RUNNING HEAD: Social Approach-Avoidance

Human social defeat and approach-avoidance: Escalating social-evaluative threat and threat of aggression increases social avoidance

Michael W Schlund [a,b] *, Hannah Carter [a], Gloria Cudd [a], Katie Murphy [a], Nebil Ahmed [a], Simon Dymond [c,d] and Erin B Tone [a]

[a] Department of Psychology, Georgia State University

[b] Department of Psychiatry and Behavioral Sciences, University of Pittsburgh

[c] Department of Psychology, Swansea University

[d] Department of Psychology, Reykjavík University

* Corresponding author:
MSchlund@gsu.edu, MichaelSchlundATL@gmail.com
Phone 410-652-8888
Department of Psychology
Georgia State University
P.O. Box 5010
Atlanta, GA 30302-5010
Word counts: Abstract 200; Introduction 1762; Discussion 1174

Abstract

Basic research on avoidance by Murray Sidman laid the foundation for advances in the classification, conceptualization and treatment of avoidance in psychological disorders. Contemporary avoidance research is explicitly translational and increasingly focused on how competing appetitive and aversive contingencies influence avoidance. In this laboratory investigation, we examined the effects of escalating social-evaluative threat and threat of social aggression on avoidance of social interactions. During social-defeat learning, thirty-eight adults learned to associate nine virtual peers with an increasing probability of receiving negative evaluations. Additionally, one virtual peer was associated with positive evaluations. Next, in an approach-avoidance task with social-evaluative threat, one peer associated with negative evaluations was presented alongside the peer associated with positive evaluations. Approaching peers produced a positive or a probabilistic negative evaluation, while avoiding peers prevented a negative evaluation (and forfeited a positive evaluation). In an approach-avoidance task with social aggression, virtual peers gave and took money away from participants. Escalating social-evaluative threat and aggression increased avoidance, ratings of feeling threatened and threat expectancy and decreased ratings of peer favorableness. These findings underscore the potential of coupling social defeat and approach-avoidance paradigms for translational research on the neurobehavioral mechanisms of social approach-avoidance decision-making and anxiety.

Key words: social approach-avoidance, social-evaluative threat, social defeat, anxiety, aggression.

Sticks and stones may break my bones

But words will never hurt me

Early theories categorized avoidance behavior as a Pavlovian conditioned reflex (Bekhterev, 1907, 1913; Watson, 1916). This view dominated psychology for nearly 50 years, until intensive nonhuman laboratory research revealed the distinct contributions of Pavlovian and instrumental learning processes (Baum, 2020; Dymond & Roche, 2009; Herrnstein, 1969; Krypotos et al., 2015; LeDoux et al., 2017; Servatius, 2016). During the mid-20th century, Murray Sidman's research on avoidance learning and generalization (Boren et al., 1959; Sidman, 1953a, 1953b, 1957, 1961, 1962; Sidman et al., 1957) challenged prevailing theories of avoidance that emphasized fear/anxiety and drive reduction (Freud, 1936; Mowrer, 1939, 1951; Miller, 1951) and helped lay the footing for more in-depth investigations of avoidance specifically and negative reinforcement more generally (Herrnstein & Hineline, 1966). Sidman's innovative methods and research findings also provided a framework for conducting research on human avoidance (Ader & Tatum, 1961; Baron & Kaufman, 1966, 1968) that spawned decades of research on uniquely human characteristics, such as instruction following, that modulate operant behavior (Baron et al., 1969; Kaufman et al., 1966).

Sidman held the view that behavioral processes derived from nonhuman experimental research could improve the understanding of human pathological behavior. Sidman (1960a) voiced concerns over the divide between clinical practice and laboratory experimentation. Implicit in his argument was the belief that better treatment outcomes were achievable if abnormal or pathological behavior was not viewed as stemming from aberrant processes, but rather that "…maladaptive behavior can result from quantitative and qualitative combinations of processes which are themselves orderly, strictly determined, and normal in origin" (p. 61). Five decades of human and nonhuman research on avoidance and negative reinforcement seems to support this view, as do the enormous advances during that period in the classification, conceptualization and treatment of avoidance in a wide range of psychological disorders.

To the extent that each version of the American Psychiatric Association's (APA) *Diagnostic and Statistical Manual of Mental Disorders* (DSM) is a snapshot of mainstream views of psychopathology, perusal of all five editions and their revised versions reveals an increasing trend towards inclusion of avoidance among diagnostic criteria. Figure 1 plots both the number of publications on avoidance (Krypotos et al., 2015) and DSM diagnoses with avoidance as a criterion for the past 60 years. In DSM-I (APA, 1952), avoidance is only noted as a coping response for people with a phobic reaction and as a consequence of social detachment in those with schizoid personality. By the third edition (DSM-III; APA, 1980), avoidance was a key element of multiple phobic and anxiety-related conditions, and that number further increased when the third edition was revised. Although the number of disorders with avoidance decreased slightly in subsequent editions (DSM-IV; APA, 1994; DSM-IV-TR, 2000; DSM-5; APA, 2013), this reflected the removal of some diagnoses entirely. It also warrants mention that criteria for several disorders (e.g., major depressive disorder) do not include the words "avoid" or "avoidance", but implicitly make clear that such behavior may characterize affected individuals. Importantly, the trend over the course of the DSM's evolution reflects progressively greater recognition and acceptance of the role of "normal" negative reinforcement processes in facilitating and maintaining pathological behavior.

Insert Figure 1 about here

Notwithstanding criticisms of syndromal classifications as captured by the editions of the DSM, the increase in number and diversity of DSM disorders associated with maladaptive avoidance represents progress towards narrowing the clinical practice/laboratory divide Sidman addressed. Today, avoidance and negative reinforcement are recognized as key features of depression (Taylor et al.,2004; Trew, 2011), obsessive compulsive disorder (Van Ameringen et al., 2014), social anxiety disorder (Bögels et al., 2010; Heimberg, et al., 2014), generalized anxiety disorder (Borkovec et al., 2004), eating disorders (Reas et al., 2005) and avoidant personality disorder (Taylor et al., 2004). Post-traumatic stress disorder now has separate classification categories that distinguish persistent avoidance of thoughts and memories from persistent avoidance of external reminders (Friedman, 2013; Miller et al., 2014). The DSM-5 has added avoidant/restrictive food intake disorder to distinguish it from eating disorders not associated with avoidance (Fisher, et al., 2014). The significance of avoidance and negative reinforcement also has risen in theories of addiction (Koob, 2020; Wise & Koob, 2014); paranoia, paranoid thinking, and beliefs in psychosis (Freeman et al., 2001; Moutoussis et al., 2007); nonsuicidal self-harm (Chapman et al., 2006); and chronic pain (Vlaeyen & Grombez, 2019).

The recent increase in avoidance research shown in Figure 1 (Krypotos et al., 2015) arguably has come with an increasing emphasis on conducting translational studies that will narrow the clinical practice/laboratory divide and provide empirical data to justify DSM classifications (Hofmann, 2014). One emerging research area focuses on how shifts in the competition between appetitive and aversive contingencies control approach and avoidance behavior. A key goal of this research is to model the emergence of avoidance seen in psychopathology, such as choices to avoid social gatherings in social anxiety disorder. In addition, work in this vein is designed to identify and manipulate variables to reduce avoidance and increase approach. To achieve these goals, many researchers employ approach-avoidance (AP-AV) or threat-of-punishment paradigms in which behavior maintained by positive reinforcement is threatened by punishment (Bublatzky et al., 2017; Kirlic et al., 2017; Pittig & Dehler, 2019; Pittig et al., 2018; Schlund et al., 2017, 2020). Emerging neurophysiological research with humans (Aupperle et al., 2015 ; Bach et al, 2014; Schlund et al., 2016; Patrick et al., 2019; Zorowitz et al., 2019) and nonhumans (Burgos-Robles et al., 2017; Capuzzo & Floresco, 2020; Jacobs & Moghaddam, 2020; Schwartz et al., 2017) is also examining how neural systems for reward and threat that historically have been studied independently interact to support approach or avoidance.

Translational research specifically on human social avoidance has been largely neglected, due at least in part to the difficulties arranging ecologically valid social threats in the laboratory. Threat of social aggression (SA) and social evaluative threat (SET) are two widely studied psychosocial stressors that have been linked to psychopathology, stress-related disease and stress-induced behavior change that could be adapted for research on human social avoidance. In preclinical studies, SA often is used with resident-intruder or social defeat paradigms. Social defeat paradigms are effective, ethologically relevant, and reliable models for inducing post-traumatic stress disorder (PTSD), and mood- and anxiety-related symptomology in rodents (Beery & Kaufer, 2015; Hammels et al., 2015; Huhman, 2006; Toth & Neumann, 2013). Social-defeat learning involves repeatedly placing an “intruder” animal in the territory of a larger, aggressive dominant “resident” animal, creating a social conflict which results in the intruder being threatened with attack and/or repeatedly attacked. During subsequent social interaction tests designed to assess approach-avoidance behavior, victimized intruders often show conditioned social avoidance of the dominant resident, as well as of subordinates (Beery & Kaufer, 2015; Hammels et al., 2015; Sandi & Haller, 2015). Pavlovian fear conditioning is the mechanism behind defeat-induced social avoidance (Ayash et al., 2020). Exposure to social defeat reliably elicits increased activation of the hypothalamic-pituitary-adrenal axis, increased blood levels of adrenocorticotropin hormone, increased blood pressure and heart rate, and produces long-lasting changes in brain neurochemistry (Buwalda et al., 2005; Martinez et al., 1998). These findings support predictions that when human participants are confronted with a social interaction involving peers associated with threat of SA, participants will engage in social avoidance and perceive peers as both threatening and unfavorable.

In contrast to SA, SET often is used in human stress research and arises from social interactions, in which there is a threat to one's social esteem, social status, or acceptance that elicits a fear of rejection (Dickerson & Kemeny, 2004). Typically, stress and threat-related responses associated with SET have been induced by requiring participants to give speeches or complete mental arithmetic while being subject to negative peer evaluation (e.g., Kirschbaum et al., 1993). Exposure to SET also reliably elicits increases in cortisol levels (Woody et al., 2018), systolic blood pressure and heart rate (Lehman et al., 2015). Although SET has not historically been used in the context of social defeat paradigms, the literature suggests that confrontation with social interactions that involve peers associated with frequent negative evaluation (i.e., high SET) should elicit similar avoidance and negative peer perceptions.

As part of a human neuroimaging investigation on relations between anxiety and avoidance, this laboratory study assessed the extent to which escalating SET and SA associated with virtual peers affect human social avoidance. Our main interest was in identifying, at the individual-subject level, the magnitude of social threat (i.e., punishment) that overrides control exerted by social reinforcement during a mock social interaction. We hypothesized that exposure to social defeat would produce differential threat (fear) conditioning. We also hypothesized that escalating SET and SA would increase avoidance, along with increases in self-report ratings of negative expectancy and feeling threatened, and decreases in peer favorability. Finally, we examined the internal consistency of the social AP-AV task through assessment of behavioral stability, with the expectation it would exhibit high internal consistency.

Method

Participants

Thirty-eight adult participants (M_{age} = 22.6, *SD* = 3.9; 25 females) were recruited by community flyers. All participants reported being free of psychiatric disorders, brain insult, neurological disorders, and use of medications capable of altering central nervous system functioning. Participants were compensated $5.00 for participation and earned an additional $15 on the AP-AV tasks completed in the 2-hr session. This investigation used deception and was approved by the Institutional Review Board for the Protection of Human Subjects at Georgia State University (GSU). All participants provided written informed consent.

Apparatus

The experiment took place in a small windowless room containing a desk, computer monitor, chair, and standard keyboard. Responses were made with the right hand on a number pad. Experimental events were programmed, and data collected, with software written in the Eprime® platform.

Stimuli

We drew neutral faces used as stimuli (described as peers) from the Georgia State University (GSU) Diverse Faces photoset, which comprises images of 117 models posing happy, sad, angry, fearful, and neutral/calm expressions. Undergraduates (N = 56) rated how much each face conveyed each emotion; faces included in the present task were rated as highly neutral on a scale from 0-100 (M = 82.01, SD=4.14) and minimally expressive of each other emotion (all other mean ratings < 18). Paired samples t-tests showed that the faces were rated as significantly more neutral than they were rated as each other emotion (all *p*'s < .01). Moreover, ratings for the neutral face associated with social reward did not differ significantly from ratings for the other 9 faces (all *p*'s > .05). There is little chance that participants knew peers because the photographs were taken 5-7 years before study participants were enrolled at GSU.

We drew positive and negative words paired with neutral faces from a pool of words that had been rated according to the valence (or pleasantness) of the emotions they invoked (from unhappy to happy) and the degree of arousal and dominance/power (the extent to which the word denotes weakness/submission or strength/dominance) associated with the word (Warriner et al., 2013). The positive (N=70) and negative words (N=85) used differed significantly on valence (*p* < .001; Negative *M* = 2.41, *SD* = .07, Positive *M* = 7.67, *SD* = .11) and dominance (*p* < .001; Negative *M* = 3.90, *SD* = .38, Positive *M* = 6.82, *SD* = .19), but not arousal (*p* = .12; Negative *M* = 4.84, *SD* =.48, Positive *M* = 4.89, *SD* =.11).

Procedure

The methods used closely modeled those used in several prior behavioral and neuroscience investigations on avoidance (Schlund et al., 2015, 2016, 2017, 2020). A within-subject design was used. Deception was used to establish stimulus faces as peers who viewed and acted differently towards participants. After social defeat pretraining, participants completed social interaction tests with an AP-AV task, once with social evaluative threat (SET) and subsequently with social aggression (SA). Participants learned Pavlovian and operant contingencies through experience rather than instructions. Total earnings depended upon AP-AV choices and a small initial stipend. Table 1 lists the order of experimental conditions and primary dependent measures.

Insert Table 1 about here

Social Deception.

A three-step deception procedure was used to increase participant engagement in the social AP-AV task. Across the three steps, the social deception induction procedure employed images of peer faces described as prior research participants (see Figure 2).

Insert Figure 2 about here

Step 1. Participants stood in front of a white wall with a white linen towel draped across their upper chest and shoulders. Participants were asked to look forward and remain expressionless while a headshot was taken with a smartphone. Instructions stated the photo, along with ratings they would provide in future tasks, would be added to our research database.

Participants then were seated and completed the Peer-Evaluation task. The purpose of the task was to have participants make evaluations of other research participants. On each trial, participants viewed an image of a female with a neutral expression, described as a prior research participant, followed by a positive or negative descriptive word (e.g. nice, honest, ugly, lazy). Participants were asked to evaluate the person and rate (yes/no) whether the word described the peer. Thirty trials were presented. The same female image and descriptors were used for all participants. The following instructions were printed on the computer screen and read aloud by the experimenter:

> "In this task you will see a person's face. Like you, this person is a participant in our study. We would like you to evaluate this person's qualities and characteristics based on your experiences with people who look like this person. Later on, they will receive your evaluation. Here is how this task works. The person will appear for 1 s. Next, you will see a descriptor, such as happy, disgusting, sad or loving. There are 30 total. Please enter 1 if you believe the descriptor is less likely to apply this person or a 2 if you think it is more likely to apply to this person. Your judgements are important."

The experimenter pressed 'start' and read the prompt to respond by entering 1 or 2 on a keypad. The experimenter remained in the room for approximately 3 trials.

Step 2. Participants then completed the Peer-Action task. The purpose of the task was to have participants decide whether they would take or give money to other research participants. On each trial, participants viewed an image of a peer for 3 s and then rated how often they would give or take away from them a small amount of money. Nine trials were presented. The same faces were presented to all participants. The following instructions were printed on the computer screen and read aloud by the experimenter:

> "You are about to see the faces of people who have participated in our studies on social decision-making. In this task, we would like you to make decisions based on the qualities and characteristics you see in nine people and on your experiences with people who look like them. Here is the task: imagine you are playing an exchange game with 25-cent rewards. On each trial you will see a person's face for 3 s. Please look at it carefully. Next, you will decide how likely it is you would give a reward to this person or take a

> reward away from this person. Your decisions will determine how much money each person receives or loses. A total of nine faces will appear."

Initially, the experimenter pressed 'start,' the peer was presented, and the experimenter read the rating scale instructions aloud:

> "Based on the qualities and characteristics I see, I would: (enter a number from 1-9 using the scale below)
>
> 1 = take a reward from them 80% of the time
>
> 2 = take a reward from them 60% of the time
>
> 3 = take a reward from them 40% of the time
>
> 4 = take a reward from them 20% of the time
>
> 5 = would not take or give a reward
>
> 6 = give a reward to them 20% of the time
>
> 7 = give a reward to them 40% of the time
>
> 8 = give a reward to them 60% of the time
>
> 9 = give a reward to them 80% of the time"

The experimenter remained in the room for 3-4 trials.

Step 3. Participants were told we use facial recognition software to map characteristics of their headshot photo and then search our research database for nine "prior participants" whose ratings from the peer evaluation task indicate that they vary in how favorably and unfavorably they view other people who resemble the participant. During the ~4 min sham search, "identified" peers appeared one by one and were displayed vertically on the computer screen. The following instructions were printed on the computer screen at the start and read aloud by the experimenter:

> "This software creates a social desirability matrix extending from positive to negative qualities across multiple dimensions. The outcome is a group of nine people that have strong positive and negative feelings about the participant. Press 5 to access the database and begin the facial-character analysis."

Several additional sentences about copyright protection, permissions and assorted bogus legalese were also printed to legitimize the software presentation.

Social-Defeat (Threat) Learning.

A modified social defeat paradigm associated each of the nine "identified" virtual peers with an increasing probability (*p* = 0.0, 0.0, .06, .12, .18, .25, .31, .43, .75) of giving participants negative evaluations (i.e., escalating SET). The aim was to produce peers associated with social punishment that increased the probability of avoidance of mock social interactions in the main AP-AV task.

Figure 3 about here

Pretest. A pretest phase was used to ensure identified peers arranged vertically on the screen were viewed as neutral stimuli and responding was undifferentiated. Pretest also served as a baseline condition for assessing social-defeat learning (see below). Figure 3A shows the vertical arrangement of peers. First, pretest ratings were obtained for each peer regarding level of threat and favorability. Second, participants completed a 90 s task where peers were paired with a neutral prompt. On each trial, a large arrow appeared next to one peer for 2 s. Next, the prompt "---------" appeared for 1 s, indicating that no evaluation occurred. During a subsequent 2-s intertrial interval, a dark blue circle appeared. Nine trials were performed, with each peer highlighted once by the arrow. Afterwards, ratings of how often a negative descriptor appeared for each peer/threat level were obtained. The following instructions were printed on the computer screen and read aloud by the experimenter:

> "This is a 90-s task. Your task is to pay attention. During this task, you will see faces on the screen. Every 5 s or so an arrow will appear next to one face. Your task is to watch where the arrow appears and to notice if you then see a negative descriptor appear (like ugly, unworthy, sad, or unfriendly). A dark blue circle will also appear in the center of the screen. It is provided to help you direct your attention—nothing more. At the end, you will be asked to rate how often you saw a negative descriptor follow each face."

Social-defeat learning. Social-defeat learning involves Pavlovian conditioning to produce differential threat responses. Figure 3B shows peers at levels 1-2 were paired with no negative descriptors (i.e., conditional stimulus, or CS-). In contrast, peers/threat levels 3-9 were paired with an increasing probability of giving participants a negative descriptor (i.e., CS+). Task instructions stated that participants would learn what negative ratings the identified peers had made about people who resemble the participant. Participants were instructed to watch which peers were highlighted by an arrow and to learn how often a negative descriptor appeared for each peer when it was highlighted. Instructions emphasized that learning this relationship would be important for doing well later. Trials consisted of a 1 s peer/threat level presentation in which an arrow appeared next to a face, a 950 ms outcome screen, and a 250 ms ITI. Each peer/threat level was presented for sixteen trials in a randomized order (144 trials). During post-testing, ratings of threat, negative expectancy and peer favorability were obtained for each peer/threat level. Learning was considered stable and ended when posttest ratings of negative expectancy showed an increasing trend across peers/threat levels. The following instructions were printed on the computer screen and read aloud by the experimenter:

> “This is a 5-min task. We call it "What are people saying about you?" During this task, you will see faces on the screen. The faces are people who made negative ratings about people who look like you. Like real life, some people made many more negative ratings about you than others. In this task, every 3 s or so an arrow will appear next to a face. Your task is watch where the arrow appears and if you see a negative descriptor about you appear (like ugly, unworthy, sad, or unfriendly). Your goal is to learn which people gave you negative descriptors AND how often they did so. At the very end, you will be tested on how often you saw negative descriptors follow each person's face. So, learn how often each person gave you a negative descriptor.”

Positive Social Evaluative Learning.

Positive social evaluative learning associated one peer with making positive evaluations about the participant (e.g., Hofmann et al., 2010). The aim was to produce a peer associated with social reinforcement that would increase the probability of approaching mock social interactions in the main AP-AV task.

Pretest. A pretest phase was designed to ensure the peer presented on the screen was viewed as a neutral stimulus and responding was undifferentiated. This pretest served as a baseline condition for positive social evaluative learning. Participants viewed the peer and provided a favorability rating.

Positive social evaluative learning. Figure 3C shows the 2 min task. On each trial, an arrow appeared next to the peer for 1 s. Next, a positive descriptor appeared for 950 ms. During a subsequent 250 ms intertrial interval, a dark blue circle appeared. Ten trials were presented.

Afterwards, ratings of how often a positive descriptor appeared and peer favorability were obtained. The following instructions were printed on the computer screen and read aloud by the experimenter:

> **"**This is a 2-min task. Your task is to pay attention. During this task, you will see a face on the screen. Every 3 s or so an arrow will appear next to the face. Your task is to watch where the arrow appears and if you then see a positive descriptor about you appear (like nice, happy, kind or friendly). These descriptors were made by this participant about people who look like you. A dark blue circle will also appear in the center of the screen. It is provided to help you direct your attention—nothing more. At the end, you will be asked to rate how often you saw a positive descriptor follow the face. So, learn how often you do and do not receive a positive descriptor."

Approach-Avoidance of Social Interactions.

Practice. Figure 3D-F provides a schematic of the discrete trial, AP-AV tasks with SET and SA used for social interaction tests. This practice phase used the AP-AV task with SET and involved trial-and-error learning of AP-AV contingencies. Four blocks of five trials were presented. Each block began with a baseline trial in which the arrow pointed to the prompt "Press #3." (These trials served as baseline trials for a subsequent functional magnetic resonance imaging study and are irrelevant to the goals of the present study.) Blocks 1 and 2 presented the peer associated with positive descriptors alongside the peer at level 1 who was never associated with negative descriptors. Blocks 3-4 presented the peer associated with positive descriptors alongside the peer at level 9 who was almost always associated with negative descriptors. Each trial consisted of a 3 s choice period, 950 ms outcome and a 1-5 s variable ITI. During the intertrial interval ITI, the screen was blank. On each trial, participants chose whether to "join" (approach by pressing button #1) or “pass” (avoid by pressing button #2) the peers. At level 1, approach produced a positive descriptor, while avoidance produced the outcome prompt “-----.“ At level 9, approach produced a positive descriptor or a negative descriptor programmed at *p* = .75, while avoidance produced the outcome prompt “-----.“ Within our discrete trial procedure, we favored the use of two choices (one button for approach and a separate button for avoidance) to distinguish between approach and avoidance This also enabled us to program a negative reinforcement contingency for avoidance in which an active (as opposed to passive) avoidance response was required to prevent contact with aversive stimuli. These features differ from other AP-AV methods that employ a single approach response and infer avoidance by way of reductions in the latency, rate or direction of approach. The following instructions were printed on the computer screen and read aloud by the experimenter:

> This task will last about 2 minutes. We call it "Choosing Social Interactions." During this task, you will decide how often you want to see positive and negative descriptors about people who look like you. That's it!
>
> Here is how it works. On each trial, you will see an arrow pointing to a person who --always-- gives you positive descriptors (left side of screen) and one of the people who gave you negative descriptors (right side of screen). Remember, these people differed in how often they gave you negative descriptors. So, use this information when choosing. When faces appear you will have to make choice.
>
> 1. You can JOIN the pair by pressing #1.
>
> Next, you may see the positive descriptor OR
> you could see the negative descriptor --if-- the second person rated
> people like you negatively. Again, remember how often these people did this.

> 2. (OR) You can PASS on the pair by pressing #2.
>
> Next, the screen will clear and you will not see a positive or negative descriptor. You are free to choose between JOIN and PASS. You choose how many positive and negative descriptors you may see. There are two more things you must know. First, please make a choice on every trial so that we know your preference. Second, once in a while you will see the arrow pointing at <Press #3>. Please press #3 when asked. It is important to the study.

AP-AV task with SET. After practice, the AP-AV task with SET was completed. The nine peer/threat levels were each presented for 8 trials, along with 8 baseline trials, in a randomized order. On each trial, an arrow pointed to the peer associated with positive descriptors and one of the nine peers associated negative descriptors. Approach produced a positive descriptor or a probabilistic negative descriptor, while avoidance prevented a negative descriptor (and forfeited a positive descriptor). Instructions were identical to those used in practice, expect the task duration was shown as 9 min.

AP-AV task with threat of SA. After completing the AP-AV task with SET, participants were instructed they would complete another version of the task in which positive and negative descriptors were replaced with money being received from or taken by peers. The following instructions were printed on the computer screen and read aloud by the experimenter:

> This task will last about 9 minutes. It is much like the other social interaction task where you had to decide how often you wanted to see positive and negative descriptors from other people. In this task, positive descriptors were replaced with giving you some money and negative descriptors were replaced with taking away some of your money. To begin, we will give you 200 cents.

Here is how it works. Once again, on each trial you will see an arrow pointing to the person who --always-- gave you positive descriptors (left side of screen) and one of the people who gave you negative descriptors (right side of screen). Remember, these people differed in how often they gave you negative descriptors. In this task, the person on the left can give you money while people on the right side of the screen can take away some of your money.

When the faces appear, you will have 3 s to make a choice. Use what you learned about these people to make decisions that will earn you the most money.

1. You can choose to JOIN the pair by pressing #1.

Next, you will either EARN MONEY (13 cents) or you will LOSE MONEY (31 cents) if person 2 (on the right side) decides to take it from you.

2. OR You can PASS on the pair by pressing #2.

Next, all the faces will be removed. You will never earn or lose money.

You are free to choose between JOIN and PASS. There are two more things you must know. First, please make a choice on every trial so that we know your preference. Second, once in a while you will see the arrow pointing at <Press #3>. Please press #3 when asked. It is important to the study.

Post-study manipulation checks and debriefing.

Because peers/threat levels were arranged vertically on the computer screen, it is possible that stimulus control could be exerted primarily by the vertical position of faces rather than facial features. Therefore, recognition memory for faces was examined using a pencil and paper assessment in which the faces of the nine virtual peers were embedded in a 3 x 6 field of novel neutral faces. Participants were asked to circle the faces of peers that appeared on the screen (no feedback was provided). Recall memory of the vertical order of face was also examined by asking participants to order a set of index cards with faces of the nine peers (and five novel neutral faces) printed on them in the order displayed on the choice screen. The percentage of correct responses on these two tests provided measures of stimulus control. These assessments parallel social recognition memory tests commonly used in rodent social defeat studies (Sandi & Haller, 2015).

During debriefing participants were fully informed about our deception. Participants were told that the peers identified were not past research participants, the pictures were the same for all participants, and we do not ask anyone to rate pictures of any people taking part in the study. Participants were then asked if they believed our deception (yes or no).

Dependent Measures.

Approach and avoidance responses were made by pressing buttons 1 or 2, respectively, on a computer keypad. Decision time was measured from the onset of the choice display to a key press. The effects of escalating threat were assessed by examining changes in the probability of avoidance and approach responses. Social conflict was assessed by examining changes in decision times. Trials with no choice were excluded from analyses.

Self-report data consisted of peer/threat level-specific ratings of feeling threatened, peer favorability and expectancy of positive and negative evaluations (Boddez et al., 2013). During pre- and posttests, each threat level was individually displayed (randomized order) and ratings were obtained in three categories: Threat ("Please rate how much you (feel/felt) threatened by this person") was measured using a 9-point scale (*0=undecided/unknown,* 1=*Little, 5=Moderate,* 9=*Most threatening*); Favorability ("Please provide a likeability/favorability rating of this person") was measured using a 9-point scale *(0=undecided/unknown, 1=Least favorable, 5=Moderate,* 9=*Most favorable*); Expectancy ("Please rate how often you (would expect to see / saw) a (negative / positive) statement appear from this person?") was measured using a 9-point scale (1=*Never Ever, 0% of the time, 5=Moderate, 40% of the time, 9=A lot, more than 80% of the time*).

Group statistical analyses.

Trials with choices were included in the calculation of all descriptive measures. For group analyses, the assumption of sphericity was tested using Mauchly's test and when it was violated (*p* < .05) a Greenhouse-Geisser correction was used. Pre-post changes in expectancy, feeling threatened and peer favorability ratings across threat levels were examined using each participant as their own control. For each participant, posttest ratings were subtracted from pretest ratings at each threat level. Changes in rating differences across threat levels were examined using one-way repeated measures analysis of variance (ANOVA). Within-condition changes in percentage avoidance across threat levels were examined using one-way repeated measures ANOVA. Decision times also were examined using each participant as their own control. Decision times from threat levels 2 through 9 were subtracted from threat level 1. Changes in the differences across threat levels were examined using one-way repeated measures ANOVA and assessed for quadratic and cubic trends. Paired one-sample t-tests were used to examine pre-post rating differences for positive learning. Significant within-condition changes for Group 3 (see below), which included five participants that responded inconsistently, were examined using the Friedman test, which is a non-parametric alternative to repeated measures ANOVA. Criterion α was set to *p* < .05.

Individual-subject analyses and post-hoc groups.

Because grouped data conceal individual-subject performance, we developed a set of three criteria to apply to individual-subject data to assess the function of evaluations and money gain/loss and changes in contingency control with escalating threat. (#1) Positive evaluations and money gains were considered positive reinforcers when the percentage of trials with approach was > = 75% at threat level 1. (#2) Negative evaluations and money losses were considered negative reinforcers when the percentage of trials with avoidance was > = 75% at level 9. The 75% criterion was used because it reflects that at threat-level 1 approach occurred on 6 of 8 trials and at threat-level 9 avoidance occurred on 6 of 8 trials. (#3) AP-AV transitions were considered to occur when the absolute difference between mean percentage of trials with avoidance at levels 7-9 and levels 1-3 was > = 50%.

Psychometric properties of the AP-AV task.

Internal consistency (or within-subject reliability) was calculated for percentage avoidance and the AP-AV transition threat level using data from both AP-AV tasks. These two dependent measures provide a comprehensive view of contingency control of AP-AV. Independent analyses of both measures used the Spearman-Brown Prophecy Formula and Flanagan-Rulan split-half method score reliabilities, with the highest reliability achievable being 1.0.

Results

Group analyses.

Overall, results of the group analyses presented in Figures 4 and 5 show successful social-defeat learning, positive social evaluative learning, and transitions from approach to avoidance with escalating SET and SA.

Social-defeat learning. Figure 4A presents results from social-defeat learning, in which nine peers were associated with increasing probability of giving participants negative evaluations. In general, pretest ratings showed no systematic variability across peers. However, after social-defeat learning ratings of negative expectancy and feeling threatened increased and peer favorability decreased with escalating threat. Analysis of pre-post rating differences across threat levels yielded evidence of significant changes in expecting negative evaluations, $F(4.62, 170.82) = 164.10$, $p < .001$, $\eta_p^2 = .816$, , 95% CI [.75, .86], feeling threatened, $F(4.25, 170.82) = 11.33$, $p < .001$, $\eta_p^2 = .234$, 95% CI [.12, .33] and peer favorability, $F(3.74, 138.41) = 24.95$, $p < .001$, $\eta_p^2 = .403$, 95% CI [.26, .51].

Figures 4 and 5 about here

Positive social evaluative learning Figure 4B shows positive social evaluative learning, in which one peer was paired with giving participants positive evaluations, was successful. Positive social learning produced a high expectancy of receiving positive evaluations (M =8.29, SD = .41) and a significant pre-post increase in peer favorability (Pretest, M = 5.94, SD = 2.07; Posttest, M = 8.18, SD = 1.25), $t(37) = 6.57$, $p < .001$), d = 0.91, 95% CI [.43, 1.37].

AP-AV performance. Overall, escalating SET and SA produced increases in social avoidance and social conflict. AP-AV performances under SET and SA appear in Figures 4C and 4D. In Figure 4C, escalating SET produced significant increases in percentage avoidance, $F(3.55, 131.54) = 26.12$, $p < .001$, $\eta_p^2 = .414$, 95% CI [.27, .53], and decision times evidenced significant change, $F(5.76, 213.21) = 2.08$, $p = .059$, $\eta_p^2 = .053$, 95% CI [.00, .10], best described by a cubic trend, $F(1, 37) = 5.277$, $p = .027$, $\eta_p^2 = .125$, 95% CI [.00, .28]. This cubic change suggests social conflict was present, as indexed by slower reaction times at middle threat levels. In Figure 4D, escalating SA also produced significant increases in percent avoidance, $F(4.62, 171.07) = 78.41$, $p < .001$, $\eta_p^2 = .679$, 95% CI [.58, .74], and decision times evidenced significant change, $F(5.66, 209.27) = 4.01$, $p = .001$, $\eta_p^2 = .098$, 95% CI [.02, .16], best described by a quadratic trend, $F(1, 37) = 10.77$, $p = .002$, $\eta_p^2 = .225$, 95% CI [.03, .46]. The negative quadratic change is also consistent with social conflict at middle threat levels.

Figure 5 shows AP-AV transitions and response outcomes for the group. Figure 5A highlights the percentage of participants that transitioned from approach to avoidance at each threat level. Transitions for AP-AV with SET occurred near level 6 (*M* = 6.4, *SD* = 1.43). The histogram also reveals that 39% (N=15) of participants did not exhibit a transition. Transitions for AP-AV with SA also occurred near level 6 (*M* = 6.3, *SD* = 1.31) and the histograms reveal that 8% (N=3) of participants did not exhibit a transition. Figure 5B shows the percentage of trials where approach produced a positive reinforcer (positive evaluation / money gain), avoidance produced a negative reinforcer (prevented negative evaluation / prevented money loss), approach produced a punisher (negative evaluation / money loss), and no choice occurred. For AP-AV with SET, the percentage of choices that produced positive reinforcement (*M* = 51, *SD* = 13.45) and negative reinforcement (*M* = 38, *SD* = 18.15) was substantial; few choices were punished (*M* = 9.5, *SD* = 7.01) or did not occur on a trial (*M* = 1.7, *SD* = 2.47). For AP-AV with SA, the percentage of choices that produced positive reinforcement (*M* = 48, *SD* = 10.46) and negative reinforcement (*M* = 45, *SD* = 12.74) was also substantial and few choices were punished (*M* = 6.1, *SD* = 4.18) or not emitted on a trial (*M* = 1.4, *SD* = 2.19).

Individual-subject analyses.

A breakdown of grouped results using our criteria identified notable between-subject performance differences during AP-AV with SET. We observed three different response patterns. Accordingly, participants were subdivided into three post-hoc groups (see Table 2). Participants in *Group 1-Avoided* (N=23, 61%) met our criteria and evidenced increasing avoidance as threat escalated. Participants in *Group 2-Approached* (N=10, 26%) did not meet criterion #1 and primarily approached as threat escalated. Participants in *Group 3-Inconsistent* (N=5, 13%) did not meet criteria and AP-AV varied unsystematically as threat escalated.

During AP-AV with SA, which was completed after the AP-AV with SET, a number of participants showed increased avoidance with escalating threats. We found 9 out of 10 participants in Group 2 (who primarily approached under SET) and 3 out of 5 participants in Group 3 (who responded inconsistently) exhibited a trend towards increasing avoidance as threat escalated (see Table 2). The observed increases in avoidance with escalating SA suggests threat of money loss was a more potent aversive stimulus than negative evaluation. However, this cannot be confirmed because tasks were not counterbalanced and practice effects cannot be ruled out. The following analyses were performed to better understand these between-subject differences.

Social-defeat learning. Analyses of ratings suggest all groups exhibited social-defeat learning, but only Group 1 (*Avoided*) reported feeling threatened by negative evaluations. Figure 6 presents individual-subject ratings during social-defeat learning by group. Figure 6A shows analysis of pre-post rating differences across threat levels yielded evidence of significant changes in expecting negative evaluations for Group 1, $F(4.15, 91.40) = 178.87$, $p < .001$, $\eta_p^2 = .890$, 95% CI [.83, .92], Group 2, $F(2.91, 26.20) = 29.48$, $p < .001$, $\eta_p^2 = .766$, 95% CI [.52, .87], and Group 3, $\chi2(8) = 25.12$, $p = 0.001$, $V = .79$, 95% CI [.51, 1.04]. Thus, each group evidenced successful social-defeat learning. However, Figure 6B shows analysis of pre-post rating differences across threat levels yielded evidence of significant changes in feeling threatened for Group 1, *F*(3.60, 79.24) = 14.75, *p* < .001, η_p^2 = .401, 95% CI [.21, .54], but not Groups 2 and 3 (Group 2: *F*(3.50, 31.51) = 1.71, *p* = .177, η_p^2 = .160, 95% CI [.00, .35]; Group 3: χ2(8) = 13.95, *p* = 0.083, *V* = .59, 95% CI [.42, .84]). Finally, Figure 6C shows analysis of pre-post rating differences across threat levels yielded evidence of significant changes in peer favorability for Group 1, *F*(4.84, 106.58) = 28.01, *p* < .001, η_p^2 = .560, 95% CI [.40, .66], and Group 2, *F*(2.91, 26.23) = 5.12, *p* < .007, η_p^2 = .363, 95% CI [.04, .59], but not Group 3, χ2(8) = 8.13, *p* = 0.421, *V* = .45, 95% CI [.42, .70].

Figure 6 about here

Positive social evaluative learning. Analyses of subgroup ratings suggested positive social learning. Figure 7 presents individual-subject ratings by group. Figure 7A shows pre-post ratings of peer favorability significantly increased for Group 1 and 2 (Group 1: Pretest *M* = 5.95, *SD* = 2.42, Posttest *M* = 8.39, *SD* = 0.89, *t*(22) = 4.85, *p* <.001, *d* = .89, 95% CI [.26, .51]; Group 2: Pretest *M* = 5.90, *SD* = 1.29, Posttest *M* = 8.20, *SD* = 1.32, *t*(9) = 5.13, *p* <.001, *d* = .75, 95% CI [.18, 1.41]) and approached significance for Group 3 (Pretest *M* = 6.0, *SD* = 2.0, Posttest *M* = 7.20, *SD* = 2.17, χ2(1) = 3.0, *p* = 0.083, *V* = .77, 95% CI [.45, 1.71]). Ratings of the expectancy of receiving positive evaluations were similarly high across groups (Group 1: *M* = 8.69, *SD* = 0.63; Group 2: *M* = 7.78, *SD* = 1.64; Group 3: *M* = 7.40, *SD* = 2.03).

Figure 7 about here

AP-AV with social evaluative threat. Figure 8 highlights the effects of escalating SET on individual-subject AP-AV. Figure 7A shows changes in the percentage of trials with avoidance across threat levels by group. Escalating SET produced a significant increase in avoidance in Group 1, $F(4.24, 93.42) = 70.45$, $p < 0.001$, $\eta_p^2 = .762$, 95% CI [.65, .83], and Group 2, $F(3.55, 32) = 3.12$, $p = 0.033$, $\eta_p^2 = .257$, 95% CI [.00, .46], but not in Group 3, $\chi2(8) = 9.6$, $p = 0.294$, $V = .490$, 95% CI [.42, .74]. Figure 8C shows changes in decision time differences (absolute differences relative to threat level 1) across threat levels by group. Escalating SET produced a significant change in decision times for Group 1, $F(4.84, 106.57) = 2.53$, $p = .035$, $\eta_p^2 = .103$, 95% CI [.00, .19], best described by a quadratic trend that is consistent with social conflict, $F(1, 22) = 7.49$, $p = .012$, $\eta_p^2 = .254$, 95% CI [.01, .56]. Group 2 showed a significant change in decision times, $F(8, 72) = 2.35$, $p <.05$, $\eta_p^2 = .207$, 95% CI [.00, .30], that was not quadratic, $F(1, 9) = .747$, $p =.410$, $\eta_p^2 = .077$, 95% CI [.00, .52]. Decision times for Group 3 did not show significant change, $\chi2(8) = 4.16$, $p <= .842$, $V = .322$, 95% CI [.42, .55].

Figure 8 about here

AP-AV with threat of social aggression. Figure 9 highlights the effects of escalating SA on individual-subject AP-AV. Figure 9A shows changes in the percentage of trials with avoidance across threat levels by group. Escalating SA produced a significant increase in avoidance in Group 1, $F(3.82, 84.03) = 84.43$, $p < 0.001$, $\eta_p^2 = .793$, 95% CI [.69, .85], Group 2, $F(3.09, 25.19) = 32.47$, $p < 0.001$, $\eta_p^2 = .802$, 95% CI [.58, .89] and Group 3, $\chi2(8) = 19.77$, $p = 0.011$, $V = .703$, 95% CI [.45, .95]. Figure 9C plots changes in decision time differences across threat levels by group. Escalating SA produced a significant change in decision times for Group 1, $F(8, 176) = 5.34$, $p < 0.001$, $\eta_p^2 = .195$, 95% CI [.07, .27] , best described by a quadratic trend that is consistent with social conflict, $F(1, 22) = 12.56$, $p = 0.002$, $\eta_p^2 = .364$, 95% CI [.07, .64]. Group 2 did not show a significant change in decision times, $F(3.25, 24.29) = 2.19$, $p = 0.106$, $\eta_p^2 = .228$, 95% CI [.00, .46] nor did Group 3, $\chi 2(8) = 2.19$, $p = 0.975$, $V = .233$, 95% CI [.42, .42.].

Figure 9 about here

AP-AV transitions and outcomes. Figure 10A plots the AP-AV transitions for each participant. Transitions under SET for Group 1 ranged between threat levels 3-9, while Groups 2 and 3 did not show clear transitions. All groups showed transitions under SA. Figures 10B-D show the percentage of trials where approach produced a positive reinforcer, avoidance produced a negative reinforcer and approach produced punishment. It is notable that under SET, Group 2 primarily engaged in approach, and therefore, there is a large percentage of trials with positive reinforcement, a small percentage of trials with negative reinforcement and a large percentage of trials with punishment. This suggests negative evaluations did not function as punishers; rather, both positive and negative evaluations functioned as positive reinforcers for Group 2. Similar distributions of outcomes were not present under SA with money gain and loss. Lastly, Figure 10D highlights that choices were omitted on a small percentage of trials under SET and SA.

Figure 10 about here

Manipulation checks and debriefing.

Results of manipulation checks appear in Figures 11A and 11B. One potential methodological weakness is choice may have been controlled by the vertical height of the arrow, rather than the faces of peers. This issue was examined using the unannounced recognition memory test, which required identifying the faces of peers embedded in a field of distractors, and a recall test of the vertical arrangement of faces shown on the choice display. Individual-subject results in Figure 11A show the majority of participants correctly recognized faces (Group 1, *M* = 96%, *SD* = 5.5; Group 2, *M* = 94.2%, *SD* = 8.0; Group 3, *M* = 93.3%, *SD* = 14.9) and correctly reproduced the vertical arrangement of faces (Group 1, *M* = 90.2%, *SD* = 13.6; Group 2, *M* = 94.5%, *SD* = 11.9; Group 3, *M* = 79.9%, *SD* = 24.3). These findings demonstrate stimulus control by peer faces and vertical arrangement.

After debriefing, we queried whether participants believed the faces were prior participants who had provided positive or negative evaluations of people who looked like them. Figure 11B shows a majority of participants reported believing the deception (Group 1 = 85.7%; Group 2 = 100%; Group 3 = 100%). The three participants that did not believe our deception were from Group 1 (*Avoided*), highlighting that contingency control was not dependent upon believing the deception.

Verbal ratings of threat expectancy are widely used in fear-conditioning studies to assess conscious knowledge (or awareness) of the CS-US contingency (Boddez et al., 2013). Therefore, we examined the correspondence between ratings and AP-AV performance. Figure 12 plots individual-subject correlation coefficients that reflect the strength of the relationship between each posttest rating and percent avoidance across threat levels. The higher the correlation, the more changes in ratings aligned with changes in percent avoidance as threat escalated. Figure

12A and 12B show relatively high positive correlations for expectancy and threat ratings and avoidance for participants in Group 1 under SET and SA and Group 2 under SA. Similarly, Figure 12C shows relatively high negative correlations for peer favorabilty ratings and avoidance for participants Group 1 under SET and SA and Group 2 under SA. (Correlations could not be calculated for some participants because avoidance never occurred or there was too little variability in ratings and/or avoidance.)

Figures 11 and 12 about here

Psychometric properties of the AP-AV task.

Internal consistency (or within-subject reliability) was calculated for percentage avoidance and the AP-AV transition threat level using data from both AP-AV tasks. Analyses excluded five inconsistent responders in Group 3 and four participants in Group 2 who exhibited less than 5% avoidance under SET. Figure 13 shows high score reliabilities for avoidance under escalating SET and SA (Spearman-Brown Prophecy Formula, .87; Flanagan-Rulan split-half method, .87). Figure 13 also shows high score reliabilities for threat levels associated with AP-AV transitions under escalating SET and SA (Spearman-Brown Prophecy Formula, .79; Flanagan-Rulan split-half method, .78).

Figure 13 about here

Discussion

In this study, escalating social-evaluative threat and threat of social aggression associated with virtual peers influenced human social avoidance. Group analyses revealed that escalating social-evaluative threat and social aggression increased avoidance, as well as self-report ratings of feeling threatened and threat expectancy. Individual-subject analyses of the effects of escalating social-evaluative threat on approach-avoidance revealed 61% of participants exhibited increasing avoidance and 26% of participants exhibited consistent approach, with little or no avoidance. The remaining 13% of participants showed inconsistent response patterns. Analyses of training data highlighted that these performance differences could not be explained by poor social-defeat learning. Under threat of social aggression, the percentage of participants exhibiting an increase in avoidance as threat escalated rose to 92%, most likely because money loss was a more aversive punisher. Manipulation checks showed high levels of accurately recognizing and recalling the faces of virtual peers and vertical alignment on the AP-AV choice screen. Correlational analyses revealed a high level of correspondence at the individual-subject level between changes in self-report ratings (threat expectancy, feeling threatened and peer favorably) and changes in avoidance. Finally, the approach-avoidance task used in social interaction tests exhibited high internal consistency, reflecting stable within-subject approach-avoidance performances.

These findings complement and extend prior avoidance research in several ways. First, they indicate the potential value of coupling social defeat and approach-avoidance paradigms for investigating human social avoidance. Two different escalating social threats increased human social avoidance in ways consistent with findings reported in nonhuman social defeat (Beery & Kaufer, 2015; Hammels et al., 2015; Huhman, 2006; Toth & Neumann, 2013), human and nonhuman AP-AV, and threat-of-punishment studies (Aupperle, et al., 2015; Bach et al, 2014;

Bublatzky et al., 2017; Burgos-Robles et al., 2017; Capuzzo & Floresco, 2020; Jacobs & Moghaddam, 2020; Pittig & Dehler, 2019; Pittig et al., 2018; Schlund et al., 2016, 2017, 2020; Schwartz et al., 2017; Zorowitz et al., 2019). The use of negative evaluations as a SET to produce avoidance yielded results consistent with those from prior investigations that have used SET to produce anxiety and stress-related responses (Dickerson & Kemeny, 2004). Pairing virtual peers with negative social evaluations during social-defeat learning and pairing one peer with positive social evaluation during positive evaluative learning successfully established peers as conditioned threats and nonthreats. Moreover, much like shock and money loss, response-contingent reductions of negative evaluations and money loss functioned as negative reinforcement for social avoidance. Although there were individual differences, it is notable that both escalating social threats were associated with decreases in ratings of peer favorability. The present findings also represent another systematic replication of AP-AV findings reported by Schlund et al. (2016, 2017, 2020).

It is worth highlighting that individual-subject analyses revealed a significant number of participants (26%) who consistently chose to engage in approach as SET escalated, resulting in a high frequency of receiving negative evaluations. Continued approach despite contact with a putative punisher (i.e., punishment insensitivity) is an established feature of impulsive and uninhibited behavior and associated with antisocial behavior and addiction, as well as extraversion (Byrd et al., 2014; Goldstein & Volkow, 2011; Newman, 1987). It is therefore plausible that our screening procedures failed to exclude participants with significant psychiatric histories or high levels of extraversion. However, when negative evaluations were replaced by money loss, the majority of participants (9 of 10) showed increasing avoidance as threat of aggression escalated, which runs counter to punishment insensitivity. Moreover, participants who exhibited inconsistent approach-avoidance under SET also showed a pattern of increasing avoidance with escalating SA. Overall, these results suggest social evaluations (positive and negative) functioned as a potent positive reinforcer for some participants. Indeed, many participants stated during debriefing that they wanted to see the negative evaluations, which in a social context may be useful, rather than ‘bad’ information (e.g. Fantino & Silberberg, 2010).

The present findings also reveal some of the pitfalls of group analyses. Results of our individual-subject analyses revealed three different behavior patterns during the AP-AV task with SET that suggest SET functioned differently between subjects. Most participants showed increasing avoidance with escalating SET, highlighting that negative social evaluations were punishers. Numerous participants, however, showed consistent approach responding, suggesting that positive and negative social evaluations functioned as positive reinforcers. Lack of experimental control over choice was also evident in several participants. “Hybrid” strategies that combine elements of individual-subject analyses with inferential statistical approaches can provide an effective way to highlight individual-subject effects and enable quantitative analyses of behavior change, especially in EAB experiments with large numbers of participants.

Further investigation is needed to address a number of potential limitations that may limit generalization of findings, but which are addressable in ways consistent with Sidman’s (1960b) views on research design and analysis. Social interaction tests using the approach-avoidance task were relatively brief and increasing exposure may reduce between-subject performance differences. Importantly, increased exposure would ensure avoidance is stable, which is a better approximation of chronic avoidance coping in social anxiety disorders. One methodological weakness was a failure to counterbalance presentations of approach-avoidance tasks with SET and SA. The vertical arrangement of faces on the choice display also was fixed rather than randomized across participants. However, results of group and individual-subject analyses highlighted significant differences between pretest and posttest self-report ratings and manipulation checks revealed accurate recall and recognition of peer faces. Inclusion of skin-conductance measures could have aided in demonstrating that parametric manipulations of social threats generated threat-induced physiological reactions. To initiate contact with the SET and SA contingencies it was necessary to provide extensive task-related instructions. While the stable, predicted patterns of responding were unlikely to have occurred solely as a result of instructional control, future studies must replicate the present findings and investigate the effects of manipulating instructions, contingencies and outcome value and evaluate the necessity of deception.

At a broader level, these findings add to the surge in translational research on avoidance that has occurred since the late 20th century (see Figure 1). This surge has paralleled a steady rise in the number of psychological disorders for which the DSM explicitly or implicitly recognizes avoidance as a key element. The results of the present study suggest, however, that this understanding remains incomplete and point to directions in which translational research on avoidance might fruitfully proceed.

For example, they underscore that avoidance does not manifest in monolithic and undifferentiated ways and suggest that translational research would benefit from careful and precise definitions of avoidance behaviors and their function. As Krypotos et al. (2015) pointed out, the DSM-5 (APA, 2013) defines avoidance in global terms, combining outcomes that may result from varied mechanisms. The DSM also fails to acknowledge the possibility of marked individual differences in the ways that people acquire and maintain avoidant patterns of behavior. Our findings indicate that people who show equivalent levels of learning about the degree to which peers are aversive or reinforcing nonetheless vary in the degree to which that valuation translates into avoidance. Continued research in this vein, which draws heavily on Pavlovian and operant traditions and methods to tease apart significant processes and individual difference variables, will be important in successfully disseminating research findings in ways that are understandable and accessible to mainstream Psychology and incorporated into diagnostic and intervention protocols.

Sidman (1960a) voiced concerns over the clinical practice/laboratory divide and suggested pathological behavior can be understood in terms of "…processes which are themselves orderly, strictly determined, and normal in origin." (p. 61). Looking back at almost 70 years of basic and translational research on avoidance and the increased presence of avoidance in the DSM, there is reason for optimism that the clinical practice/laboratory divide has narrowed. The growing number of scientific disciplines recognizing and investigating avoidance and negative reinforcement in human psychopathology offers additional reasons for optimism.

Conflict of interest

All authors have no conflict of interest.

Funding

Data analyses and manuscript preparation were supported in part by grants from the National Institutes of Mental Health (R01 MH120448-01A1) and Georgia State University Brains and Behavior Program to MS and ET.

Acknowledgments

We thank Anglos Krypotos and Tom Beckers for providing data on publication rates for Figure 1. This paper is dedicated to the memory of Murray Sidman.

References

Ader, R., & Tatum, R. (1961). Free-operant avoidance conditioning in human subjects. *Journal of the Experimental Analysis of Behavior*, *4*(3), 275. https://doi.org/10.1901/jeab.1961.4-275

American Psychiatric Association (1952). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Revision). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

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

Ayash, S., Schmitt, U., & Müller, M. B. (2020). Chronic social defeat-induced social avoidance as a proxy of stress resilience in mice involves conditioned learning. *Journal of Psychiatric Research*, *120*, 64-71. https://doi.org/10.1016/j.jpsychires.2019.10.001

Bach, D. R., Guitart-Masip, M., Packard, P. A., Miró, J., Falip, M., Fuentemilla, L., & Dolan, R. J. (2014). Human hippocampus arbitrates approach-avoidance conflict. *Current Biology*, *24*(5), 541-547. https://doi.org/10.1016/j.cub.2014.01.046

Baron, A., & Kaufman, A. (1966). Human free-operant avoidance of "time out" from monetary reinforcement. *Journal of the Experimental Analysis of Behavior*, *9*(5), 557-565. https://doi.org/10.1901/jeab.1966.9-557

Baron, A., & Kaufman, A. (1968). Facilitation and suppression of human loss-avoidance by signaled, unavoidable loss. *Journal of the Experimental Analysis of Behavior*, *11*(2), 177-185. https://doi.org/10.1901/jeab.1968.11-177

Buwalda, B., Kole, M. H., Veenema, A. H., Huininga, M., de Boer, S. F., Korte, S. M., & Koolhaas, J. M. (2005). Long-term effects of social stress on brain and behavior: a focus on hippocampal functioning. *Neuroscience & Biobehavioral Reviews*, *29*(1), 83-97. https://doi.org/10.1016/j.neubiorev.2004.05.005

Baum, W. M. (2020). Avoidance, induction, and the illusion of reinforcement. *Journal of the Experimental Analysis of Behavior*, *114*(1), 116-141. https://doi.org/10.1002/jeab.615

Bekhterev,V. (1907). *ObjectivePsychology*. St.Petersburg,FL:Soikin.

Bekhterev,V. (1913). *LaPsychologieObjective*. Paris:Alcan.

Beery, A. K., & Kaufer, D. (2015). Stress, social behavior, and resilience: insights from rodents. *Neurobiology of Stress*, *1*, 116-127. https://doi.org/10.1016/j.ynstr.2014.10.004

Boddez, Y., Baeyens, F., Luyten, L., Vansteenwegen, D., Hermans, D., & Beckers, T. (2013). Rating data are underrated: Validity of US expectancy in human fear conditioning. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(2), 201-206. https://doi.org/10.1016/j.jbtep.2012.08.003

Bögels, S. M., Alden, L., Beidel, D. C., Clark, L. A., Pine, D. S., Stein, M. B., & Voncken, M. (2010). Social anxiety disorder: questions and answers for the DSM-V. *Depression and Anxiety*, *27*(2), 168-189. https://doi.org/10.1002/da.20670

Boren, J. J., Sidman, M., & Herrnstein, R. J. (1959). Avoidance, escape, and extinction as functions of shock intensity. *Journal of Comparative and Physiological Psychology*, *52*(4), 420–425. https://doi.org/10.1037/h0042727

Borkovec, T. D., Alcaine, O. M., & Behar, E. (2004). *Avoidance Theory of Worry and Generalized Anxiety Disorder.* In R. G. Heimberg, C. L. Turk, & D. S. Mennin (Eds.), *Generalized anxiety disorder: Advances in research and practice* (p. 77–108). The Guilford Press.

Bublatzky, F., Alpers, G. W., & Pittig, A. (2017). From avoidance to approach: The influence of threat-of-shock on reward-based decision making. *Behaviour Research and Therapy*, *96*, 47-56. https://doi.org/ 10.1016/j.brat.2017.01.003

Burgos-Robles, A., Kimchi, E. Y., Izadmehr, E. M., Porzenheim, M. J., Ramos-Guasp, W. A., Nieh, E. H., ... & Anandalingam, K. K. (2017). Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment. *Nature Neuroscience*, *20*(6), 824-835. https://doi.org/10.1038/nn.4553

Byrd, A. L., Loeber, R., & Pardini, D. A. (2014). Antisocial behavior, psychopathic features and abnormalities in reward and punishment processing in youth. *Clinical Child and Family Psychology Review*, *17*(2), 125-156. https://doi.org/10.1007s10567-013-0159-6.

Capuzzo, G., & Floresco, S. B. (2020). Prelimbic and infralimbic prefrontal regulation of active and inhibitory avoidance and reward-seeking. *Journal of Neuroscience, 40*(24), 4773-4787. https://doi.org/10.1523/jneurosci.0414-20.2020

Chapman, A.L., Gratz, K.L., & Brown, M.Z. (2006). Solving the puzzle of deliberate selfharm: The experiential avoidance model. *Behaviour Research and Therapy, 44*, 371–394. http://dx.doi.org/10.1016/j.brat.2005.03.005

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355. https://doi.org/10.1037/0033-2909.130.3.355

Dymond, S. (2019). Overcoming avoidance in anxiety disorders: The contributions of Pavlovian and operant avoidance extinction methods. *Neuroscience & Biobehavioral Reviews*, *98*, 61-70. https://doi.org/10.1016/j.neubiorev.2019.01.007

Dymond, S., & Roche, B. (2009). A contemporary behavior analysis of anxiety and avoidance. *The Behavior Analyst*, *32*(1), 7-27. https://doi.org/10.1007/bf03392173

Fantino, E., & Silberberg, A. (2010). Revisiting the role of bad news in maintaining human observing behavior. *Journal of the Experimental Analysis of Behavior*, *93*(2), 157-170.

Fisher, M. M., Rosen, D. S., Ornstein, R. M., Mammel, K. A., Katzman, D. K., Rome, E. S., ... & Walsh, B. T. (2014). Characteristics of avoidant/restrictive food intake disorder in children and adolescents: a “new disorder” in DSM-5. *Journal of Adolescent Health*, *55*(1), 49-52. https://doi.org/10.1016/j.jadohealth.2013.11.013

Freeman, D., Garety, P. A., & Kuipers, E. (2001). Persecutory delusions: developing the understanding of belief maintenance and emotional distress. *Psychological Medicine*, *31*(7), 1293.https://doi.org/10.1017/S003329170100455X

Freud, S. (1936). The problem of anxiety. New York: Norton and Co.

Friedman, M. J. (2013). Finalizing PTSD in DSM-5: Getting here from there and where to go next. *Journal of Traumatic Stress*, *26*(5), 548-556. https://doi.org/10.1002/jts.21840

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews Neuroscience*, *12*(11), 652-669. https://doi.org/10.1038/nrn3119

Hammels, C., Pishva, E., De Vry, J., van den Hove, D. L., Prickaerts, J., van Winkel, R., ... & van Os, J. (2015). Defeat stress in rodents: from behavior to molecules. *Neuroscience & Biobehavioral Reviews*, *59*, 111-140. https://doi.org/10.1016/j.neubiorev.2015.10.006

Heimberg, R. G., Hofmann, S. G., Liebowitz, M. R., Schneier, F. R., Smits, J. A., Stein, M. B., ... & Craske, M. G. (2014). Social anxiety disorder in DSM-5. *Depression and Anxiety*, *31*(6), 472-479. https://doi.org/10.1002/da.22231

Herrnstein, R. J. (1969). Method and theory in the study of avoidance. *Psychological Review*, *76*(1), 49. https://doi.org/10.1037/h0026786

Herrnstein, R. J., & Hineline, P. N. (1966). Negative reinforcement as shock-frequency reduction. *Journal of the Experimental Analysis of Behavior*, *9*(4), 421-430. https://doi.org/10.1901/jeab.1966.9-421

Hofmann, S. G. (2014). Toward a cognitive-behavioral classification system for mental disorders. *Behavior Therapy*, *45*(4), 576-587. https://doi.org/10.1016/j.beth.2014.03.001

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin*, *136*(3), 390. https://doi.org/10.1037/a0018916

Huhman, K. L. (2006). Social conflict models: can they inform us about human psychopathology?. *Hormones and Behavior*, *50*(4), 640-646. https://doi.org/10.1016/j.yhbeh.2006.06.022

Jacobs, D. S., & Moghaddam, B. (2020). Prefrontal cortex representation of learning of punishment probability during reward-motivated actions. *Journal of Neuroscience*. https://doi.org/10.1523/ jneurosci.0310-20.2020

Kirlic, N., Young, J., & Aupperle, R. L. (2017). Animal to human translational paradigms relevant for approach avoidance conflict decision making. *Behaviour Research and Therapy, 96*, 14-29. https://doi.org/10.1016/j.brat.2017.04.010

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1-2), 76-81. https://doi.org/10.1159/000119004

Koob, G. F. (2020). Neurobiology of opioid addiction: opponent process, hyperkatifeia, and negative reinforcement. *Biological Psychiatry*, *87*(1), 44-53. https://doi.org/10.1016/j.biopsych.2019.05.023

Krypotos, A. M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: a review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*, *9*, 189. https://doi.org/10.3389/fnbeh.2015.00189

LeDoux, J. E., Moscarello, J., Sears, R., & Campese, V. (2017). The birth, death and resurrection of avoidance: a reconceptualization of a troubled paradigm. *Molecular Psychiatry, 22,* 24-36. https://doi.org/10.1038/mp.2016.166.

Lehman, B. J., Cane, A. C., Tallon, S. J., & Smith, S. F. (2015). Physiological and emotional responses to subjective social evaluative threat in daily life. *Anxiety, Stress, & Coping, 28*(3), 321-339. https://doi.org/10.1080/10615806.2014.968563

Martinez, M., Calvo-Torrent, A., & Pico-Alfonso, M. A. (1998). Social defeat and subordination as models of social stress in laboratory rodents: a review. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *24*(4), 241-256. https://doi.org/10.1002/(SICI)1098-2337(1998)24:4<241::AID-AB1>3.0.CO;2-M

Miller, N. E. (1951). *Learnable drives and rewards.* In S. S. Stevens (Ed.), *Handbook of experimental psychology* (p. 435–472). Wiley.

Miller, M. W., Wolf, E. J., & Keane, T. M. (2014). Posttraumatic stress disorder in DSM-5: New criteria and controversies. *Clinical Psychology: Science and Practice*, *21*(3), 208-220. https://doi.org/10.1111/cpsp.12070

Moutoussis, M., Williams, J., Dayan, P., & Bentall, R. P. (2007). Persecutory delusions and the conditioned avoidance paradigm: towards an integration of the psychology and biology of paranoia. *Cognitive Neuropsychiatry*, *12*(6), 495-510. https://doi.org/10.1080/13546800701566686

Mowrer, O. H. (1939). A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychological Review, 46*(6), 553. https://doi.org/10.1037/h0054288

Mowrer, O. H. (1951). Two-factor learning theory: summary and comment. *Psychological Review*, *58*(5), 350. https://doi.org/10.1037/h0058956

Newman, J. P. (1987). Reaction to punishment in extraverts and psychopaths: Implications for the impulsive behavior of disinhibited individuals. *Journal of Research in Personality, 21*(4), 464-480. https://doi.org/10.1016/0092-6566(87)90033-X

Patrick, F., Kempton, M. J., Marwood, L., Williams, S. C., Young, A. H., & Perkins, A. M. (2019). Brain activation during human defensive behaviour: a systematic review and preliminary meta-analysis. *Neuroscience & Biobehavioral Reviews, 98,* 71-84. https://doi.org/10.1016/j.neubiorev.2018.12.028

Pittig, A., & Dehler, J. (2019). Same fear responses, less avoidance: Rewards competing with aversive outcomes do not buffer fear acquisition, but attenuate avoidance to accelerate subsequent fear extinction. *Behaviour Research and Therapy, 112*, 1-11. https://doi.org/10.1016/j.brat.2018.11.003

Pittig, A., Hengen, K., Bublatzky, F., & Alpers, G. W. (2018). Social and monetary incentives counteract fear-driven avoidance: Evidence from approach-avoidance decisions. *Journal of Behavior Therapy and Experimental Psychiatry*, *60*, 69-77. https://doi.org/10.1016/j.jbtep.2018.04.002

Reas, D. L., Grilo, C. M., Masheb, R. M., & Wilson, G. T. (2005). Body checking and avoidance in overweight patients with binge eating disorder. *International Journal of Eating Disorders*, *37*(4), 342-346. https://doi.org/10.1002/eat.20092

Sandi, C., & Haller, J. (2015). Stress and the social brain: behavioural effects and neurobiological mechanisms. *Nature Reviews Neuroscience*, *16*(5), 290-304. https://doi.org/10.1038/nrn3918

Schlund, M. W., Brewer, A. T., Richman, D. M., Magee, S. K., & Dymond, S. (2015). Not so bad: avoidance and aversive discounting modulate threat appraisal in anterior cingulate and medial prefrontal cortex. *Frontiers in Behavioral Neuroscience, 9*, 142. https://doi.org/10.3389/fnbeh.2015.00142

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

Schlund, M. W., Treacher, K., Preston, O., Magee, S. K., Richman, D. M., Brewer, A. T., ... & Dymond, S. (2017). “Watch out!”: Effects of instructed threat and avoidance on human free-operant approach–avoidance behavior. *Journal of the Experimental Analysis of Behavior*, *107*(1), 101-122. https://doi.org/10.1002/jeab.238

Schlund, M. W., Ludlum, M., Magee, S. K., Tone, E. B., Brewer, A., Richman, D. M., & Dymond, S. (2020). Renewal of fear and avoidance in humans to escalating threat: Implications for translational research on anxiety disorders. *Journal of the Experimental Analysis of Behavior*, *113*(1), 153-171. https://doi.org/10.1002/jeab.565

Schwartz, N., Miller, C., & Fields, H. L. (2017). Cortico-accumbens regulation of approach-avoidance behavior is modified by experience and chronic pain. *Cell reports*, *19*(8), 1522-1531. https://doi.org/10.1016/j.celrep.2017.04.073

Servatius, R. J. (2016). Avoidance: from basic science to psychopathology. *Frontiers in Behavioral Neuroscience, 10*, 15. https://doi.org/10.3389/fnbeh.2016.00015

Sidman, M. (1953a). Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science, 118*, 157–158. https://doi.org/10.1126/science.118.3058.157

Sidman, M. (1953b). Two temporal parameters of the maintenance of avoidance behavior by the white rat. *Journal of Comparative and Physiological Psychology, 46*(4), 253–261. https://doi.org/10.1037/h0060730

Sidman, M. (1955). On the persistence of avoidance behavior. *The Journal of Abnormal and Social Psychology, 50*(2), 217–220. https://doi.org/10.1037/h0039805

Sidman, M. (1957). Conditioned reinforcing and aversive stimuli in an avoidance situation. *Transactions of the New York Academy of Sciences*, *19*(6), 534. https://doi.org/10.1111/j.2164-0947.1957.tb00547.x

Sidman, M. (1960a). Normal sources of pathological behavior. *Science, 132*(3419), 61-68. https://doi.org/10.1126/science.132.3419.61

Sidman, M. (1960b). *Tactics of scientific research.* New York : Basic.

Sidman, M. (1961). Stimulus generalization in an avoidance situation. *Journal of the Experimental Analysis of Behavior, 4*(2), 157. https://doi.org/10.1901/jeab.1961.4-157

Sidman, M. (1962). Reduction of shock frequency as reinforcement for avoidance behavior. *Journal of the Experimental Analysis of Behavior*, 5(2), 247-257. https://doi.org/10.1901/jeab.1962.5-247

Sidman, M., Herrnstein, R. J., & Conrad, D. G. (1957). Maintenance of avoidance behavior by unavoidable shocks. *Journal of Comparative and Physiological Psychology, 50*(6), 553–557. https://doi.org/10.1037/h0043500

Taylor, C. T., Laposa, J. M., & Alden, L. E. (2004). Is avoidant personality disorder more than just social avoidance?. *Journal of Personality Disorders*, *18*(6), 571-594. https://doi.org/10.1521/pedi.18.6.571.54792

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical Psychology Review, 31*(7), 1156-1168. https://doi.org/10.1016/j.cpr.2011.07.007

Toth, I., & Neumann, I. D. (2013). Animal models of social avoidance and social fear. *Cell and Tissue Research*, *354*(1), 107-118. https://doi.org/10.1007/s00441-013-1636-4

Van Ameringen, M., Patterson, B., & Simpson, W. (2014). DSM-5 obsessive-compulsive and related disorders: Clinical implications of new criteria. *Depression and Anxiety, 31*(6), 487-493. https://doi.org/10.1002/da.22259

Vlaeyen, J. W., & Crombez, G. (2019). Behavioral conceptualization and treatment of chronic pain. *Annual Review of Clinical Psychology, 16*. https://doi.org/10.1146/annurev-clinpsy-050718-095744

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191-1207. https://doi.org/10.3758/s13428-012-0314-x

Watson, J. B. (1916). The place of the conditioned-reflex in psychology. *Psychological Review, 23*(2), 89–116. https://doi.org/10.1037/h0070003

Wise, R. A., & Koob, G. F. (2014). The development and maintenance of drug addiction. *Neuropsychopharmacology*, *39*(2), 254-262. https://doi.org/10.1038/npp.2013.261

Woody, A., Hooker, E. D., Zoccola, P. M., & Dickerson, S. S. (2018). Social-evaluative threat, cognitive load, and the cortisol and cardiovascular stress response. *Psychoneuroendocrinology*, *97*, 149-155. https://doi.org/10.1016/j.psyneuen.2018.07.009

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

Figure Captions

Figure 1. *Frequency of publications on avoidance and DSM diagnoses with avoidance as a criterion.* *Avoidance is not named, but is implied in criteria for separation anxiety disorder and dependent personality disorder; Avoidant disorder of childhood removed from DSM after DSM-III-R. **Avoidance explicitly mentioned in criteria for separation anxiety disorder, panic attacks, and agoraphobia; sexual aversion disorder removed from DSM; avoidant/restrictive food intake disorder added. (Frequency data on publications were reproduced with permission from Krypotos et al., 2015.)

Figure 2. *Three step deception procedure.* Deception was used to instill the belief that neutral faces used as stimuli were prior research participants (peers) that varied in how favorably and unfavorably they viewed and acted towards people who look like the participant. [A] In the Peer-Evaluation task, participants viewed an image of a prior research participant followed by a positive or negative descriptor (e.g. honest, nice, ugly, lazy). Participants rated (yes/no) whether the word described the peer. [B] In the Peer-Action task, participants viewed an image of a prior research participant and then rated how often they would give or take away a small amount of money from them. [C] Participants were told facial recognition software uses their image and task ratings to search our research database for nine peers that varied in how favorably and unfavorably they rated other people who resemble the participant. During the ~4 min shame search, nine virtual peers were “identified” and displayed vertically one-by-one on the computer screen. (Faces were not masked during the experiment.)

Figure 3. *Virtual peers used as social threats, pretraining conditions and approach-avoidance tests of social interactions.* [A] Nine virtual peers were used as social threats. [B] Social defeat (threat) learning paired nine peers with increasing probabilities of giving

participants negative evaluations (*threat levels 1-9)*. [C] Positive social evaluative learning paired one peer with giving participants positive evaluations. [D] Approach under social-evaluative threat produced either a positive peer evaluation (positive reinforcer) or a probabilistic negative peer evaluation (punisher). [E] Approach under threat of social aggression (SA) resulted in receiving money from a peer (positive reinforcer) or a probabilistic money loss from a peer (punisher). [F] Avoidance prevented negative evaluations / money loss (negative reinforcement). (Faces were not masked during the experiment.)

Figure 4. *Group social-defeat learning and approach-avoidance performance.* [A] Social-defeat learning pre-post ratings showed escalating social-evaluative threat produced significant increases in ratings of the expectancy of receiving a negative evaluation and feeling threatened, and decreases in ratings of peer favorability. [B] Positive social learning generated high ratings of the expectancy of receiving a positive evaluation and significant pre-post increases in ratings of peer favorability. During social interaction tests with an approach-avoidance task, results showed escalating [C] social-evaluative threat and [D] threat of social aggression produced significant increases in avoidance and decreases in approach. Decision time differences also showed significant cubic/quadratic changes consistent with social conflict. [Eval. = Evaluation. Heavy black lines are group means. Vertical bars are 95% confidence intervals.]

Figure 5. *Group approach-avoidance transitions and outcomes.* The left panel shows results for approach-avoidance with social-evaluative threat (SET). The right panel shows results for approach-avoidance with threat of aggression (SA). [A] Percentage of participants that transitioned from approach to avoidance at each threat level and group mean (*M*). The *None* bin captures participants that never avoided or exhibited unsystematic approach-avoidance. [B]

Percentage of trials in which approach produced a positive reinforcer (SET: received a positive evaluation; SA: received money), avoidance produced a negative reinforcer (SET: prevented negative evaluation; SA: prevented money from being taken by a peer), approach produced a punisher (SET: received a negative evaluation; SA: money taken by a peer) and no choice was emitted. [RF = reinforcement. Vertical bars are 95% confidence intervals.]

Figure 6. *Social-defeat learning.* Analysis of individual-subject performances revealed the sample comprised three different subgroups. We found negative evaluations functioned as negative reinforcers that maintained avoidance for *Group 1-Avoided* (N=23 (61%)), positive reinforcers that maintained approach for *Group 2-Approached* (N=10 (26%)) or failed to maintain consistent approach-avoidance for *Group 3-Inconsistent* (N=5 (13%)). Plots in [A-C] show pre-post changes in ratings with escalating social-evaluative threat by group following social-defeat learning. Escalating social-evaluative threat produced [A] increases in expectancy of receiving a negative evaluation from peers, highlighting successful probability learning for all groups. [B] Ratings of feeling threatened significantly increased only for *Group 1*. [C] Ratings of peer favorability significantly decreased for *Groups 1* and *2*. [Light gray lines represent participants. Heavy black lines are group means. Vertical bars are 95% confidence intervals. Horizontal bars signify significant pre-post differences.]

Figure 7. *Positive social evaluative learning.* [A] Pre-post ratings of peer favorability significantly increased for all groups. [B] Following learning, ratings of the expectancy of receiving a positive evaluation were generally high across groups. [Light gray lines and filled circles represent individual subjects. Heavy black lines are group means. Vertical bars are 95% confidence intervals.]

Figure 8. *Effects of escalating social-evaluative threat on approach-avoidance.* [A] Percentage of trials with avoidance at each threat level. [B] Percentage of trials with approach at each threat level. [C] Decision times differences (absolute differences from threat level 1) at each threat level. [Light gray lines represent participants. Heavy black lines are group means. Vertical bars are 95% confidence intervals.]

Figure 9. *Effects of escalating threat of social aggression on approach-avoidance.* [A] Percentage of trials with avoidance at each threat level. [B] Percentage of trials with approach at each threat level. [C] Decision times differences (absolute differences from threat level 1) at each threat level. Results showed 9 of 10 participants in *Group 2,* who previously approached social-evaluative threat (Figure 8 middle panel), and 3 of 5 participants in *Group 3*, who previously exhibited inconsistent responding (Figure 8 right panel), showed a trend towards increasing avoidance. [Light gray solid (Group 1 and 2) and dashed lines (Group 3) represent participants. Heavy black lines are group means. Vertical bars are 95% confidence intervals.]

Figure 10. *Individual-subject approach-avoidance transitions and outcomes.* The left panel shows results for approach-avoidance with social-evaluative threat. The right panel shows results for approach-avoidance with threat of aggression. [A] Threat levels associated with approach-avoidance transitions. Percentage of trials where choosing [B] approach produced a positive reinforcer, [C] avoidance produced a negative reinforcer, and [D] approach produced a punisher (for more details see Figure 2D-F). [E] Percentage of trials without a choice. [Vertical bars represent participants. Gray horizontal lines highlight group means.]

Figure 11. *Manipulation check: identification of faces and deception.* [A] Percent correctly identifying the faces of nine virtual peers used as threats embedded in a field of eighteen distractor faces (recognition) and ordering virtual peers by threat level (recall). [B]

Percentage of participants that reported believing our deception after debriefing. [Circles represent participants.]

Figure 12. *Manipulation check: correspondence between posttest ratings and percent avoidance across threat levels.* The left panel shows results for approach-avoidance with social-evaluative threat. The right panel shows results for approach-avoidance with threat of aggression. Individual-subject correlations coefficients plotted reflect the strength of the relationship between posttest ratings and percent avoidance across threat levels. [A] Correlations for expectancy ratings and percent avoidance. [B] Correlations for ratings of feeling threatened and percent avoidance. [C] Correlations for favorability ratings and percent avoidance. [Bars represent participants. The absence of data for some participants resulted from never avoiding or little variability in ratings and/or avoidance. Gray horizontal lines highlight group means.]

Figure 13. *Internal consistency of the approach-avoidance task.* (Left) Relationship between percentage of avoidance choices under escalating social-evaluative threat and social aggression. (Right) Relationship between threat levels associated with approach-to-avoidance transitions under escalating social-evaluative threat and social aggression. (Analyses excluded five inconsistent responders in Group 3 and four participants in Group 2 that exhibited less than 5% avoidance under social-evaluative threat. Circles represent participants.)