anxiety disorders*, *Vol. 73*, 111–123.

Wong, Q. J., & Rapee, R. M. (2016). The aetiology and maintenance of social anxiety disorder: A synthesis of complementary theoretical models and formulation of a new integrated model. *Journal of Affective Disorders, 203*, 84–100.