# Social avoidance behavior modulates motivational responses to social reward-threat conflict signals: A preliminary fMRI study

Travis C. Evans[1,2,3] · Michael Esterman[1,2,4] · Jennifer C. Britton[3]

## Abstract

Social avoidance behavior (SAB) produces impairment in multiple domains and contributes to the development and maintenance of several psychiatric disorders. Social behaviors such as SAB are influenced by approach-avoidance (AA) motivational responses to affective facial expressions. Notably, affective facial expressions communicate varying degrees of social reward signals (happiness), social threat signals (anger), or social reward-threat conflict signals (co-occurring happiness and anger). SAB is associated with dysregulated modulation of automatic approach-avoidance (AA) motivational responses exclusively to social reward-threat conflict signals. However, no neuroimaging research has characterized SAB-related modulation of automatic and subjective AA motivational responses to social reward-threat conflict signals. We recruited 30 adults reporting clinical, moderate, or minimal SAB based on questionnaire cutoff scores. SAB groups were matched on age range and gender. During fMRI scanning, participants completed implicit and subjective approach-avoidance tasks (AATs), which involved more incidental or more explicit evaluation of facial expressions that parametrically varied in social reward signals (e.g., $50\%_{\text{Happy}}$), social threat signals (e.g., $50\%_{\text{Angry}}$), or social reward-threat conflict signals (e.g., $50\%_{\text{Happy}} + 50\%_{\text{Angry}}$). In the implicit AAT, SAB was associated with slower automatic avoidance actions and weaker amygdala-pgACC connectivity exclusively as a function of social reward-threat conflict signals. In the subjective AAT, SAB was associated with smaller increases in approach ratings, smaller decreases in avoidance ratings, and weaker dlPFC-pgACC connectivity exclusively in response to social reward-threat conflict signals. Thus, SAB is associated with dysregulated modulation of automatic and subjective AA motivational sensitivity to social reward-threat conflict signals, which may be facilitated by overlapping neural systems.

**Keywords** Social · Approach · Avoidance · fMRI · Reward · Threat

## Introduction

Social avoidance behavior (SAB) disrupts both the formation and maintenance of social relationships, which plays an important role in the development of mood, anxiety, and psychotic disorders (Dunbar & Shultz, 2007; Keltner & Kring, 1998; Umberson & Montez, 2010). Broadly, SAB

consists of withdrawing during social interactions, prematurely terminating social interactions, and/or avoiding social interactions entirely (Blalock & Joiner, 2000). From a functional perspective, SAB reduces negative affect associated with potential or anticipated social exclusion (Cacioppo & Cacioppo, 2014; Cacioppo & Hawkley, 2009; Kupferberg et al., 2016). However, SAB also produces and/or exacerbates social isolation, which erodes social relationships and maintains chronic distress (Hawkley et al., 2007; Masi et al., 2011). As a result, SAB putatively serves as a transdiagnostic risk factor for the development and maintenance of multiple psychiatric disorders (for reviews, see Porcelli et al., 2019; Cotter et al., 2018). For example, patients with Major Depressive Disorder (MDD) or Social Anxiety Disorder (SAD) exhibit distinct clinical profiles, but demonstrate equivalent levels of SAB (Ottenbreit et al., 2014). Even within the same disorder, such as SAD, patients exhibit varying degrees of SAB ranging from prototypical social

✉ Travis C. Evans
t.evans1@umiami.edu

Jennifer C. Britton
j.britton@miami.edu

[1] VA Boston Healthcare System, Boston, MA, USA

[2] Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA

[3] Department of Psychology, University of Miami, Coral Gables, FL, USA

[4] National Center for PTSD, Boston, MA, USA

avoidance to atypical risky social approach (Kashdan & Hofmann, 2008). Therefore, it is important to characterize mechanisms that contribute to SAB specifically, rather than psychopathology more generally.

Social behaviors, such as SAB, are guided in part by approach-avoidance (AA) motivational responses to affective facial expressions (Ambadar et al., 2005; Barrett et al., 2019; Frith, 2009; Rilling & Sanfey, 2011; Strack & Deutsch, 2004). For example, happy facial expressions frequently serve as social reward signals that communicate an opportunity for social affiliation. As a result, happy facial expressions typically activate approach motivational responses (Stins et al., 2011). In contrast, angry facial expressions frequently serve as social threat signals that communicate an opportunity for social exclusion. As such, angry facial expressions typically activate avoidance motivational responses (Marsh et al., 2005; Vrana & Gross, 2004). However, it is important to note that social reward signals or social threat signals conveyed by affective facial expressions are not always perceived in this manner. For example, happy facial expressions may activate avoidance motivational responses if perceived as mocking and/or an opportunity for social exclusion (Cacioppo & Hawkley, 2009). Additionally, angry facial expressions may activate approach motivational responses if perceived as an opportunity to establish social dominance over another individual (Krieglmeyer & Deutsch, 2013). Thus, affective facial expressions may activate differing motivational responses across individuals and/or environmental contexts.

To measure AA motivational responses to affective facial expressions, previous research has employed various versions of the Approach-Avoidance Task (Heuer et al., 2007; Rinck & Becker, 2007). In the AAT, participants make behavioral responses (e.g., pushing or pulling a joystick) that increases or decreases stimulus size to simulate approach and avoidance in response to a stimulus (van Peer et al., 2010). Previous research has utilized both implicit and explicit versions of the AAT paradigm to measure AA motivational responses to affective facial expression (Roelofs et al., 2009). In the implicit AAT, participants are instructed to make AA motivational responses based on a contingency that is *independent* of the emotion conveyed by a facial expression (e.g., male face = pull; female face = push). In the explicit AAT, participants are instructed to make AA motivational responses based on a contingency that is *dependent* on the emotion conveyed by a facial expression (e.g., happy face = pull; angry face = push). In this manner, AA motivational responses can be directly compared when facial affect is incidentally or explicitly evaluated. However, it should be noted that AAT paradigms cannot fully disentangle the contribution of Pavlovian, habitual, and instrumental processes (Huys et al., 2011). Thus, rather than assessing dualistic motivational systems, implicit and explicit AAT paradigms

may assess relatively more automatic or relatively more controlled AA motivational responses, respectively (Rotteveel et al., 2015; Rotteveel & Phaf, 2004).

In AAT paradigms, the strength of AA motivational responses is typically quantified as the reaction time (RT) required to "approach" or "avoid" affective facial expressions. Specifically, RTs are compared between AA motivational responses that are congruent (e.g., happy = approach) or incongruent (e.g., happy = avoid) with the emotion conveyed by an affective facial expression (Roelofs et al., 2005; Rotteveel & Phaf, 2004). In explicit AAT paradigms, affective facial expressions reliably elicit slower RTs during emotion incongruent compared to emotion congruent conditions (Roelofs et al., 2005; Rotteveel et al., 2015). In implicit AAT paradigms, however, affective facial expressions less reliably elicit differences in RTs between emotion incongruent and emotion congruent conditions (Roelofs et al., 2009; Rotteveel & Phaf, 2004). Mirroring these behavioral effects, multiple neuroimaging studies using explicit AAT paradigms demonstrate that emotion incongruent trials recruit greater activation within anterior prefrontal cortex (aPFC) regions, such as the ventrolateral prefrontal cortex (vlPFC), compared with emotion congruent trials (Bramson et al., 2018; Kaldewaij et al., 2017; Kaldewaij et al., 2021; Roelofs et al., 2009). Moreover, both neuroimaging and neuromodulation evidence suggests that these aPFC regions exert top-down control over emotion-relevant processing within the amygdala (Bramson, den Ouden, et al., 2020a; Bramson, Folloni, et al., 2020b; Volman, Toni, et al., 2011b). Together, these results suggest that individuals exercise cognitive control over more automatic AA motivational responses elicited by affective facial expressions when necessary to maintain goal-directed behavior (Koch et al., 2018).

It is important to note, however, that facial expressions rarely communicate "pure" social reward signals (e.g., $100\%_{Happy}$) or "pure" social threat signals (e.g., $100\%_{Angry}$; Matsumoto & Hwang, 2014; Carrol & Russell, 1997). Instead, facial expressions typically communicate varying degrees of social reward signals (e.g., $50\%_{Happy}$), social threat signals (e.g., $50\%_{Angry}$), or *co-occurring* signals of social reward and social threat (e.g., $50\%_{Happy} + 50\%_{Angry}$; Matsumoto & Hwang, 2014; Barrett et al., 2019; Carrol & Russell, 1997; Beaver et al., 2008). Consistent with a greater degree of ecological validity, these types of ambiguous facial expressions elicit more pronounced individual differences in perceptual processes relative to unambiguous, "pure" facial expressions (Staugaard, 2010). Notably, individual differences are particularly pronounced when social reward signals and social threat signals simultaneously co-occur to generate social reward-threat conflict signals, which activates competing motivations to approach *and* avoid (Evans & Britton, 2020; Gutierrez-Garcia & Calvo, 2014; Gutiérrez-García & Calvo, 2016). To prevent behavioral inaction during these types of approach-avoidance conflicts, AA motivational responses

must be flexibly modulated to effectively guide social behavior (Fishbach & Shah, 2006; Krieglmeyer et al., 2013; Strack & Deutsch, 2004). Thus, maladaptive social behaviors such as SAB may be associated with the degree to which individuals modulate AA motivational responses as a function of varying social signals conveyed by ambiguous facial expressions.

Consistent with this conceptualization, previous research using implicit AAT paradigms demonstrate that SAB is selectively associated with modulation of automatic AA motivational responses to varying degrees of social reward-threat conflict (Evans & Britton, 2020). In this study, SAB was characterized by a U-shaped pattern of modulation in which automatic avoidance actions were comparatively faster to social reward-threat conflict signals relative to unambiguous social reward signals and unambiguous social threat signals (e.g., $50\%_{Happy}$ *and* $50\%_{Angry}$ < $100\%_{Happy}$ *or* $100\%_{Angry}$). In contrast, SAB was not associated with modulation of automatic approach actions as a function of varying social reward-threat conflict. Moreover, SAB did not modulate automatic approach or avoidance actions as a function of varying degrees of social reward signals or social threat signals. Thus, within implicit AAT paradigms that assess more automatic AA motivational responses, previous research suggest that SAB is selectively associated with dysregulated motivational responses as a function of social reward-threat conflict signals. However, this previous study did not examine SAB-related modulation of more controlled AA motivational responses. Therefore, it remains unclear if SAB is associated with dysregulated modulation of both automatic and controlled AA motivational responses as a function of social reward-threat conflict signals.

However, assessing controlled AA motivational responses to ambiguous facial expressions with traditional explicit AAT paradigms poses challenges to categorizing emotion incongruent and emotion congruent conditions. In previous research using unambiguous facial expressions ($100\%_{Happy}$ or $100\%_{Angry}$), it was possible to unequivocally categorize AA motivational responses as either emotion incongruent (e.g., happy = avoid) or emotion congruent (e.g., happy = approach). However, it is not possible to definitively categorize AA motivational responses to ambiguous facial expressions as emotion incongruent or emotion congruent. For example, social reward-threat conflict facial expressions simultaneously communicate both social reward signals and social threat signals ($50\%_{Happy}$ + $50\%_{Angry}$), which participants perceive as expressing simultaneous happiness and anger (Evans & Britton, 2020). During an emotion congruent condition, some individuals might generate approach motivational responses due to perceiving these faces as predominantly happy, whereas other individuals might generate avoidance motivational responses due to perceiving these faces as predominantly angry. Further complicating this issue, individuals systematically vary in emotion

categorization of ambiguous facial expressions based on factors such as depressive and anxiety symptoms (Gutierrez-Garcia & Calvo, 2014; Gutiérrez-García & Calvo, 2016; Joormann & Gotlib, 2006). Therefore, using a traditional explicit AAT paradigm in conjunction with ambiguous facial expressions may confound individual differences in emotional categorization and AA motivational responses.

One potential way to circumvent this issue is to utilize subjective AAT paradigms to characterize modulation of more controlled AA responses to ambiguous facial expressions. Subjective AAT paradigms measure more controlled AA motivational responses based on self-reported or behavioral AA motivation responses (Aupperle et al., 2015; Aupperle & Paulus, 2010; Evans & Britton, 2020; Schlund et al., 2011; Schlund et al., 2016). Like explicit AAT paradigms, individuals generate AA motivational responses in subjective AAT paradigms based on explicitly evaluating the affective properties of a stimulus. Unlike explicit AAT paradigms, however, subjective AA motivational responses are not associated with visual feedback that simulate approach or avoidance actions (e.g., increasing/decreasing stimulus size). Thus, it is not possible to compare directly the AA motivational responses between subjective AAT paradigms and implicit AAT paradigms. Although direct comparisons are not possible, previous research nevertheless demonstrates unique patterns of individual differences in AA motivational responses measured with implicit and subjective AAT paradigms (Basanovic et al., 2022; Heuer et al., 2007; Lange et al., 2008; Rinck & Becker, 2007). Therefore, by using implicit and subjective AAT paradigms, it may be possible to simultaneously characterize SAB-related modulation of more automatic and more controlled AA motivational responses as a function of social reward-threat conflict signals.

Given that neuromodulation techniques demonstrate promise as an intervention targeting AA motivational responses (Bramson et al., 2018; Bramson, den Ouden, et al., 2020a; Bramson, Folloni, et al., 2020b; Volman, Roelofs, et al., 2011a), it is also important to characterize the neural mechanisms underlying SAB-related modulation of AA motivational responses. At the neural level, multiple neuroimaging studies using explicit AAT paradigms consistently demonstrate that exerting emotional control over AA motivational responses recruits aPFC regions such as the vlPFC and frontal pole to exert top-down control over the amygdala (for a review, see Koch et al., 2018). In contrast, neuroimaging studies using implicit AAT paradigms demonstrate more mixed and inconsistent findings across studies. Although somewhat mixed, automatic approach motivational responses to rewarding stimuli are associated with greater ventral striatum activation, whereas more automatic avoidance motivational responses to threating stimuli are associated with greater amygdala activation and/or greater

ventral striatum activation (Derntl et al., 2011; Gellner et al., 2021; Kaldewaij et al., 2016; Porcelli et al., 2019; Radke et al., 2015; Wiers et al., 2014). In subjective AAT paradigms, self-reported AA motivation and decision-making are associated with diffuse activation across a widely distributed set of regions including the dorsolateral prefrontal cortex (dlPFC), anterior cingulate cortex (ACC), insula, and caudate (Aupperle et al., 2015; Schlund et al., 2016; Zorowitz et al., 2019). However, no research to date has utilized implicit and subjective AAT paradigms to characterize SAB-related modulation of neural activation or neural connectivity as a function of social reward-threat conflict signals.

The primary goals of the current study were to characterize SAB-related modulation of automatic and subjective AA motivational responses to social reward-threat conflict as well as the neural mechanisms underlying SAB-related modulation of these processes. To this end, adults ranging from clinical to minimal levels of SAB completed implicit and subjective AAT paradigms that presented matched ambiguous facial expressions during fMRI scanning. In both the implicit and subjective AAT paradigms, facial expressions parametrically varied in degrees of social reward, social threat, or social reward-threat conflict. In the implicit paradigm, we hypothesized that SAB would be associated with *relatively* faster automatic avoidance actions as a function of social reward-threat conflict (i.e., a U-shaped pattern). At the neural level, we hypothesized that SAB would be associated with greater amygdala and/or ventral striatum activation during automatic avoidance actions (i.e., an inverse U-shaped pattern) as a function of social reward-threat conflict. Based on a preliminary study validating the subjective AAT

paradigm in an unselected sample (Evans & Britton, 2020), we hypothesized that SAB would be associated with weaker approach motivation and/or stronger avoidance motivation as social reward decreased relative to co-occurring social threat (i.e., $100\%_{\text{Happy}} + 0\%_{\text{Angry}} \rightarrow 50\%_{\text{Happy}} + 50\%_{\text{Angry}} \rightarrow 0\%_{\text{Happy}} + 100\%_{\text{Angry}}$). Based on previous fMRI research using explicit motivation paradigms, we hypothesized SAB would be associated with differential patterns of dlPFC, ACC, insula, and/or caudate activation/connectivity, which may vary *linearly* or *nonlinearly* as social reward decreases relative to co-occurring social threat (Schlund et al., 2016). Thus, we hypothesized that SAB would be associated with modulation of neural activation characterized by either: 1) weaker *reward-related* activation/connectivity as social reward decreased relative to co-occurring social threat ($100\%_{\text{Happy}} + 0\%_{\text{Angry}} \rightarrow 50\%_{\text{Happy}} + 50\%_{\text{Angry}} \rightarrow 0\%_{\text{Happy}} + 100\%_{\text{Angry}}$), or 2) weaker *conflict-related* activation/connectivity as a function of social reward-threat conflict ($0\%_{\text{Conflict}} \rightarrow 100\%_{\text{Conflict}} \rightarrow 0\%_{\text{Conflict}}$).

## Methods

### Participants

We strategically recruited a sample of 32 adults to approximate a full distribution of self-reported SAB across the sample (Table 1). To screen participants based on SAB, we used the social avoidance scale of the Liebowitz Social Anxiety Scale (LSAS) without any reference to social anxiety symptoms (i.e., how frequently participants avoided social

**Table 1** Characteristics and comparisons of social avoidance behavior groups

| Measure | Clinical SAB (n = 10) | Moderate SAB (n = 10) | Minimal SAB (n = 10) | Difference |
|---|---|---|---|---|
| Age | 22.44 (3.28) | 20.50 (3.14) | 20.80 (2.35) | $p = 0.32$ |
| Gender | | | | |
| % Female | 60.00% | 50.00% | 50.00% | $p = 0.88$ |
| Racial Identity | | | | |
| % Caucasian | 30.00% | 70.00% | 20.00% | $p = 0.12$ |
| % Asian | 30.00% | 00.00% | 00.00% | |
| % Black | 20.00% | 20.00% | 60.00% | |
| % Other/Multiple | 20.00% | 10.00% | 20.00% | |
| Ethnicity | | | | |
| % Hispanic | 60.00% | 50.00% | 30.00% | $p = 0.54$ |
| CBAS-SAB | 24.10 (9.65) | 15.90 (5.47) | 9.10 (1.66) | $p < 0.001$ |
| LSAS-Avoid | 34.50 (17.78) | 13.80 (8.36) | 4.10 (3.78) | $p < 0.001$ |
| DASS-21 | 19.60 (12.99) | 11.50 (10.10) | 7.70 (8.36) | $p = 0.06$ |

Social avoidance behavior (SAB) groups did not significantly differ on demographic characteristics. SAB groups significantly differed in CBAS-SAB and LSAS-Avoid screening scores. SAB groups differed in DASS-21 internalizing symptoms at trend level

*CBAS-SAB* Cognitive Behavioral Avoidance Scale, Social Behavioral Avoidance sub-scale. *LSAS-Avoid* Liebowitz Social Avoidance Scale, Avoidance sub-scale. *DASS-21* Depression Anxiety and Stress Scales

situations more generally, rather than due to social anxiety symptoms specifically). Using previously validated cutoff scores (Rytwinski et al., 2009), we recruited participants who reported clinical levels of SAB (LSAS-Avoid > 23), moderate levels of SAB (LSAS-Avoid ≤23 & ≥7), or minimal levels of SAB (LSAS-A <7; see Table 1 for SAB group characteristics and comparisons). Importantly, participants in each SAB category were matched on age range (18–30 years old) and gender (5 men and 5 women).

We determined our sample size in part based on previous research characterizing SAB-related modulation of automatic action tendencies using the same implicit AAT paradigm (I-AAT; Evans & Britton, 2020). In this previous study, SAB significantly modulated automatic action tendencies in response to social reward-threat conflict in two relatively small participant samples ($n = 45$ and $n = 58$) with an overall medium-large effect size ($\eta^2 = 0.10$). Based on this effect size, a sample size of 34 participants would be sufficient to detect SAB-related modulation of automatic action tendencies with 80% power. Given that this previous study did not examine SAB-related modulation of subjective motivational responses, we were not able to conduct *a priori* power analyses for the subjective AAT (S-AAT). To address this issue, we conducted an independent replication of SAB-related modulation of subjective AA motivational responses using a modified, online version of the S-AAT (see Supplemental Information).

To be included in the current study, participants were required to report normal color vision and proficiency in English. Participants were excluded from participation based on the following criteria: 1) Significant medical conditions (e.g., cardiovascular disease) or other conditions (e.g., neurological disorder, schizophrenia, brain trauma history, etc.); 2) Prescribed or nonprescribed use of psychotropic medication during the previous 3 months; 3) Clinically significant suicidality or homicidality; 4) Substance disorder in the past 6 months; and 5) Contraindications for MRI scanning.

## Study procedure

All participants provided written informed consent prior to study procedures. All study procedures were conducted in accordance with the local Institutional Review Board. Participants were compensated with either monetary payment and/or course credit.

Following a phone screening session to establish initial eligibility criteria and preliminarily assess SAB, participants completed two separate study visits. In the first study visit, participants completed an assessment battery that included the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998), various self-report questionnaires, the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 2011), the Ishihara Test of Color

Deficiency (Ishihara, 1917), several attention paradigms (e.g., dot-probe task), and a mock MRI scan to acclimate participants to the scanning environment. In the second study visit, participants completed a 1-hour fMRI scanning protocol that included: a resting state scan, implicit Approach-Avoidance Task, an MPRAGE scan, and subjective Approach-Avoidance Task. Following the fMRI session, participants rated the emotion conveyed by facial expressions presented in the AAT paradigms.

## Questionnaires

### Liebowitz social anxiety scale

We used the social avoidance scale of the Liebowitz Social Anxiety Scale (LSAS) to initially screen participants based on SAB (Liebowitz, 1987). The LSAS is most commonly used to measure fear and avoidance of social situations specifically due to social anxiety symptoms. Given our interest in SAB independent of internalizing symptoms, however, we asked participants to rate avoidance of social situations without any reference to social anxiety symptoms. Specifically, individuals reported the frequency to which they avoided 24 different social situations more generally (e.g., meeting strangers; 0 = Never; 3 = Usually). The LSAS avoidance scale ranges from 0 to 72 and demonstrated excellent internal consistency in the current study ($a = 0.94$).

### Cognitive-behavioral avoidance scale

Consistent with our previous research examining SAB-related modulation of motivational responses, we used the social behavioral avoidance subscale from the Cognitive Behavioral Avoidance Scale (CBAS; Ottenbreit & Dobson, 2004). The CBAS is a 31-item questionnaire that assesses 4 distinct types of avoidance. Specifically, the CBAS is comprised of four subscales that assess social behavioral avoidance (e.g., avoid attending social activities), social cognitive avoidance (e.g., avoid thinking about relationship problems), non-social behavioral avoidance (e.g., avoid challenging activities), and non-social cognitive avoidance (e.g., avoid thinking about the future). For all CBAS subscales, items are rated on a 5-point Likert-type scale (1 = *Not at all true for me;* 5 = *Extremely true for me*). In line with our previous research (Evans & Britton, 2020), we utilized the social behavioral avoidance scale as the primary measure of SAB, which demonstrated excellent internal consistency in the current study ($a = 0.93$).

### Depression, anxiety, and stress scale

The Depression, Anxiety, and Stress Scale (DASS-21) assesses internalizing symptoms with subscales measuring

depressive symptoms, anxiety symptoms, and general stress reactivity (Lovibond & Lovibond, 1995). All items in the DASS-21 are measured on a 4-point Likert-type scale (0 = *Did not apply to me at all;* 3 = *Applied to me very much or most of the time*). When summed together as a single total score, DASS-21 scores ranging from 0 to 63. In the current study, the DASS-21 total score demonstrated excellent internal consistency ($a = 0.94$).

## Task paradigms

### Morphed facial expressions

As in our previous work, we used Morpheus software (Broad Institute) to generate three sets of ambiguous facial expressions that conveyed different types of social signals (e.g., social reward) at varying intensities (e.g., 50%). Specifically, we morphed stereotypic facial expressions (i.e., $100\%_{\text{Happy}}$, $100\%_{\text{Angry}}$, and $100\%_{\text{Neutral}}$) to parametrically modulate the type and intensity of social signals. To create varying intensities of social reward signals, for example, we visually morphed $100\%_{\text{Neutral}}$ and $100\%_{\text{Happy}}$ facial expressions to parametrically modulate social reward signal intensity in 25% increments (i.e., $0\%_{\text{Happy}}$, $25\%_{\text{Happy}}$, $50\%_{\text{Happy}}$, $75\%_{\text{Happy}}$, and $100\%_{\text{Happy}}$). In this manner, ambiguous facial expressions parametrically varied in social reward signals (e.g., $50\%_{\text{Happy}}$), social threat signals (e.g., $50\%_{\text{Angry}}$), or social reward-threat conflict signals (e.g., $50\%_{\text{Happy}} + 50\%_{\text{Angry}}$; Fig. 1). In total, we generated morphed facial expressions for six male and six female actors using the NimStim stimulus set (Tottenham et al., 2009), which were subsequently used in the Implicit AAT and Subjective AAT paradigms.

### Implicit approach-avoidance task

To measure automatic AA motivational responses, we utilized an implicit AAT paradigm (Heuer et al., 2007). In the Implicit Approach-Avoidance Task (I-AAT) paradigm, facial expressions are presented on a blue or green background (Fig. 1b). Based on the background color of the facial expressions, participants were instructed to repeatedly press one of two buttons on an MRI-safe controller (e.g., blue background = left button, green background = right button). Background color assignment and button assignment were each counterbalanced across participants. By using a response contingency that is orthogonal to facial expressions (i.e., background color), facial affect is proposed to implicitly influence approach and avoidance response latencies.

On each trial in the I-AAT, participants made five "approach" or five "avoid" button presses. With each approach or avoidance button press, the image either increased (approach) or decreased (avoid) in size by 20% increments until the image disappeared from the screen after five correct button responses (Evans et al., 2021). For each trial, participants were provided with a 2,000 ms response window to make five correct responses. Across three task runs, participants completed a total of 288 trials (144 Approach and 144 Avoid) in which each morphed facial expression was randomly presented 12 times as an approach trial and 12 times as an avoid trial. In addition to approach/avoid trials, 72 null trials (blank screen) also were randomly presented to facilitate modeling the resolution of the hemodynamic response. All trials were separated by presented an average jittered intertrial interval of 500 (range: 250–750) ms.

### Subjective approach-avoidance task

In the Subjective Approach-Avoidance Task (S-AAT), facial expressions are presented in the center of the screen (Fig. 1c). For each trial, participants rate the degree to which they would feel motivated to approach or avoid the facial expression in a social situation. To provide motivation ratings, facial expressions were presented with a 7-point dynamic virtual analogue scale (0 = *Not at all*; 6 = *Extremely*). At the start of each trial, the slider rating was positioned in the center of the scale (i.e., 3 = *Somewhat*). Using the left or right buttons on the fMRI controller, participants decreased or increased the slider value, which dynamically updated with each button press. Upon reaching the desired rating, participants pressed a third button to confirm their rating selection. For each trial, participants were provided with a 4,000-ms response window to select and confirm their rating.

Unlike the I-AAT, participants did not receive visual feedback when rating approach or avoidance motivation in the S-AAT. Given that participants dynamically moved the rating slider between lower and higher ratings, facial expressions would dynamically increase and decrease in size as participants selected among rating options. As a result, larger approach motivation ratings would be confounded with larger amounts of visual information (increasing stimulus size), whereas larger avoidance motivation ratings would be confounded with smaller amounts of visual information (decreasing stimulus size). To prevent a confound between motivational decision-making and visual information, the S-AAT did not provide dynamic visual feedback (i.e., increases or decreases in stimulus size). Given these paradigm-related differences in visual feedback and response system, it is not possible to directly compare SAB-related modulation within the I-AAT and S-AAT.

Across four runs, participants completed a total of 192 trials (96 Approach and 96 Avoid) in which all morphed facial expressions were presented 8 times as an approach trial and 8 times as an avoid trial. Additionally, the paradigm presented 48 null trials (blank screen) to allow periodic resolution of the hemodynamic response. To minimize task switching, participants only completed approach ratings
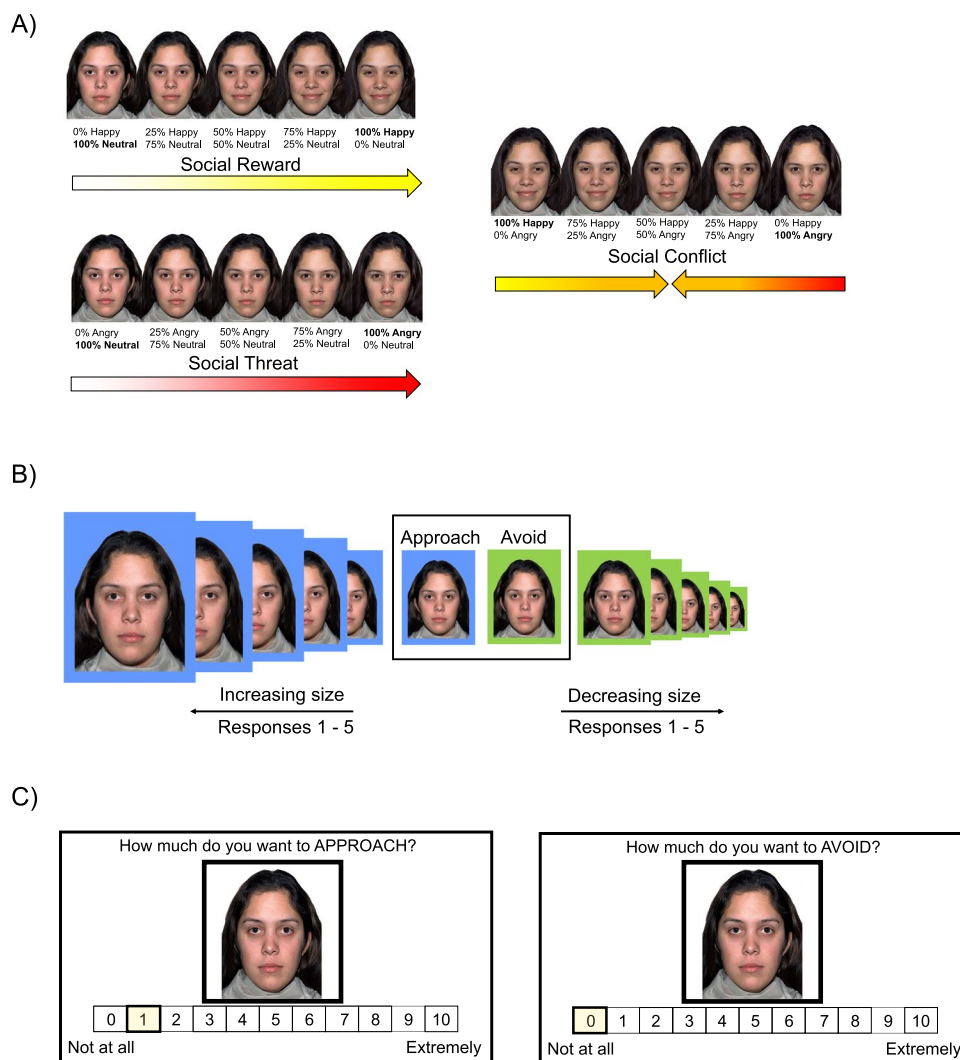
**Fig. 1 Implicit and subjective approach-avoidance task schematics with morphed facial expressions**. **a** Facial expressions were parametrically morphed in 25% increments to vary in social reward ($0\%_{Happy}$, $25\%_{Happy}$, $50\%_{Happy}$, $75\%_{Happy}$, or $100\%_{Happy}$), social threat ($0\%_{Angry}$, $25\%_{Angry}$, $50\%_{Angry}$, $75\%_{Angry}$, or $100\%_{Angry}$), or social reward-threat conflict ($100\%_{Happy}$ + $0\%_{Angry}$, $75\%_{Happy}$ + $25\%_{Angry}$, $50\%_{Happy}$ + $50\%_{Angry}$, $25\%_{Happy}$ + $75\%_{Angry}$, or $0\%_{Happy}$ + $100\%_{Angry}$, $0\%_{Happy}$, $25\%_{Happy}$, $50\%_{Happy}$, $75\%_{Happy}$, or $100\%_{Happy}$). **b** In the Implicit Approach-Avoidance Task (I-AAT), facial expressions appear on a blue or green background. Participants were instructed to press a left or right button based on the background color of the facial expression. With each button response, the size of the image increased in size by 20% (approach trials) or decreased in size by 20% (avoid trials). After making the fifth and final response for a trial, the image disappeared from the screen. **c** In the Subjective Approach-Avoidance Task (S-AAT), facial expressions appear above a visual rating scale ranging from 0 (*Not at All*) to 6 (*Extremely*). Using this scale, participants rate the degree to which they would feel motivated to approach (approach trials) or avoid (avoid trials) the individual displaying the facial expression

(Approach runs) or avoid ratings (Avoid runs) within each task run. All trials were separated by an average jittered inter-trial interval of 500 (range: 250–750) ms.

## fMRI data acquisition

For both tasks, neural data were acquired using the same 3-Tesla General Electric Discovery MR750 scanner with a 32-channel head coil. Blood-oxygenation-level-dependent (BOLD) activation was measured with a series of 47 contiguous 3-mm, interleaved axial slices acquired in a $96 \times 96$ matrix resolution with EPI sequencing (TR = 2,300 ms; TE = 25 ms; FOV = 240 mm, Flip Angle = 50$^{o}$). An MPRAGE, high-resolution, T1-weighted, volumetric scan of the whole brain was acquired between the task paradigm scans for co-registration and normalization of functional data.

## fMRI data processing

### Pre-processing

Prior to analysis, fMRI data were preprocessed using standard procedures with Analysis of Functional NeuroImages (AFNI) software. First, EPI images were slice-time corrected and realigned to the first image of each time-series. Following these steps, the EPI images were co-registered to the anatomical image and subsequently normalized within Talairach space. Next, functional data were smoothed with a 6-mm, full-width-at-half maximum, isotropic, Gaussian filter. Each voxel timeseries was scaled to a mean of 100. Next, motion parameters were examined to identify participants who exhibited excessive head motion during the scan (>3-mm translation or >3° rotation across >30% of TRs).

### Defining regions of interest

For the I-AAT, our *a priori* hypotheses focused on examining differential activation of the amygdala and ventral striatum during the generation of automatic AA motivational responses. To this end, we utilized the Talairach-Daemon atlas to generate anatomically derived Regions of Interest (ROIs) for the amygdala and ventral striatum.

For the S-AAT, our *a priori* hypotheses focused on activation within several distributed neural regions including the bilateral PFC, ACC, insula, and caudate. However, cross-study differences demonstrate a heterogeneous topography of conflict-related neural activation within these distally distributed regions (Aupperle et al., 2015; Roelofs et al., 2009; Rolle et al., 2022; Schlund et al., 2016; Zorowitz et al., 2019). Given this spatial heterogeneity, we used the Talairach-Daemon atlas to define a search territory based on previous studies (Fig. S5). We then examined SAB-related differences in modulation of neural activation using small-volume correction (SVC) across the masked search territory.

### Neural activation

For first-level models, trial onsets were subsequently modelled as 2-second blocks (I-AAT) or 4-second blocks (S-AAT) based on stimulus duration. Next, task regressors were convolved with a gamma variate function to approximate the hemodynamic response. For both tasks, we modelled 24 task regressors (12 [Morphs] × 2 [Approach, Avoid]). For the I-AAT paradigm, we also modelled error responses and RT outliers as a separate error regressor similar to previous research. Six rigid-body motion regressors modelled degrees of translation and rotation. Additionally, we modelled both linear and non-linear, low-frequency,

temporal drift during task runs. Finally, TRs that exceeded framewise displacement of >0.5 mm and the preceding TR were censored due to motion.

For the I-AAT and S-AAT, we examined both task-related and SAB-related modulation of neural activation within the *a priori* search territory comprised of the bilateral PFC, ACC, insula, and caudate. To correct for multiple comparisons across this search territory, we used a combined voxel-wise and cluster threshold approach. To obtain a cluster threshold at α = 0.05, 10,000 Monte Carlo simulations were run using the recently developed nonparametric ClustSim function within AFNI (Cox et al., 2017). Based on a nominal threshold of $p = 0.005$ and the observed smoothness of estimated residuals (I-AAT: ACF parameters = 0.50 4.95 12.16; S-AAT: ACF parameters = 0.50, 4.89, 12.18), a 44-voxel (687.50 mm$^3$) cluster level threshold corrected for multiple comparisons across the masked search territory for both the I-AAT and S-AAT (FWE $p < 0.05$). Peak activation voxel coordinates are reported in LPI (Left, Posterior, Inferior) orientation.

### Exploratory neural connectivity

To model task-related connectivity, we utilized a generalized form of context-dependent psychophysiological interaction analyses (gPPI; McLaren et al., 2012). For gPPI analyses, we computed interaction terms between the time series of each neural seed region and task regressors. To account for desynchronization between TRs and stimulus onsets, we upsampled both the neuronal time series and task regressors. After deconvolving the hemodynamic timeseries to estimate the underlying neural response function, we subsequently convolved the upsampled neuronal response function with the upsampled task regressors. After computing gPPI interaction regressors in this manner, gPPI regressors were downsampled back to the original TR resolution (2.3 seconds). For all gPPI models, we included event-related regressors and the mean seed region timeseries to ensure that differences in connectivity could not be attributed to task-related activation or intrinsic connectivity. Finally, gPPI models utilized the same nuisance regressors (motion parameters and drift parameters) included in activation models.

For these exploratory gPPI analyses, we selected seed regions in a post-hoc manner based on regions exhibiting significant SAB-related differences in task activation. To correct for multiple comparisons across the whole brain search territory, we used the same nonparametric cluster correction approach with a nominal statistical threshold of $p = 0.005$ and the observed smoothness of estimated residuals within each paradigm. Based on this combined threshold, a 119-voxel (1859.38 mm$^3$) or 114-voxel

(1781.25 mm$^3$) cluster level threshold corrected for multiple comparisons across the whole brain search territory for both the I-AAT and S-AAT paradigms (FWE $p <$ 0.05), respectively.

## Data reduction

### Participant exclusions

One participant was excluded due to falling asleep during the scan session and one participant was excluded due to prematurely discontinuing the task/scan session. Following these exclusions, all behavioral analyses were conducted in the same final sample of 30 participants for both the I-AAT and S-AAT. For neural analyses, one participant was additionally excluded due to removal of >30% of TRs due to excessive motion during the Subjective AAT. Following these exclusions, all neural analyses were conducted on a final sample of 30 participants for the I-AAT and 29 subjects for the S-AAT.

### Behavioral data exclusions

For the I-AAT, trials in which participants failed to complete five responses in the correct direction were categorized as errors and subsequently excluded from all analyses. After removing error trials, RTs greater than 2.5 standard deviations from a participant's mean approach RT or avoid RT were classified as outliers and removed. Across the final sample, these additional data cleaning procedures removed 6.74% of trials.

## Data analytic strategy

To ensure that unambiguous and ambiguous stimuli were presented with equal frequency, the same unambiguous stimuli trials (100%$_{Happy}$, 100%$_{Angry}$, 100%$_{Neutral}$) were used as the endpoints of the continuum across the social reward-threat conflict, social reward, and social threat models (Fig. 1a). Thus, motivational responses to unambiguous facial expressions are not statistically independent across models, which precludes a direct comparison between the social reward-threat conflict, social reward, and social threat models.

Additionally, we also characterized task-related effects in the absence of SAB-related modulation. To test task-related effects, we utilized GLMMs to test 2-way interactions using a 2 (Condition: Approach vs. Avoid) × Linear/Non-Linear omnibus model ($a \leq 0.05$). Following significant 2-way interactions within omnibus models ($a \leq 0.05$), we then examined linear/non-linear patterns separately within the Approach condition and Avoid condition. All analyses were conducted by using SPSS software ver. 24.0

(IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM). To compute $p$-values and degrees of freedom for GLMMs, we used restricted maximum likelihood (REML) in conjunction with the Satterthwaite approximation (Luke, 2017).

To test SAB-related modulation, all analyses (behavioral, neural activation, and neural connectivity) utilized GLMMs to test 3-way interactions using a 2 (Condition: Approach vs. Avoid) × SAB × Linear/Non-Linear omnibus model. For all GLMMs, SAB was mean-centered and modeled as a continuous covariate of interest. Additionally, we confirmed that significant SAB-related modulation was independent of internalizing symptoms by including DASS-21 total scores in the first step of omnibus models as a continuous covariate of non-interest. Following significant 3-way interactions within omnibus models ($a \leq 0.05$), we then examined 2-way interactions (SAB × linear/non-linear) within the Approach condition and Avoid condition, separately. As indicated by significant 2-way interactions, we conducted continuous simple slopes analyses within the Approach or Avoid condition, which examined linear/non-linear patterns of modulation at high (+1 SD) and low (−1 SD) levels of SAB. For all significant 3-way interactions, we also provide scatterplots to display the distribution of linear and non-linear polynomial contrasts across participants (see Supplemental Information).

Given our relatively modest sample size, we were not able to conduct maximum model GLMMs that included random slopes and interactions among random effects due to model convergence issues. Although some GLMMs reached model convergence when random slopes for linear/non-linear trends across morphed stimuli were included, other GLMMs did not meet convergence criteria when random slopes were included. To standardize model complexity across behavioral analysis GLMMs, we elected to exclusively model random intercepts to account for individual differences in overall RT (I-AAT) or subjective ratings (S-AAT). For behavioral GLMMs in which model convergence criteria were not successfully satisfied when random intercepts were included, we confirmed primary results after removing the random intercept to eliminate redundancy in the covariance structure. For neural activation and neural connectivity analyses, we modelled both random intercepts and random slopes and report corrected degrees of freedom. When model convergence was not obtained, we removed random slopes and report non-corrected degrees of freedom.

### Implicit approach-avoidance task

Consistent with our previous research (Evans & Britton, 2020), we tested SAB-related differences in modulation of automatic action tendencies using quadratic contrasts (U-shaped) for the social reward-threat conflict model

(0%$_{Conflict}$, 50%$_{Conflict}$, 100%$_{Conflict}$, 50%$_{Conflict}$, 0%$_{Conflict}$) and linear contrasts for the social reward model and social threat model (e.g., 0%$_{Happy}$, 25%$_{Happy}$, 50%$_{Happy}$, 75%$_{Happy}$, 100%$_{Happy}$). To quantify these patterns of SAB-related modulation, we utilized orthogonal polynomial contrasts that tested quadratic and linear trends in automatic action tendencies, respectively. For social reward-threat conflict models, we simultaneously modeled both quadratic and linear trends to ensure that SAB-related differences in quadratic response patterns were independent of linear response patterns.

Consistent with previous research (Buetti et al., 2012; Evans & Britton, 2020; Veenstra et al., 2017), all behavioral analyses were conducted on *initial* RTs from the *first* response of each trial (of the 5 required responses for each trial). Given the non-normal distribution of RTs, we separately performed a natural log transformation on each participant's avoid trial RTs and approach trial RTs (following removal of error and outlier RTs). We utilized log-normal transformed RTs for all I-AAT analyses to approximate assumptions of normality in GLMMs, whereas figures present nontransformed RTs for comparative purposes.

### Subjective approach-avoidance task

Our previous research using the S-AAT suggests that subjective approach and avoidance ratings vary linearly as a function of social reward signals (0%$_{Happy}$ ➜ 100%$_{Happy}$), social threat signals (0%$_{Angry}$ ➜ 100%$_{Angry}$), and social reward-threat conflict signals (100%$_{Happy}$ + 0%$_{Angry}$ ➜ 0%$_{Happy}$ + 100%$_{Angry}$; Evans & Britton, 2020). Thus, for behavioral analyses, we tested SAB-related differences in modulation of subjective approach and avoidance ratings using linear polynomial trends. At the neural level, however, previous research suggests that neural activation patterns may vary linearly or nonlinearly as a function of reward-threat conflict (Schlund et al., 2016). Thus, for neural analyses, we tested SAB-related differences in modulation of neural activation/connectivity using both linear and quadratic polynomial trends.

## Results

### Implicit approach-avoidance task

#### Task-related effects

We did not observe task-related effects on automatic action tendencies as a function of social reward-threat conflict signals, social reward signals, or social threat signals (all $p$s > 0.30; see Supplemental Information). Similarly, we did not observe task-related effects on amygdala or ventral striatum activation as a function of social reward-threat conflict

signals, social reward signals, or social threat signals (all $p$s > 0.18; see Supplemental Information). Finally, no regions survived small-volume correction for task-related effects within the *a priori* search territory as a function of social reward-threat conflict signals, social reward signals, or social threat signals.

### SAB-related modulation

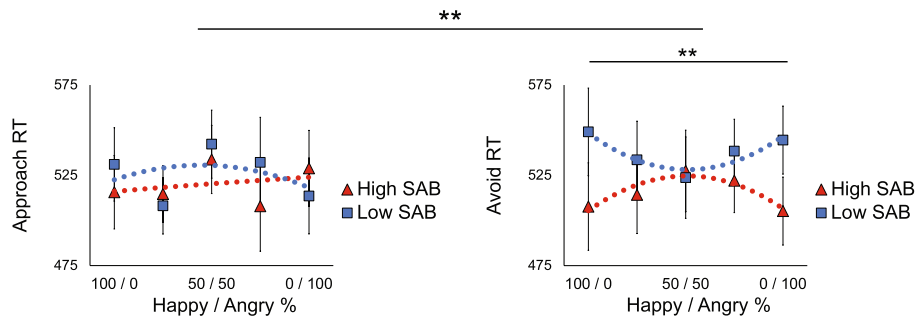### Social reward-threat conflict model

**Behavioral** As hypothesized, we observed SAB-related modulation of automatic action tendencies as a function of social reward-threat conflict, which significantly differed between Approach and Avoid conditions (Condition × SAB × Quadratic: $B = 0.002$, SE $= 0.0007$; $F_{(1,259)} = 6.41$, $p = 0.01$). After controlling for internalizing symptoms, this pattern of SAB-related modulation remained unchanged ($p = 0.01$).

Consistent with our previous research, SAB significantly modulated automatic avoidance actions (SAB × Quadratic: $B = -0.001$, SE $= 0.0005$; $F_{(1,115)} = 7.77$, $p = 0.006$) but not automatic approach actions (SAB × Quadratic: $B = 0.0005$, SE $= 0.0005$; $F_{(1,115)} = 0.88$, $p = 0.35$; Fig. 2). Contrary to our hypotheses and previous findings, higher levels of SAB were characterized by significantly *slower* automatic avoidance actions as a function of social reward-threat conflict, which produced an inverted U-shaped pattern (Quadratic: $B = 0.03$, SE $= 0.01$; $F_{(1,115)} = 6.71$, $p = 0.01$). In contrast, lower levels of SAB were characterized by the opposite pattern (i.e., U-shaped pattern), which was not statistically significant (Quadratic: $B = 0.01$, SE $= 0.005$; $F_{(1,115)} = 3.06$, $p = 0.08$) .

Although SAB-related differences were descriptively largest in response to unambiguous facial expressions (i.e., 100%$_{Happy}$ and 100%$_{Angry}$; Fig. 2), SAB was not significantly associated with automatic approach actions (both $r_s <$ |0.21|, both $p$s > 0.28) or automatic avoidance actions (both $r_s <$ |0.15|, both $p$s > 0.42) to unambiguous facial expressions. Thus, SAB-related modulation of automatic action tendencies was not driven by a particular facial expression (e.g., 100%$_{Happy}$ or 100%$_{Angry}$) but was instead characterized by a quadratic *pattern* of modulation.

*Amygdala activation.* Similar to our behavioral results, we observed SAB-related modulation of amygdala activation that significantly differed between approach and avoid conditions as a function of social reward-threat conflict signals (Left Amygdala: Condition × SAB × Quadratic: $B = -0.002$, SE $= 0.0008$; $F_{(1, 182.39)} = 5.24$, $p = 0.02$; Right Amygdala: Condition × SAB × Quadratic: $B = -0.002$, SE $= 0.001$; $F_{(1, 199.85)} = 5.41$, $p = 0.03$; Fig. 3). After controlling for internalizing symptoms (DASS-21), SAB-related
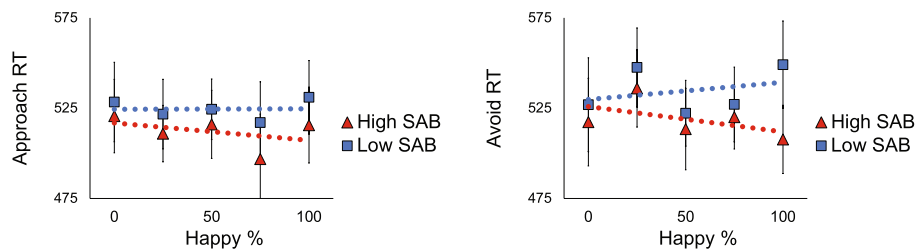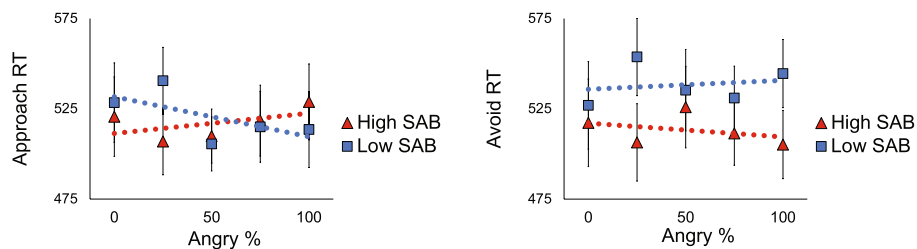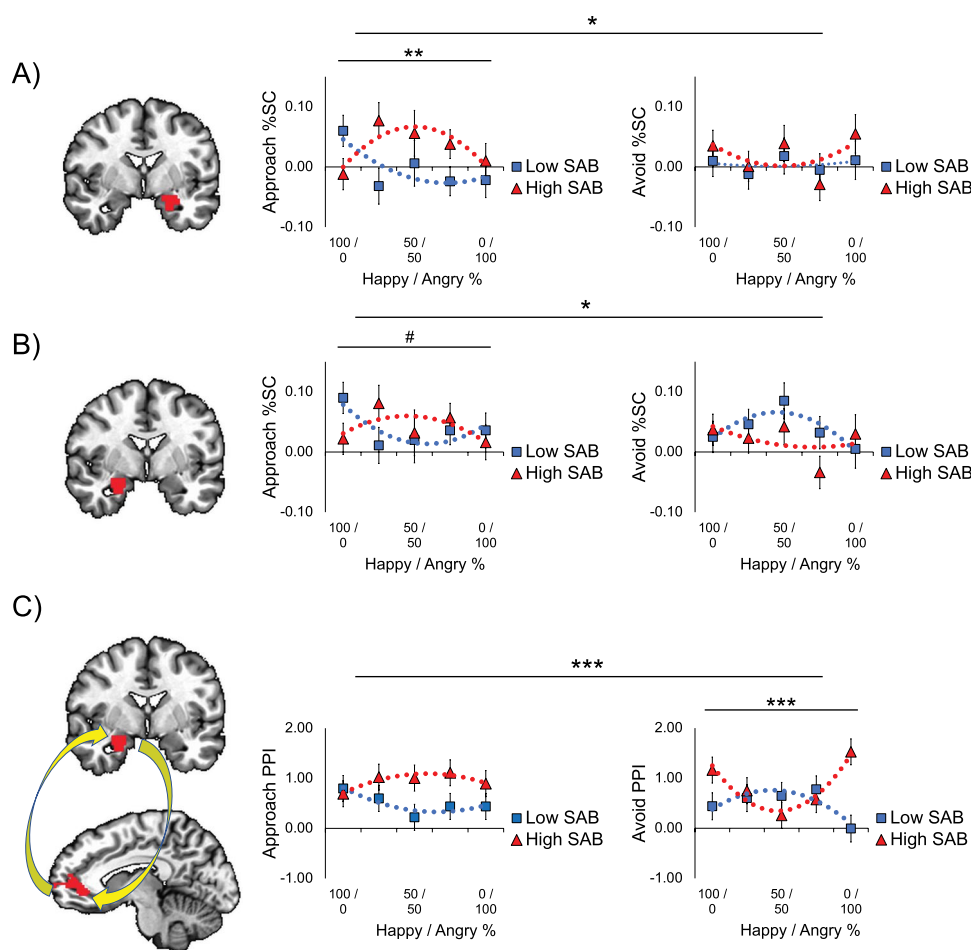
**Fig. 2 Social avoidance behavior modulates automatic motivational responses to social reward-threat conflict signals**. Social Avoidance Behavior (SAB)-related modulation of automatic approach actions (left column) and automatic avoidance actions (right column). Based on continuous simple slope effects, behavioral effects are depicted at high levels of SAB (+1SD; red triangles and red dotted lines) and low levels of SAB (−1SD; blue squares and blue dotted lines). As a function of varying degrees of social reward-threat conflict relative to unambiguous social reward or social threat (*top row*), SAB was not associated with differences in automatic approach actions (**a**), but was associated with slower automatic avoidance actions to social reward-threat conflict (**b**). As a function of social reward (*Middle Row*) or social threat (*Bottom Row*), SAB did not modulate automatic approach actions or automatic avoidance actions (C, D, E, & F). Note: **$p \leq 0.01$

modulation remained significant for the left and right amygdala ROIs (both $ps < 0.05$).

Contrary to our hypotheses, however, SAB did not significantly modulate amygdala activation during automatic *avoidance* actions (Left Amygdala: SAB × Quadratic: $B = 0.0004$, SE $= 0.0006$; $F_{(1, 115)} = 0.62$, $p = 0.43$; Right Amygdala: SAB × Quadratic: $B = 0.001$, SE $= 0.0007$; $F_{(1, 115)} = 1.92$, $p = 0.17$). Instead, SAB primarily modulated amygdala activation during automatic *approach* actions (Left Amygdala: SAB × Quadratic: $-B = 0.001$, SE $= 0.0006$; $F_{(1, 115)}$

$= 6.32$, $p = 0.01$; Right Amygdala: SAB × Quadratic: $B = -0.001$, SE $= 0.0007$; $F_{(1, 115)} = 2.94$, $p = 0.09$). During automatic *approach* actions, higher levels of SAB were characterized by stronger amygdala activation as a function of social reward-threat conflict signals, which produced an inverted U-shaped pattern (Left Amygdala: Quadratic: $B = 0.03$, SE $= 0.01$; $F_{(1, 115)} = 4.44$, $p = 0.04$; Right Amygdala: Quadratic: $B = 0.03$, SE $= 0.02$; $F_{(1, 115)} = 2.89$, $p = 0.09$). In contrast, lower levels of SAB were not associated with modulation of amygdala activation as a function of social

**Fig. 3 Social avoidance behavior modulates amygdala activation and connectivity during automatic motivational responses to social reward-threat conflict signals**. Social Avoidance Behavior (SAB)-related modulation of amygdala activation and amygdala connectivity during automatic approach actions (left column) and automatic avoidance actions (right column). Neural regions are depicted in radiological convention (left = right). Based on continuous simple slope effects, neural activation/connectivity effects are depicted at high levels of SAB (+1SD; red triangles and red dotted lines) or low levels of SAB (−1SD; blue squares and blue dotted lines). (**a**; *top row*) SAB was associated with greater left amygdala activation to social reward-threat conflict facial expressions relative to unambiguous social reward or social threat facial expressions during automatic approach actions (left column), but not during automatic avoidance actions (right column). (**b**; *middle row*) SAB was associated with greater right amygdala activation to social reward-threat conflict facial expressions relative to unambiguous social reward or social threat facial expressions during automatic approach actions (left column), but not during automatic avoidance actions (right column). (**c**; *bottom row*) SAB was not associated with differences in amygdala-ACC connectivity during automatic approach actions (left column), but was associated with weaker amygdala-ACC connectivity during automatic avoidance actions to social reward-threat conflict facial expressions relative to unambiguous social reward or social threat facial expressions (right column). Note: ***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$; #$p < 0.10$

reward-threat conflict signals (Left Amygdala: Quadratic: $B = -0.01$, SE = 0.01; $F_{(1, 115)} = 1.11$, $p = 0.30$; Right Amygdala: Quadratic: $B = 0.01$, SE = 0.008; $F_{(1, 115)} = 1.73$, $p = 0.19$).

*Ventral striatum activation*. We did not observe SAB-related modulation of ventral striatum activation that differed between Approach and Avoidance conditions as a function of social reward-threat conflict signals all $ps > 0.26$, see Supplemental Information).

*Small volume activation*. No regions exhibited SAB-related modulation as a function of social reward-threat conflict signals that survived small-volume correction within the *a priori* search territory.

*Exploratory neural connectivity*. Given that SAB-related modulation of neural activation was exclusive to the amygdala, we elected to model amygdala ROIs as seed regions for exploratory gPPI analyses.

For the right amygdala seed, we observed SAB-related modulation of connectivity that survived whole-brain correction with a cluster centered on the pregenual ACC (pgACC; [6, −39, 1], $k = 163$]; Fig. 3), which significantly differed between approach and avoid conditions (Condition

× SAB × Quadratic: $B = -0.03$, SE = 0.007; $F_{(1, 218.76)}$ = 20.25, $p < 0.001$, uncorrected; FWE $p < 0.05$). After controlling for internalizing symptoms (DASS-21), SAB-related modulation of right amygdala-pgACC connectivity continued to survive whole-brain correction.

During automatic *approach* actions, SAB did not modulate right amygdala-pgACC connectivity as a function of social reward-threat conflict signals (SAB × Quadratic: $B = -0.007$, SE = 0.004; $F_{(1, 105.76)} = 1.55$, $p = 0.22$, uncorrected). During automatic *avoidance* actions, however, SAB significantly modulated right amygdala-pgACC connectivity as a function of social reward-threat conflict signals (SAB × Quadratic: $B = 0.03$, SE = 0.006; $F_{(1, 34.37)} = 10.84$, $p = 0.002$, uncorrected). Specifically, higher levels of SAB were characterized by significantly *lower* amygdala-pgACC connectivity as a function of social reward-threat conflict signals, which produced a U-shaped pattern across morphed stimuli (Quadratic: $B = -0.57$, SE = 0.20; $F_{(1, 34.37)} = 7.57$, $p = 0.009$, uncorrected). In contrast, lower levels of SAB were not associated with amygdala-pgACC connectivity as a function of social reward-threat conflict signals (Quadratic: $B = -0.12$, SE = 0.09; $F_{(1, 34.37)} = 1.84$, $p = 0.18$).

For the left amygdala seed, we did not observe SAB-related differences in task-related connectivity that survived whole-brain correction.

**Social reward and social threat models** Consistent with previous findings, we did not observe SAB-related modulation of automatic action tendencies as a function of either the social reward or social threat models (both $ps > 0.25$; see Supplemental Information). Similarly, SAB did not modulate amygdala activation or ventral striatum activation as a function of varying social reward signals or social threat signals (all $ps > 0.71$; see Supplemental Information). Additionally, no regions exhibited SAB-related modulation that survived small-volume correction within the *a priori* search territory. Finally, no regions survived whole-brain correction for SAB-related modulation of left or right amygdala connectivity as a function of varying social reward signals or social threat signals.

## Subjective approach-avoidance task

### Task-related effects

#### Social reward-threat conflict model

**Behavioral** Subjective motivation ratings significantly differed between Approach and Avoid conditions as a function of social reward-threat conflict signals (Condition × Linear: $B = -1.95$, SE = 0.09; $F_{(1, 296)} = 523.63$, $p < 0.001$). As social reward signals decreased relative to co-occurring social threat signals, approach ratings significantly *decreased* (Linear: $B = -0.97$, SE = 0.05; $F_{(1, 119)} = 459.16$, $p < 0.001$), whereas subjective avoidance ratings significantly *increased* (Linear: $B = 0.98$, SE = 0.05; $F_{(1, 119)} = 407.92$, $p < 0.001$).

*Small volume activation.* No clusters survived small-volume correction for task-related neural activation across morphed stimuli between the approach and avoidance conditions (i.e., no Condition × Linear/Quadratic interaction). However, five clusters survived multiple comparison correction for task-related effects on neural activation that differed as a function of approach and avoidance ratings more generally (i.e., a main effect of Condition; Fig. S7). Specifically, we observed greater activation during approach ratings compared with avoidance ratings within a posterior right dlPFC cluster [$k = 256$; 21, 9, 46], posterior left dlPFC cluster [$k = 60$; −39, 11, 36], anterior left dlPFC cluster [$k = 657$; −11, 9, 46], and right insula cluster [$k = 56$; 39, 4, 1]. Additionally, we observed greater deactivation during avoidance ratings compared to approach ratings within a bilateral ACC cluster [$k = 269$; 14, 39, 29].

### Social reward and social threat models

**Behavioral** Subjective motivation ratings significantly differed between Approach and Avoid conditions as a function of increasing social reward signals (Condition × Linear: B = 1.01, SE = 0.08; $F_{(1, 267)} = 159.40$, p < 0.001) and increasing social threat signals (Condition × Linear: B = −0.85, SE = 0.07; $F_{(1, 267)} = 158.06$, p < 0.001).

*Small volume activation.* Similar to the social reward-threat conflict model, no clusters survived small-volume correction for task-related effects on neural activation across morphed stimuli between the approach and avoidance conditions (i.e., no Condition × Linear/Quadratic interaction). For both the social reward model and social threat model, task-related effects on neural activation differed more generally as a function of approach and avoidance ratings (i.e., a main effect of Condition; Figs. S9 and S10). In both the social reward and social threat models, we observed dlPFC and ACC clusters that overlapped with regions observed in the social reward-threat model and exhibited similar patterns of task-related activation (see Supplemental Information). Additionally, we also observed caudate and vlPFC clusters that did not overlap with regions observed in the social reward-threat conflict model (see Supplemental Information).
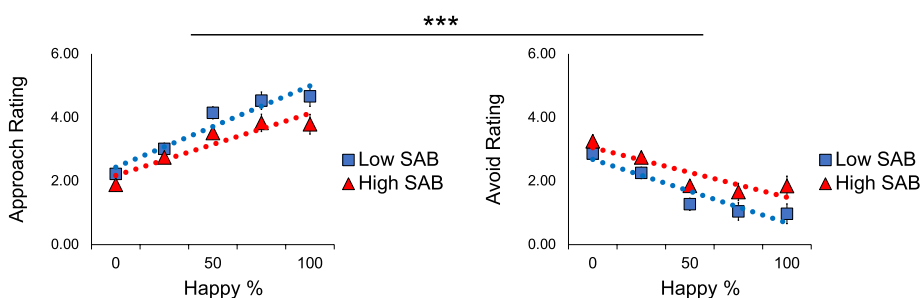
### SAB-related modulation

#### Social reward-threat conflict model

**Behavioral** As hypothesized, we observed SAB-related modulation of subjective motivation ratings as a function
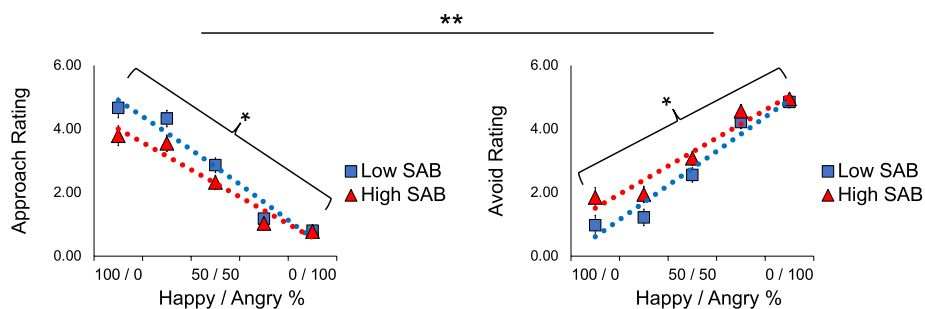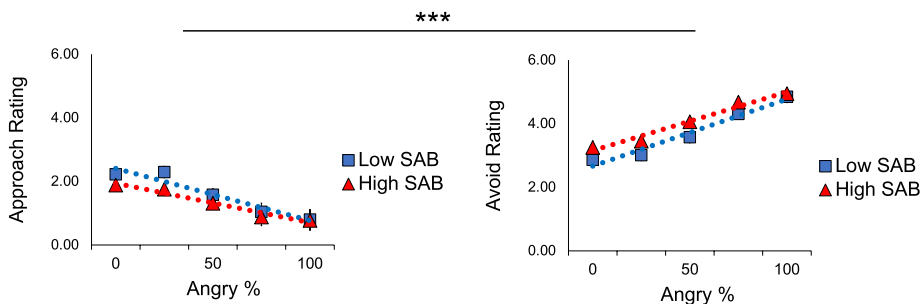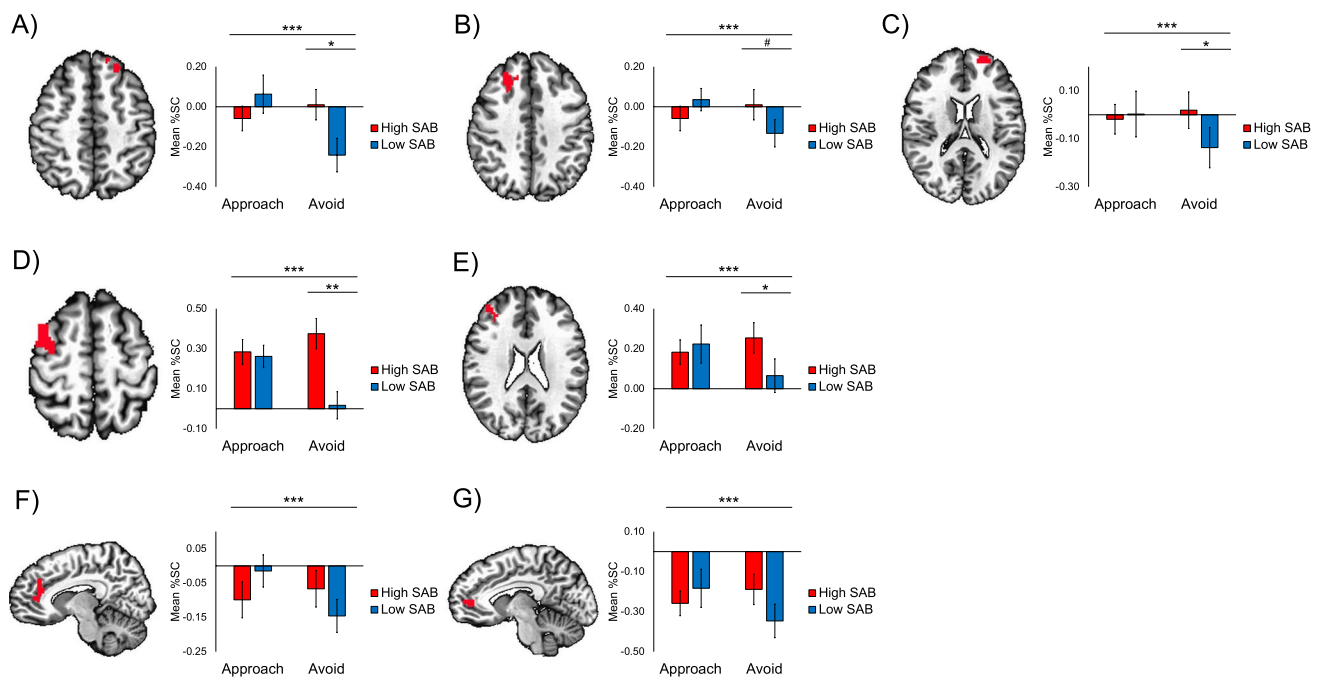
**Fig. 4 Social avoidance behavior modulates subjective motivational responses to social reward-threat conflict signals.** Social Avoidance Behavior (SAB)-related modulation of subjective approach motivation (left column) and subjective avoidance motivation (right column). Based on continuous simple slope effects, behavioral effects are depicted at high levels of SAB (+1SD; red triangles and red dotted lines) or low levels of SAB (−1SD; blue squares and blue dotted lines). (**a**; *top row*) SAB was associated with weaker linear increases in approach motivation (left column) and weaker linear decreases in avoidance motivation (right column) as social reward increased relative to co-occurring social threat. (**b**; *middle row*) SAB was associated with generally weaker approach motivation (left column) and stronger avoidance motivation (right column), which did not vary as a function of social reward. (**c**; *bottom row*) SAB was associated with generally weaker approach motivation (left column) and stronger avoidance motivation (right column), which did not vary as a function of social threat. Note: ***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$

of social reward-threat conflict, which significantly differed between approach and avoid conditions (Condition × SAB × Linear: $B = 0.02$, SE = 0.009; $F_{(1, 292)} = 6.47$, $p = 0.01$; Fig. 4). After controlling for internalizing symptoms (DASS-21), this pattern of SAB-related modulation remained unchanged ($p = 0.01$).

For subjective approach ratings, greater SAB was characterized by a smaller linear *increase* in approach motivation as social reward signals increased relative to co-occurring social threat signals (SAB × Linear: $B = 0.01$, SE = 0.005; $F_{(1, 118)} = 5.77$, $p = 0.02$). For subjective avoid ratings, greater SAB was characterized by a smaller linear *decrease* in avoidance motivation as social reward signals increased

**Fig. 5 Social avoidance behavior modulates neural activation during subjective motivational responses to social reward-threat conflict signals.** Social Avoidance Behavior (SAB)-related modulation of neural activation during subjective approach motivation ratings and avoidance motivation ratings. Neural regions are depicted in radiological convention (left = right). Based on continuous simple slope effects, motivation-related differences in neural activation are depicted at high levels of SAB (+1SD; red bars) or low levels of SAB (−1SD; blue bars). (*Top row*) Within the a) left anterior dorsolateral prefrontal cortex (dlPFC), b) right anterior dlPFC, and c) left vent-

rolateral prefrontal cortex (vlPFC), SAB was associated with greater deactivation during avoidance motivation ratings, but not approach motivation ratings. (*Middle row*) Within the d) right posterior dlPFC and e) right vlPFC, SAB was associated with greater activation during avoidance motivation ratings, but not approach motivation ratings. (*Bottom row*) Within the f) right pregenual anterior cingulate cortex (pgACC) and g) left pgACC, SAB was associated with lower deactivation during avoidance motivation ratings and greater deactivation during approach motivation ratings. Note: ***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$; #$p < 0.10$

relative to co-occurring social threat signals (SAB × Linear: $B = -0.01$, SE = 0.006; $F_{(1, 118)} = 3.96$, $p = 0.049$). Moreover, we replicated this pattern of SAB-related modulation in a larger independent sample with GLMMs that modelled both random intercepts and random slopes (see Supplemental Information)

Although SAB-related differences were descriptively largest in response to unambiguous social reward signals (100%$_{Happy}$), SAB was not significantly associated with approach ratings (both $r_s < |0.30|$, both $ps > 0.11$) or avoidance ratings (both $r_s < |0.32|$, both $ps > 0.08$) to unambiguous social reward. Thus, SAB-related modulation of subjective motivational responses was not driven by a particular intensity of social reward signal (e.g., 100%$_{Happy}$) but was instead characterized by a linear *pattern* of modulation as a function of social reward signals that co-occurred with social threat signals.

*Small volume activation.* Contrary to our hypotheses, no clusters survived small-volume correction for SAB-related modulation of neural activation across morphed stimuli between the approach and avoidance conditions (i.e., no Condition × SAB × Linear/Quadratic interaction). Instead,

seven clusters survived multiple comparison correction for SAB-related modulation of neural activation that differed as a function of approach and avoidance ratings regardless of morphed stimuli (i.e., Condition × SAB interaction; see Fig. 5). Importantly, these clusters continued to survive small-volume correction after controlling for internalizing symptoms (DASS-21).

Specifically, SAB was associated with smaller deactivation during avoidance ratings, but not approach ratings, within a left dlPFC cluster [k = 64; 19, −36, 36], anterior right dlPFC cluster [k = 182; -24, -31, 41], and left vlPFC cluster [k = 47; 19, −56, 16]. Additionally, SAB was associated with greater activation during avoidance ratings, but not approach ratings, within a posterior right dlPFC cluster [k = 193; −36, 4, 61] and right vlPFC cluster [k = 67; −39, -44, 19]. Finally, SAB was associated with smaller *de*activation during avoidance ratings, but greater *de*activation during approach ratings, within a right pgACC cluster [k = 87; −9, −34, 19] and left pgACC cluster [k = 52; 6, −49, 9].

To confirm that SAB modulated neural regions that were also task-relevant, we conducted a series of post-hoc analyses. Specifically, we examined SAB-related modulation
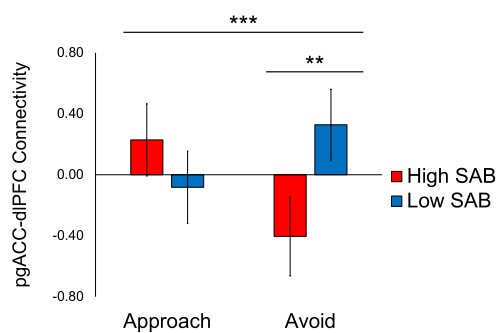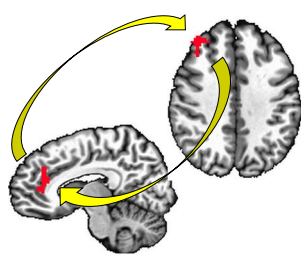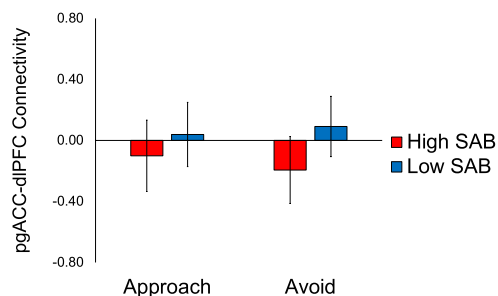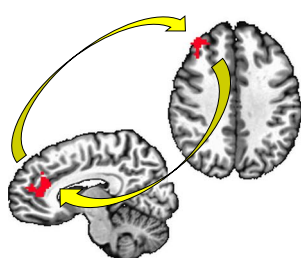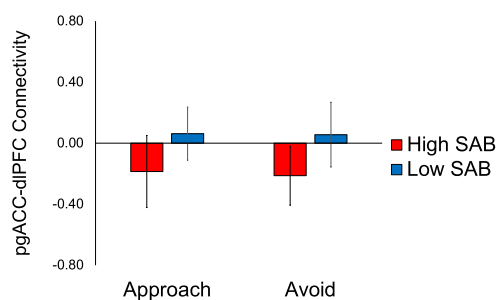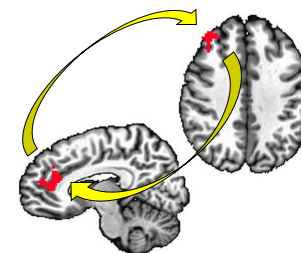
**Fig. 6 Social avoidance behavior modulates anterior cingulate connectivity during subjective motivational responses to social reward-threat conflict signals**. Social Avoidance Behavior (SAB)-related modulation of connectivity between the pregenual anterior cingulate cortex (pgACC) and dorsolateral prefrontal cortex (dlPFC) during subjective approach motivation ratings and avoidance motivation ratings. Neural regions are depicted in radiological convention (left = right). Based on continuous simple slope effects, motivation-related differences in neural activation are depicted at high levels of SAB (+1SD; red bars) or low levels of SAB (−1SD; blue bars). **a** In response to social reward-threat conflict, SAB was associated with weaker pgACC-dlPFC connectivity during avoidance motivation ratings, but not during approach motivation ratings. **b** & **c** In response to social reward or social threat, SAB was not associated with differential pgACC-dlPFC connectivity during avoidance motivation ratings or approach motivation ratings. Note: $***p \leq 0.001$; $**p \leq 0.01$

within the 5 clusters that survived small-volume correction for task-related effects within the *a priori* search territory. In these post-hoc analyses, we observed similar patterns of significant SAB-related modulation within the five task-relevant clusters (see Supplemental Information and Fig. S8). Thus, SAB modulated neural activation in regions that were also functionally relevant to subjective approach and avoidance ratings.

*Exploratory neural connectivity.* Given that previous findings have most consistently implicated ACC regions and the right dlPFC in reward-threat conflict processing, we selected functionally defined pgACC and right dlPFC clusters as seed regions for gPPI analyses.

For the right pgACC seed region, we observed SAB-related modulation of connectivity with a right dlPFC cluster that survived whole-brain correction ($k = 151$, [−34, −39, 34]; Condition × SAB: $B = 0.05$, SE = 0.01; $F_{(1, 264.71)} = 20.64$, $p < 0.001$, uncorrected; FWE $p < 0.05$; Fig. 6a). After controlling for internalizing symptoms (DASS-21), SAB-related modulation of right dlPFC-pgACC connectivity continued to survive whole-brain correction. During subjective avoidance ratings, SAB was associated with significantly lower connectivity between the pgACC and right dlPFC (SAB: $B = −0.04$, SE = 0.008; $F_{(1, 88.62)} = 18.49$, $p < 0.001$, uncorrected). During subjective approach ratings, however, SAB was associated with non-significantly stronger

connectivity between the pgACC and right dlPFC (SAB: $B = 0.02$, SE = 0.009; $F_{(1, 80.16)} = 3.36$, $p = 0.07$, uncorrected). Importantly, we also observed SAB-related modulation of pgACC-dlPFC connectivity that survived whole-brain correction when utilizing the bilateral ACC identified in task-modulation analyses (see Supplemental Information). For the left pgACC and right dlPFC seed regions, however, we did not observe any SAB-related differences in task-related connectivity that survived whole-brain correction.

### Social reward and social threat models

**Behavioral** We did not observe significant SAB-related modulation of subjective motivation ratings as a function of social reward signals (Condition × SAB × Linear: B = −0.01, SE = 0.009; $F_{(1, 292)} = 2.78$, p = 0.10) or social threat signals (Condition × SAB × Linear: B = 0.009, SE = 0.008; $F_{(1, 292)} = 1.49$, p = 0.22). Instead, SAB was associated with generally lower approach motivation ratings and greater avoidance motivation ratings, which was consistent across varying social reward signals and social threat signals (Condition × SAB: both ps < 0.001; Fig. 4).

*Small volume activation.* Similar to the social reward-threat conflict model, no clusters survived small-volume correction for SAB-related modulation of neural activation across morphed stimuli between the approach and avoidance conditions (i.e., no Condition × SAB × Linear/Quadratic interaction). For both the social reward model and social threat model, SAB-related modulation of neural activation differed more generally as a function of approach and avoidance ratings (i.e., a Condition × SAB interaction; Figs. S11 and S12). In both the social reward and social threat models, we observed dlPFC and ACC clusters that overlapped with regions observed in the social reward-threat model and exhibited similar patterns of SAB-related modulation (see Supplemental Information). Additionally, we observed caudate and insula clusters that did not overlap with regions observed in the social reward-threat conflict model (see Supplemental Information).

*Exploratory neural connectivity.* We did not observe SAB-related differences in task-related connectivity that survived whole-brain correction in either the social reward model or social threat model (see Supplemental Information). Moreover, confirmatory analyses using the same pgACC and right dlPFC clusters identified in the social reward-threat conflict model confirmed that SAB did not modulate pgACC-dlPFC connectivity in either the social reward or social threat models (both *p*s > 0.48, see Supplemental Information).

## Discussion

Across multiple levels of analysis, SAB selectively modulated automatic and subjective motivational responses as a function of social reward-threat conflict. For automatic motivational responses, SAB was associated with slower automatic *avoidance* actions as a function of social reward-threat conflict signals (i.e., an inverted U-shaped pattern). As a function of social reward-threat conflict signals, SAB was also associated with relatively stronger amygdala activation during automatic *approach* actions and relatively lower amygdala-pgACC connectivity during automatic *avoidance* actions. For subjective motivational responses, SAB was associated with smaller *increases* in approach motivation ratings and smaller *decreases* in avoidance motivation ratings as social reward signals increased relative to co-occurring social threat signals. Contrary to our hypotheses, however, SAB did not modulate neural activation specifically as a function of social reward-threat conflict signals. Instead, SAB was associated with similar patterns of neural activation during subjective approach and avoidance ratings across social reward, social threat, and social reward-threat conflict more generally. In exploratory connectivity analyses, however, SAB was characterized by weaker pgACC-dlPFC connectivity during subjective avoidance motivation ratings specifically in response to social reward-threat conflict signals. Importantly, SAB-related modulation of motivational responses was independent of more general internalizing symptoms and was generally not observed in response to increasing intensities of social reward signals or social threat signals.

In contrast to our previous behavioral study in which SAB was characterized by *faster* automatic avoidance actions as a function of social reward-threat conflict signals (i.e., a U-shape pattern; Evans & Britton, 2020), SAB was associated with *slower* automatic avoidance actions as a function of social reward-threat conflict signals in the current neuroimaging study (i.e., an inverted U-shape pattern). Although counter to our hypotheses, the specificity of SAB-related modulation to social reward-threat conflict signals suggests that cross-study differences in the direction of SAB-related modulation are not simply attributable to Type I error. In both studies, SAB modulated automatic avoidance actions as a function of social reward-threat conflict signals, but did not modulate automatic approach actions. Moreover, in both studies, SAB did not modulate automatic actions as a function of social reward signals or social threat signals. Together, these results suggest that SAB may be most accurately characterized by dysregulated modulation of automatic avoidance actions as a function of social reward-threat conflict signals, which varies in direction based on contextual factors.

Depending on several contextual factors, SAB may be associated with faster *or* slower automatic avoidance actions as a function of social reward-threat conflict signals. For example, the direction of SAB-related modulation may systematically vary depending on the time point at which automatic action tendencies are measured. Specifically,

participants in the current fMRI study exhibited markedly slower initiation of automatic motivational responses ($M$ = 522.92 ms, $SD$ = 65.91 ms) compared to participants in our previous behavioral study ($M$ = 477.71 ms, $SD$ = 65.01 ms). Previous research using the I-AAT suggests that automatic action tendencies vary in magnitude over the time course of consecutive responses, which contributes to opposing patterns of individual differences during the initiation and subsequent execution of automatic actions (Evans et al., 2021). In the current study, it is possible that the fMRI scanning environment introduced a cognitive load effect, which slowed the initiation of automatic motivational responses and altered the direction of SAB-related modulation. Although future studies will be required to empirically evaluate this interpretation, our behavioral results are nevertheless consistent with previous research demonstrating that SAB exclusively modulates automatic avoidance actions as a function of social reward-threat conflict signals.

Partially consistent with our neural activation hypotheses, SAB exclusively modulated amygdala activation as a function of social reward-threat conflict signals, but not as a function of social reward signals or social threat signals. In contrast to our hypotheses, however, SAB modulated amygdala activation as a function of social reward-threat conflict signals during automatic *approach* actions, but not during automatic *avoidance* actions. Given that SAB did not modulate amygdala activation during automatic avoidance actions, it seems unlikely that SAB-related modulation of automatic avoidance actions is directly attributable to dysregulated amygdala activation. Instead, divergent SAB-related modulation of amygdala activation and automatic action tendencies may be consistent with the amygdala's role in monitoring actions during various types of cognitive conflict (i.e., action-stimulus conflict; Polli et al., 2009; Kim et al., 2004; Etkin et al., 2010; Etkin et al., 2006). Based on this conceptualization, SAB may be associated with greater amygdala activation during automatic approach actions due to action-stimulus conflict between the *selected* action (approach) and *prepotent* action (avoid) as a function of social reward-threat conflict signals (Barbour et al., 2020). Although amygdala signaling facilitates detection of action-stimulus conflict, amygdala signaling does not modulate action selection during cognitive conflict in isolation. Instead, modulation of actions during cognitive conflict is governed by a neural circuit comprised of the right amygdala and ACC (Etkin et al., 2006; Etkin et al., 2010; Etkin et al., 2011; Passamonti et al., 2008). During these types of cognitive conflicts, stronger connectivity between the right amygdala and ACC is associated with more effective modulation of actions (Lütcke & Frahm, 2007; Polli et al., 2009). Thus, SAB-related modulation of automatic avoidance actions may not be attributable to amygdala activation, but instead be attributable to disrupted amygdala-ACC connectivity as a function of social reward-threat conflict signals.

Consistent with this conceptualization, exploratory gPPI analyses demonstrated that SAB was associated with weaker connectivity between the right amygdala and pgACC during automatic avoidance actions as a function of social reward-threat conflict signals. Broadly, amygdala-pgACC connectivity in response to affective facial expressions is proposed to facilitate implicit emotion regulation processes during cognitive conflict (Etkin et al., 2011; Gyurak et al., 2011). Notably, amygdala-pgACC connectivity also regulates negative affect to maintain adaptive action selection and action execution during cognitive conflicts instantiated by social threat signals more specifically (Egner et al., 2008; Kienast et al., 2008; Passamonti et al., 2008). In disorders characterized by social dysfunction such as SAD and MDD (Ottenbreit et al., 2014), amygdala-pgACC connectivity is diminished in response to affective facial expressions, which is proposed to reflect a failure to implicitly regulate negative affect via top-down control processes (Carballedo et al., 2011; Prater et al., 2013; Robert et al., 2021; Wacker-hagen et al., 2019). Therefore, SAB-related modulation of automatic avoidance actions and weaker amygdala-pgACC connectivity may reflect a failure to engage top-down control processes as a function of social reward-threat conflict signals.

Inconsistent with our neural activation hypotheses, however, SAB did not modulate ventral striatum activation as a function of social reward-threat conflict signals. Previous neuroimaging research examined relationships between ventral striatum activation during automatic action tendencies to affective facial expressions and more global measures of trait AA motivation (Radke et al., 2016). In the current study, however, we did not observe SAB-related associations with ventral striatum activation during automatic action tendencies to affective facial expressions. Previous studies have similarly documented brain-behavior relationships with more global measures of trait AA motivation, but not when using more symptom-specific measures (Morys et al., 2020). Thus, it is possible that ventral striatum activation during automatic action tendencies is modulated by trait AA motivation more generally, rather than being modulated by SAB specifically. Instead, SAB may primarily modulate amygdala activation and amygdala connectivity during automatic action tendencies. Therefore, it will be important for future research to dissociate the degree to which trait AA motivation and SAB exert shared and/or distinct influences on neural activation during automatic action tendencies to affective facial expressions.

Within the subjective AAT paradigm, SAB was characterized by weaker *increases* in subjective approach motivation and weaker *decreases* in subjective avoidance motivation as social reward signals increased relative to co-occurring

social threat signals. In contrast, SAB did not parametrically modulate subjective motivational ratings to varying social reward signals or social threat signals. Instead, SAB, was associated with generally lower approach motivation ratings and greater avoidance motivation ratings regardless of the specific intensity of social reward signals or social threat signals. Moreover, although SAB-related differences in subjective motivation responses were maximal at unambiguous social reward signals ($100\%_{Happy}$), SAB was nevertheless not significantly associated with motivational responses to unambiguous social reward signals. Together, this distinct pattern of results suggests that SAB is not associated with dysregulated motivational sensitivity to varying social reward signals or social threat signals. Instead, our results suggest that SAB was specifically characterized by dysregulated modulation of motivational sensitivity as social reward signals increased *relative* to co-occurring social threat signals (i.e., social reward-threat conflict signals). Thus, SAB may be most accurately characterized by a failure to adaptively titrate motivational sensitivity as a function of co-occurring social reward and social threat signals.

At the neural level, SAB was not associated with parametric modulation of neural activation as a function of social reward-threat conflict signals, but was instead with differential activation during subjective approach and avoidance motivation ratings more generally. For example, SAB was associated with greater dlPFC and vlPFC activation during avoidance motivation ratings compared to approach motivation ratings, which did not differ as a function of varying social reward signals and/or social threat signals. In addition to  processes involved in arbitrating approach-avoidance conflict, these regions also play important roles in downregulating negative affect both more generally and in response to social exclusion more specifically (He et al., 2018; He et al., 2019; Kelley et al., 2019; Zhao et al., 2021). Thus, greater SAB-related dlPFC and vlPFC activation during avoidance motivation ratings may reflect the utilization of SAB as a regulatory strategy that reduces negative affect by reducing opportunities for social exclusion (Cacioppo et al., 2011). Within the pgACC, SAB was associated with weaker *de*activation during avoidance motivation ratings and greater *de*activation during approach motivational ratings. Within approach-avoidance conflict paradigms, greater pgACC deactivation may indicate failed integration of reward and/or threat information during decision-making (Ironside et al., 2020). Based on this interpretation, SAB may be associated with weaker reward-threat integration during approach motivational ratings (greater pgACC deactivation) relative to avoidance motivational ratings (lower pgACC deactivation). Finally, it is important to note that the current study utilized a subjective AAT paradigm with ambiguous facial expressions, rather than a more traditional explicit AAT paradigm with unambiguous facial expressions. Thus, it is

perhaps not surprising that we did not observe SAB-related modulation or task-related effects within aPFC regions that have been replicated across previous AAT studies (Roelofs et al., 2009).

More directly aligned with subjective behavioral findings, SAB was associated with lower connectivity between the right dlPFC and pgACC during subjective avoidance motivational ratings in response to social reward-threat conflict. Demonstrating specificity to social reward-threat conflict signals, SAB-related modulation of dlPFC-pgACC connectivity was not observed in response to social reward signals or social threat signals. Previous research suggests that the right dlPFC plays a causal role in titrating reward sensitivity when rewards and threats simultaneously co-occur, which is partly based on computations performed within the ACC (Ballard et al., 2011; Bicks et al., 2015; McDonald et al., 2020; Rolle et al., 2021). Specifically, disrupting the right dlPFC via non-invasive neuromodulation causally reduces reward sensitivity during AA decisions, which is at least partly dependent on dlPFC-ACC connectivity (Rolle et al., 2021). In the current study, SAB was associated with lower dlPFC-pgACC connectivity during avoidance motivational responses exclusively in response to social reward-threat conflict signals. Thus, SAB may be associated with reduced reward sensitivity specifically when social reward signals simultaneously *co-occur* with social threat signals, rather than in response to isolated social reward signals or social threat signals. Consistent with this interpretation, our behavioral results demonstrated that SAB was exclusively associated with parametric changes in motivational sensitivity as the degree of social reward signals decreased relative to *co-occurring* social threat signals. As a whole, these findings dovetail with the functional role of SAB in reducing the probability of social exclusion  in the context of co-occurring social reward and social threat (Cacioppo et al., 2011).

Nevertheless, these fMRI findings should be considered preliminary given our relatively modest sample size. Although the current study was adequately powered to detect SAB-related modulation of behavioral metrics, it is unlikely that our relatively modest sample size was fully powered to detect more subtle patterns of neural activation and/or neural connectivity. A large body of research demonstrates that modest sample sizes reduce the probability that fMRI effects can be successfully replicated in independent samples (Button et al., 2013; Elliott et al., 2020; Grady et al., 2021; Poldrack et al., 2017; Szucs & Ioannidis, 2020; Turner et al., 2019). Although there is no single minimum sample size that could ensure adequate statistical power across all possible fMRI studies, acceptably reproducible fMRI effects may require sample sizes of approximately 80–100 participants (Grady et al., 2021; Turner et al., 2019). Based on these recent estimates, it is

unlikely that the sample size reported in the current study ($n = 30$) is fully powered to detect reproducible patterns of SAB-related modulation or task-related effects in neural activation/connectivity. Relatedly, our neuroimaging results were small-volume corrected for multiple comparisons *within* a family of analyses (e.g., neural activation), but we did not employ a strict multiple comparison *across* each family of analyses. Given these limitations, our fMRI results should be cautiously interpreted as preliminary findings pending an replication in a larger independent sample.

In addition to our relatively modest sample size, several additional limitations should be considered when interpreting the results of the current study. First, we systematically stratified our sample based on SAB to optimize detection of SAB-related modulation, which may obfuscate task-related fMRI effects by introducing heterogeneity in behavioral and neural responses. In the I-AAT, for example, we observed significant SAB-related modulation of amygdala activation as a function of social reward-threat conflict signals. However, we did not observe reliable task-related effects on amygdala activation as a function of social reward-threat conflict signals, which complicates the interpretation of SAB-related modulation. Thus, it will be necessary for future research to utilize the I-AAT and S-AAT within more homogenous, healthy control samples to better characterize task-related effects on neural activation as a function of social reward-threat conflict signals. Second, we morphed static facial expressions to parametrically modulate social reward, social threat, and social reward-threat conflict. However, facial expressions are rarely static in social interactions, but instead vary dynamically as social communication unfolds. Therefore, the current study cannot establish whether parametrically morphed, static facial expressions adequately capture dynamic information conveyed during social interactions. Finally, we exclusively presented Caucasian faces in the I-AAT and S-AAT to minimize the potential confound of perceived racial identity on motivational responses (Paulus & Wentura, 2014). Thus, it will be important for future studies to characterize to what degree perceived racial identity of facial expressions interacts with SAB-related modulation.

Despite these limitations, we believe these findings offer important insights into SAB-related dysregulation of motivational processes with clinical implications. In summary, we observed that SAB was characterized by dysregulated modulation of both automatic and subjective motivational responses, which occurred exclusively as a function of social reward-threat conflict. Most closely aligned with these behavioral findings, we observed that SAB was respectively associated with disrupted amygdala-pgACC connectivity during automatic motivational responses and disrupted dlPFC-pgACC connectivity during subjective motivational responses. Together, these results suggest that SAB is characterized by dysregulated automatic and subjective motivational responses to social reward-threat conflict, which may be partly facilitated by dysregulated pgACC connectivity.

# References

Ambadar, Z., Schooler, J., & Cohn, J. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science, 16*(5), 403–410. https://doi.org/10.1111/j.0956-7976.2005.01548.x

Aupperle, R. L., & Paulus, M. P. (2010). Neural systems underlying approach and avoidance in anxiety disorders. *Dialogues in Clinical Neuroscience, 12*(4), 517–531.

Aupperle, R. L., Melrose, A. J., Francisco, A., Paulus, M. P., & Stein, M. B. (2015). Neural substrates of approach-avoidance conflict decision-making. *Human Brain Mapping, 36*(2), 449–462. https://doi.org/10.1002/hbm.22639

Ballard, I. C., Murty, V. P., Carter, R. M., MacInnes, J. J., Huettel, S. A., & Adcock, R. A. (2011). Dorsolateral prefrontal cortex drives mesolimbic dopaminergic regions to initiate motivated behavior. *The Journal of Neuroscience, 31*(28), 10340–10346. https://doi.org/10.1523/JNEUROSCI.0895-11.2011

Barbour, T., Holmes, A. J., Farabaugh, A. H., DeCross, S. N., Coombs, G., Boeke, E. A., Wolthusen, R. P. F., Nyer, M., Pedrelli, P., Fava, M., & Holt, D. J. (2020). Elevated amygdala activity in young adults with familial risk for depression: A potential marker of low resilience. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 5*(2), 194–202. https://doi.org/10.1016/j.bpsc.2019.10.010

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Basanovic, J., Page, J., & MacLeod, C. (2022). The attenuation of spider avoidance action tendencies in spider-fearful individuals and its impact on explicit evaluation of spider stimuli. *Behaviour Research and Therapy, 151*, 104052. https://doi.org/10.1016/j.brat.2022.104052

Beaver, J. D., Lawrence, A. D., Passamonti, L., & Calder, A. J. (2008). Appetitive motivation predicts the neural response to facial signals of aggression. *The Journal of Neuroscience, 28*(11), 2719–2725. https://doi.org/10.1523/JNEUROSCI.0033-08.2008

Bicks, L. K., Koike, H., Akbarian, S., & Morishita, H. (2015). Prefrontal cortex and social cognition in mouse and man. *Frontiers in Psychology, 6*, 1805. https://doi.org/10.3389/fpsyg.2015.01805

Blalock, J. A., & Joiner, T. E. (2000). Interaction of cognitive avoidance coping and stress in predicting depression/anxiety. *Cognitive Therapy and Research, 24*(1), 47–65. https://doi.org/10.1023/A:1005450908245

Bramson, B., Jensen, O., Toni, I., & Roelofs, K. (2018). Cortical oscillatory mechanisms supporting the control of human social–emotional actions. *Journal of Neuroscience, 38*(25), 5739–5749.

Bramson, B., den Ouden, H. E., Toni, I., & Roelofs, K. (2020a). Improving emotional-action control by targeting long-range phase-amplitude neuronal coupling. *Elife, 9*, e59600.

Bramson, B., Folloni, D., Verhagen, L., Hartogsveld, B., Mars, R. B., Toni, I., & Roelofs, K. (2020b). Human lateral frontal pole contributes to control over emotional approach–avoidance actions. *Journal of Neuroscience, 40*(14), 2925–2934.

Buetti, S., Juan, E., Rinck, M., & Kerzel, D. (2012). Affective states leak into movement execution: Automatic avoidance of threatening stimuli in fear of spider is visible in reach trajectories. *Cognition & Emotion, 26*(7), 1176–1188. https://doi.org/10.1080/02699931.2011.640662

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cacioppo, J. T., & Cacioppo, S. (2014). Social relationships and health: The toxic effects of perceived social isolation. *Social and Personality Psychology Compass, 8*(2), 58–72. https://doi.org/10.1111/spc3.12087

Cacioppo, J. T., & Hawkley, L. C. (2009). Perceived social isolation and cognition. *Trends in Cognitive Sciences, 13*(10), 447–454. https://doi.org/10.1016/j.tics.2009.06.005

Cacioppo, J. T., Hawkley, L. C., Norman, G. J., & Berntson, G. G. (2011). Social isolation. *Annals of the New York Academy of Sciences, 1231*(1), 17–22. https://doi.org/10.1111/j.1749-6632.2011.06028.x

Carballedo, A., Scheuerecker, J., Meisenzahl, E., Schoepf, V., Bokde, A., Möller, H.-J., Doyle, M., Wiesmann, M., & Frodl, T. (2011). Functional connectivity of emotional processing in depression. *Journal of Affective Disorders, 134*(1-3), 272–279.

Carrol, J. M., & Russell, J. A. (1997). Facial expressions in hollywood's portrayal of emotion. *Journal of Personality and Social Psychology, 72*(1), 164–176. https://doi.org/10.1037/0022-3514.72.1.164

Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C. Y., & Barnett, J. H. (2018). Social cognitive dysfunction as a clinical marker: A systematic review of meta-analyses across 30 clinical conditions. *Neuroscience and Biobehavioral Reviews, 84*, 92–99. https://doi.org/10.1016/j.neubiorev.2017.11.014

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). Fmri clustering in afni: False-positive rates redux. *Brain Connectivity, 7*(3), 152–171. https://doi.org/10.1089/brain.2016.0475

Derntl, B., Seidel, E.-M., Eickhoff, S. B., Kellermann, T., Gur, R. C., Schneider, F., & Habel, U. (2011). Neural correlates of social approach and withdrawal in patients with major depression. *Social Neuroscience, 6*(5-6), 482–501. https://doi.org/10.1080/17470919.2011.579800

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science, 317*(5843), 1344–1347. https://doi.org/10.1126/science.1145463

Egner, T., Etkin, A., Gale, S., & Hirsch, J. (2008). Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cerebral Cortex, 18*(6), 1475–1484. https://doi.org/10.1093/cercor/bhm179

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional mri measures? New empirical evidence and a meta-analysis. *Psychological Science, 31*(7), 792–806. https://doi.org/10.1177/0956797620916786

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron, 51*(6), 871–882. https://doi.org/10.1016/j.neuron.2006.07.029

Etkin, A., Prater, K. E., Hoeft, F., Menon, V., & Schatzberg, A. F. (2010). Failure of anterior cingulate activation and connectivity with the amygdala during implicit regulation of emotional processing in generalized anxiety disorder. *American Journal of Psychiatry, 167*(5), 545–554. https://doi.org/10.1176/appi.ajp.2009.09070931

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences, 15*(2), 85–93. https://doi.org/10.1016/j.tics.2010.11.004

Evans, T. C., & Britton, J. C. (2020). Social avoidance behaviour modulates automatic avoidance actions to social reward-threat conflict. *Cognition and Emotion, 34*(8), 1711–1720. https://doi.org/10.1080/02699931.2020.1787353

Evans, T. C., Taylor, C. T., & Britton, J. C. (2021). Characterizing the time course of automatic action tendencies to affective facial expressions and its dysregulation in social anxiety disorder. *Journal of Anxiety Disorders, 78*, 102363. https://doi.org/10.1016/j.janxdis.2021.102363

Fishbach, A., & Shah, J. Y. (2006). Self-control in action: Implicit dispositions toward goals and away from temptations. *Journal of Personality and Social Psychology, 90*(5), 820–832. https://doi.org/10.1037/0022-3514.90.5.820

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364*(1535), 3453–3458. https://doi.org/10.1098/rstb.2009.0142

Gellner, A.-K., Voelter, J., Schmidt, U., Beins, E. C., Stein, V., Philipsen, A., & Hurlemann, R. (2021). Molecular and neurocircuitry mechanisms of social avoidance. *Cellular and Molecular Life Sciences, 78*(4), 1163–1189. https://doi.org/10.1007/s00018-020-03649-x

Grady, C. L., Rieck, J. R., Nichol, D., Rodrigue, K. M., & Kennedy, K. M. (2021). Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fmri data. *Human Brain Mapping, 42*(1), 204–219. https://doi.org/10.1002/hbm.25217

Gutierrez-Garcia, A., & Calvo, M. G. (2014). Social anxiety and interpretation of ambiguous smiles. *Anxiety, Stress, and Coping, 27*(1), 74–89. https://doi.org/10.1080/10615806.2013.794941

Gutiérrez-García, A., & Calvo, M. G. (2016). Social anxiety and perception of (un)trustworthiness in smiling faces. *Psychiatry Research, 244*, 28–36. https://doi.org/10.1016/j.psychres.2016.07.004

Gyurak, A., Gross, J. J., & Etkin, A. (2011). Explicit and implicit emotion regulation: A dual-process framework. *Cognition & Emotion, 25*(3), 400–412. https://doi.org/10.1080/02699931.2010.544160

Hawkley, L. C., Preacher, K. J., & Cacioppo, J. T. (2007). Multilevel modeling of social interactions and mood in lonely and socially connected individuals: The MacArthur social neuroscience studies. In A. D., Ong, & M. H. M., van Dulmen (Eds.), *Oxford handbook of methods in positive psychology*. (pp. 559–575). Oxford University Press.

He, Z., Lin, Y., Xia, L., Liu, Z., Zhang, D., & Elliott, R. (2018). Critical role of the right vlPFC in emotional regulation of social exclusion: A tdcs study. *Social Cognitive and Affective Neuroscience, 13*(4), 357–366. https://doi.org/10.1093/scan/nsy026

He, Z., Liu, Z., Zhao, J., Elliott, R., & Zhang, D. (2019). Improving emotion regulation of social exclusion in depression-prone individuals: A tdcs study targeting right vlPFC. *Psychological*

*Medicine, 50*(16), 2768–2779. https://doi.org/10.1017/S0033291719002915

Heuer, K., Rinck, M., & Becker, E. S. (2007). Avoidance of emotional facial expressions in social anxiety: The approach-avoidance task. *Behavior Research and Therapy, 45*(12), 2990–3001. https://doi.org/10.1016/j.brat.2007.08.010

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Computational Biology, 7*(4), e1002028. https://doi.org/10.1371/journal.pcbi.1002028

Ironside, M., Amemori, K. I., McGrath, C. L., Pedersen, M. L., Kang, M. S., Amemori, S., Frank, M. J., Graybiel, A. M., & Pizzagalli, D. A. (2020). Approach-avoidance conflict in major depressive disorder: Congruent neural findings in humans and nonhuman primates. *Biological Psychiatry, 87*(5), 399–408. https://doi.org/10.1016/j.biopsych.2019.08.022

Ishihara, S. (1917). *Test for colour-blindness.* Handaya, Tokyo, Hongo Harukicho.

Joormann, J., & Gotlib, I. H. (2006). Is this happiness i see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology, 115*(4), 705–714. https://doi.org/10.1037/0021-843x.115.4.705

Kaldewaij, R., Koch, S. B., Volman, I., Toni, I., & Roelofs, K. (2016). *On the control of social approach–avoidance behavior: Neural and endocrine mechanisms Social behavior from rodents to humans* (pp. 275–293). Springer.

Kaldewaij, R., Koch, S. B., Volman, I., Toni, I., & Roelofs, K. (2017). On the control of social approach-avoidance behavior: Neural and endocrine mechanisms. *Current Topics in Behavioral Neurosciences, 30*, 275–293. https://doi.org/10.1007/7854_2016_446

Kaldewaij, R., Koch, S. B., Hashemi, M. M., Zhang, W., Klumpers, F., & Roelofs, K. (2021). Anterior prefrontal brain activity during emotion control predicts resilience to post-traumatic stress symptoms. *Nature Human Behaviour, 5*(8), 1055–1064.

Kashdan, T. B., & Hofmann, S. G. (2008). The high-novelty-seeking, impulsive subtype of generalized social anxiety disorder. *Depression and Anxiety, 25*(6), 535–541. https://doi.org/10.1002/da.20382

Kelley, N. J., Gallucci, A., Riva, P., Romero Lauro, L. J., & Schmeichel, B. J. (2019). Stimulating self-regulation: A review of non-invasive brain stimulation studies of goal-directed behavior. *Frontiers in Behavioral Neuroscience, 12*(337). https://doi.org/10.3389/fnbeh.2018.00337

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology, 2*(3), 320–342. https://doi.org/10.1037/1089-2680.2.3.320

Kienast, T., Hariri, A. R., Schlagenhauf, F., Wrase, J., Sterzer, P., Buchholz, H. G., Smolka, M. N., Gründer, G., Cumming, P., Kumakura, Y., Bartenstein, P., Dolan, R. J., & Heinz, A. (2008). Dopamine in amygdala gates limbic processing of aversive stimuli in humans. *Nature Neuroscience, 11*(12), 1381–1382. https://doi.org/10.1038/nn.2222

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience, 16*(10), 1730–1745. https://doi.org/10.1162/0898929042947865

Koch, S. B. J., Mars, R. B., Toni, I., & Roelofs, K. (2018). Emotional control, reappraised. *Neuroscience and Biobehavioral Reviews, 95*, 528–534. https://doi.org/10.1016/j.neubiorev.2018.11.003

Krieglmeyer, R., & Deutsch, R. (2013). Approach does not equal approach: Angry facial expressions evoke approach only when it serves aggression. *Social Psychological and Personality Science, 4*(5), 607–614. https://doi.org/10.1177/1948550612471060

Krieglmeyer, R., De Houwer, J., & Deutsch, R. (2013). On the nature of automatically triggered approach-avoidance behavior. *Emotion Review, 5*(3), 280–284. https://doi.org/10.1177/1754073913477501

Kupferberg, A., Bicks, L., & Hasler, G. (2016). Social functioning in major depressive disorder. *Neuroscience and Biobehavioral Reviews, 69*, 313–332. https://doi.org/10.1016/j.neubiorev.2016.07.002

Lange, W.-G., Keijsers, G., Becker, E. S., & Rinck, M. (2008). Social anxiety and evaluation of social crowds: Explicit and implicit measures. *Behaviour Research and Therapy, 46*(8), 932–943. https://doi.org/10.1016/j.brat.2008.04.008

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry, 22*, 141–173. https://doi.org/10.1159/000414022

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the depression anxiety stress scales (dass) with the beck depression and anxiety inventories. *Behaviour Research and Therapy, 33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Lütcke, H., & Frahm, J. (2007). Lateralized anterior cingulate function during error processing and conflict monitoring as revealed by high-resolution fmri. *Cerebral Cortex, 18*(3), 508–515. https://doi.org/10.1093/cercor/bhm090

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion, 5*(1), 119–124. https://doi.org/10.1037/1528-3542.5.1.119

Masi, C. M., Chen, H.-Y., Hawkley, L. C., & Cacioppo, J. T. (2011). A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc, 15*(3), 219–266. https://doi.org/10.1177/1088868310377394

Matsumoto, D., & Hwang, H. C. (2014). Judgements of subtle facial expressions of emotion. *Emotion, 14*(2), 349–357. https://doi.org/10.1037/a0035237

McDonald, K. R., Pearson, J. M., & Huettel, S. A. (2020). Dorsolateral and dorsomedial prefrontal cortex track distinct properties of dynamic social behavior. *Social Cognitive and Affective Neuroscience, 15*(4), 383–393. https://doi.org/10.1093/scan/nsaa053

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gppi): A comparison to standard approaches. *Neuroimage, 61*(4), 1277–1286. https://doi.org/10.1016/j.neuroimage.2012.03.068

Morys, F., Janssen, L. K., Cesnaite, E., Beyer, F., Garcia-Garcia, I., Kube, J., Kumral, D., Liem, F., Mehl, N., Mahjoory, K., Schrimpf, A., Gaebler, M., Margulies, D., Villringer, A., Neumann, J., Nikulin, V. V., & Horstmann, A. (2020). Hemispheric asymmetries in resting-state eeg and fmri are related to approach and avoidance behaviour, but not eating behaviour or bmi. *Human Brain Mapping, 41*(5), 1136–1152. https://doi.org/10.1002/hbm.24864

Ottenbreit, N. D., & Dobson, K. S. (2004). Avoidance and depression: The construction of the cognitive–behavioral avoidance scale. *Behavior Research and Therapy, 42*(3), 293–313. https://doi.org/10.1016/s0005-7967(03)00140-2

Ottenbreit, N. D., Dobson, K. S., & Quigley, L. (2014). An examination of avoidance in major depression in comparison to social anxiety disorder. *Behaviour Research and Therapy, 56*, 82–90. https://doi.org/10.1016/j.brat.2014.03.005

Passamonti, L., Rowe, J. B., Ewbank, M., Hampshire, A., Keane, J., & Calder, A. J. (2008). Connectivity from the ventral anterior cingulate to the amygdala is modulated by appetitive motivation

in response to facial signals of aggression. *NeuroImage, 43*(3), 562–570. https://doi.org/10.1016/j.neuroimage.2008.07.045

Paulus, A., & Wentura, D. (2014). Threatening joy: Approach and avoidance reactions to emotions are influenced by the group membership of the expresser. *Cognition and Emotion, 28*(4), 656–677. https://doi.org/10.1080/02699931.2013.849659

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J. B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews. Neuroscience, 18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Polli, F. E., Wright, C. I., Milad, M. R., Dickerson, B. C., Vangel, M., Barton, J. J. S., Rauch, S. L., & Manoach, D. S. (2009). Hemispheric differences in amygdala contributions to response monitoring. *Neuroreport, 20*(4), 398–402. https://doi.org/10.1097/WNR.0b013e328324edb8

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., Mogavero, F., Lobo, A., Olivera, F. J., Lobo, E., Posadas, M., Dukart, J., Kozak, R., Arce, E., Ikram, A., Vorstman, J., Bilderbeck, A., Saris, I., Kas, M. J., & Serretti, A. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews, 97*, 10–33. https://doi.org/10.1016/j.neubiorev.2018.09.012

Prater, K. E., Hosanagar, A., Klumpp, H., Angstadt, M., & Phan, K. L. (2013). Aberrant amygdala-frontal cortex connectivity during perception of fearful faces and at rest in generalized social anxiety disorder. *Depression and Anxiety, 30*(3), 234–241. https://doi.org/10.1002/da.22014

Radke, S., Volman, I., Mehta, P., van Son, V., Enter, D., Sanfey, A., Toni, I., de Bruijn, E. R. A., & Roelofs, K. (2015). Testosterone biases the amygdala toward social threat approach. *Science Advances, 1*(5), e1400074. https://doi.org/10.1126/sciadv.1400074

Radke, S., Seidel, E. M., Eickhoff, S. B., Gur, R. C., Schneider, F., Habel, U., & Derntl, B. (2016). When opportunity meets motivation: Neural engagement during social approach is linked to high approach motivation. *Neuroimage, 127*, 267–276. https://doi.org/10.1016/j.neuroimage.2015.12.014

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*, 23–48. https://doi.org/10.1146/annurev.psych.121208.131647

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry, 38*(2), 105–120. https://doi.org/10.1016/j.jbtep.2006.10.001

Robert, G., Bannier, E., Comte, M., Domain, L., Corouge, I., Dondaine, T., Batail, J.-M., Ferre, J.-C., Fakra, E., & Drapier, D. (2021). Multimodal brain imaging connectivity analyses of emotional and motivational deficits in depression among women. *Journal of Psychiatry & Neuroscience : JPN, 46*(2), E303–E312. https://doi.org/10.1503/jpn.200074

Roelofs, K., Elzinga, B. M., & Rotteveel, M. (2005). The effects of stress-induced cortisol responses on approach–avoidance behavior. *Psychoneuroendocrinology, 30*(7), 665–677.

Roelofs, K., Minelli, A., Mars, R. B., Van Peer, J., & Toni, I. (2009). On the neural control of social emotional behavior. *Social Cognitive and Affective Neuroscience, 4*(1), 50–58.

Rolle, C. E., Pedersen, M. L., Johnson, N., Amemori, K.-I., Ironside, M., Graybiel, A. M., Pizzagalli, D. A., & Etkin, A. (2021). The role of the dorsal–lateral prefrontal cortex in reward sensitivity during approach–avoidance conflict. *Cerebral Cortex*, bhab292. https://doi.org/10.1093/cercor/bhab292

Rolle, C. E., Pedersen, M. L., Johnson, N., Amemori, K. I., Ironside, M., Graybiel, A. M., Pizzagalli, D. A., & Etkin, A. (2022). The role of the dorsal-lateral prefrontal cortex in reward sensitivity during approach-avoidance conflict. *Cerebral Cortex, 32*(6), 1269–1285. https://doi.org/10.1093/cercor/bhab292

Rotteveel, M., & Phaf, R. H. (2004). Automatic affective evaluation does not automatically predispose for arm flexion and extension. *Emotion, 4*(2), 156–172. https://doi.org/10.1037/1528-3542.4.2.156

Rotteveel, M., Gierholz, A., Koch, G., van Aalst, C., Pinto, Y., Matzke, D., Steingroever, H., Verhagen, J., Beek, T. F., Selker, R., Sasiadek, A., & Wagenmakers, E.-J. (2015). On the automatic link between affect and tendencies to approach and avoid: Chen and bargh (1999) revisited. *Frontiers in Psychology, 6*, 335. https://doi.org/10.3389/fpsyg.2015.00335

Rytwinski, N. K., Fresco, D. M., Heimberg, R. G., Coles, M. E., Liebowitz, M. R., Cissell, S., Stein, M. B., & Hofmann, S. G. (2009). Screening for social anxiety disorder with the self-report version of the liebowitz social anxiety scale. *Depression and Anxiety, 26*(1), 34–38. https://doi.org/10.1002/da.20503

Schlund, M. W., Magee, S., & Hudgins, C. D. (2011). Human avoidance and approach learning: Evidence for overlapping neural systems and experiential avoidance modulation of avoidance neurocircuitry. *Behavioral Brain Research, 225*(2), 437–448. https://doi.org/10.1016/j.bbr.2011.07.054

Schlund, M. W., Brewer, A. T., Magee, S. K., Richman, D. M., Solomon, S., Ludlum, M., & Dymond, S. (2016). The tipping point: Value differences and parallel dorsal-ventral frontal circuits gating human approach-avoidance behavior. *Neuroimage, 136*, 94–105. https://doi.org/10.1016/j.neuroimage.2016.04.070

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The mini-international neuropsychiatric interview (m.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10. *Journal of Clinical Psychiatry, 59 Suppl 20*, 22–33 quiz 34-57.

Staugaard, S. R. (2010). Threatening faces and social anxiety: A literature review. *Clinical Psychology Review, 30*(6), 669–690. https://doi.org/10.1016/j.cpr.2010.05.001

Stins, J. F., Roelofs, K., Villan, J., Kooijman, K., Hagenaars, M. A., & Beek, P. J. (2011). Walk to me when i smile, step back when i'm angry: Emotional faces modulate whole-body approach-avoidance behaviors. *Experimental Brain Research, 212*(4), 603–611. https://doi.org/10.1007/s00221-011-2767-z

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990-2012) and of latest practices (2017-2018) in high-impact journals. *Neuroimage, 221*, 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & 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

Turner, B. O., Santander, T., Paul, E. J., Barbey, A. K., & Miller, M. B. (2019). Reply to: Fmri replicability depends upon sufficient individual-level data. *Communications Biology, 2*(1), 129. https://doi.org/10.1038/s42003-019-0379-5

Umberson, D., & Montez, J. K. (2010). Social relationships and health: A flashpoint for health policy. *Journal of Health and Social Behavior, 51*(Suppl), S54–S66. https://doi.org/10.1177/0022146510383501

van Peer, J. M., Rotteveel, M., Spinhoven, P., Tollenaar, M. S., & Roelofs, K. (2010). Affect-congruent approach and withdrawal movements of happy and angry faces facilitate affective categorisation. *Cognition and Emotion, 24*(5), 863–875.

Veenstra, L., Schneider, I. K., Bushman, B. J., & Koole, S. L. (2017). Drawn to danger: Trait anger predicts automatic approach behaviour to angry faces. *Cognition and Emotion, 31*(4), 765–771. https://doi.org/10.1080/02699931.2016.1150256

Volman, I., Roelofs, K., Koch, S., Verhagen, L., & Toni, I. (2011a). Anterior prefrontal cortex inhibition impairs control over social emotional actions. *Current Biology, 21*(20), 1766–1770.

Volman, I., Toni, I., Verhagen, L., & Roelofs, K. (2011b). Endogenous testosterone modulates prefrontal–amygdala connectivity during social emotional behavior. *Cerebral Cortex, 21*(10), 2282–2290.

Vrana, S. R., & Gross, D. (2004). Reactions to facial expressions: Effects of social context and speech anxiety on responses to neutral, anger, and joy expressions. *Biological Psychology, 66*(1), 63–78. https://doi.org/10.1016/j.biopsycho.2003.07.004

Wackerhagen, C., Veer, I. M., Erk, S., Mohnke, S., Lett, T. A., Wüstenberg, T., Romanczuk-Seiferth, N. Y., Schwarz, K., Schweiger, J. I., Tost, H., Meyer-Lindenberg, A., Heinz, A., & Walter, H. (2019). Amygdala functional connectivity in major depression – disentangling markers of pathology, risk and resilience. *Psychological Medicine, 50*(16), 2740–2750. https://doi.org/10.1017/S0033291719002885

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence–second edition*. NCS Pearson.

Wiers, C. E., Stelzel, C., Park, S. Q., Gawron, C. K., Ludwig, V. U., Gutwinski, S., Heinz, A., Lindenmeyer, J., Wiers, R. W., Walter, H., & Bermpohl, F. (2014). Neural correlates of alcohol-approach bias in alcohol addiction: The spirit is willing but the flesh is weak for spirits. *Neuropsychopharmacology, 39*(3), 688–697. https://doi.org/10.1038/npp.2013.252

Zhao, J., Mo, L., Bi, R., He, Z., Chen, Y., Xu, F., Xie, H., & Zhang, D. (2021). The vlpfc versus the dlpfc in downregulating social pain using reappraisal and distraction strategies. *The Journal of Neuroscience, 41*(6), 1331. https://doi.org/10.1523/JNEUROSCI.1906-20.2020

Zorowitz, S., Rockhill, A. P., Ellard, K. K., Link, K. E., Herrington, T., Pizzagalli, D. A., Widge, A. S., Deckersbach, T., & Dougherty, D. D. (2019). The neural basis of approach-avoidance conflict: A model based analysis. *Eneuro, 6*(4), ENEURO.0115-0119.2019. https://doi.org/10.1523/eneuro.0115-19.2019

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.