UCLA
UCLA Electronic Theses and Dissertations

Title

Neural, Physiological, and Behavioral Correlates of Anhedonia – Associations with Pavlovian Learning and Exposure Therapy

Permalink

https://escholarship.org/uc/item/1926p3m9

Author

Rosenberg, Benjamin

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Neural, Physiological, and Behavioral

Correlates of Anhedonia – Associations with

Pavlovian Learning and Exposure Therapy

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy in Psychology

by

Benjamin Rosenberg

2023

ABSTRACT OF THE DISSERTATION

Neural, Physiological, and Behavioral

Correlates of Anhedonia – Associations with

Pavlovian Learning and Exposure Therapy

by

Benjamin Rosenberg

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2023

Professor Michelle Craske, Chair

This dissertation comprises three studies aiming to evaluate the relationship between anhedonia and safety learning. Study 1 asks if anhedonia, over and above other symptom dimensions, is associated with distinct patterns of brain activity during fear extinction. Study 2 asks if anhedonia, or the brain patterns associated with anhedonia, is further related to self-reported or physiological indices of Pavlovian fear learning. Study 3 asks if low positive affect, a core feature of anhedonia, is associated with aberrant prediction error during exposure therapy for social anxiety disorder. Taken together, these studies aim to generate a clearer understanding of associations between anhedonia and fear extinction to inform future innovations in the treatment of anxiety and depression. If symptoms of anhedonia are associated with deficits in

behavioral and biological indices of fear extinction, such deficits may impact the learning

mechanisms central to the efficacy of exposure-based treatments for anxiety disorders.

The dissertation of Benjamin Rosenberg is approved.

Jennifer Ashley Silvers

Jesse A. Rissman

Katherine H. Karlsgodt

Michelle Craske, Committee Chair

University of California, Los Angeles

2023

DEDICATION

This work is dedicated to my *zeyde*, Nathan Rosenberg, who was the first member of his family

to attend college and became my very first academic role model. I wish that he could read this.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

There are so many people who have helped make this work possible. To my grandparents, who made it possible for me to choose this path and consistently set the highest standard for my education – thank you. To my parents, Nancy Small and Gordon Rosenberg, who have always supported me, encouraged me, and given me the opportunity to explore my passions – this work is as much yours as it is mine. To my brother, Zachary Rosenberg, who has always shown me the way – you have no need for any more dissertations in your life. Please disregard this one. To all of my academic mentors throughout the years, Drs. Andy Reed, Laura Carstensen, Katharina Kircanski, Ian Gotlib, Amit Etkin, Desmond Oathes, Roselinde Kaiser, Hakwan Lau, Michelle Craske, and many others – you have inspired me, advocated for me, and fostered my growth as a scientist, thinker, and mentor to others. I simply would not be here without you. I would like to especially acknowledge Dr. Michelle Craske, who welcomed me into her lab and has taught me so many things I cannot list them all here. It is perhaps the most fortunate accident of my career that I ended up working with such a brilliant, generous, kind, and inspiring mentor during my doctoral training. Thank you for being you. Finally, I would like to thank my committee members, Drs. Jennifer Silvers, Jesse Rissman, and Katherine Karlsgodt, for advising and supporting me throughout the completion of my dissertation.

BIOGRAPHICAL SKETCH

Benjamin M. Rosenberg, M.A., C.Phil.

EDUCATION

Graduate   University of California, Los Angeles
                   Doctoral Student, Clinical Psychology
                   Minor: Behavioral Neuroscience

Undergraduate Stanford University
                   Bachelor of Arts with Honors, Psychology

RESEARCH FELLOWSHIPS, HONORS, & AWARDS

2021 – 2022	UCLA Dissertation Year Fellowship
2018 – 2021	National Science Foundation Graduate Research Fellowship
2018	UCLA Graduate Summer Research Mentorship Fellowship
2017 – 2018	UCLA Distinguished University Fellowship
2017	UCLA Doctoral Student Travel Grant
2014	Stanford University Award of Excellence
2013	Stanford University Undergraduate Advising & Research Grant
2012	Stanford University Psych-Summer Research Internship Grant

PEER-REVIEWED PUBLICATIONS

Kuhn, T., Spivak, N.M., Dang, B.H., Becerra. S., Halavi, S.E., Rotstein, N., **Rosenberg, B.M.**, Hiller, S., Swenson, A., … & Bookheimer, S. (2023). Transcranial focused ultrasound selectively increases perfusion and modulates functional connectivity of deep brain regions in humans. *Frontiers in Neural Circuits*, *17*. https://doi.org/10.3389/fncir.2023.1120410

Kaiser, R.H., Peterson, E.C., Moser, A., Neilson, C., Hough, C.M., **Rosenberg, B.M.**, Sandman, C.F., Schneck, C., Miklowitz, D., & Friedman, N.P. (2023). Mood Symptom Dimensions and Developmental Differences in Neurocognition in Adolescence. *Clinical Psychological Science*, *11*(2), 308-325. https://doi.org/10.1177/21677026221111389

Peterson, E.C., Snyder, H.R., Neilson, C., **Rosenberg, B.M.**, Hough, C.M., Sandman, C.F., Ohanian, L., Garcia, S., Kotz, J., … & Kaiser, R.H. (2022). General and Specific Dimensions of Mood Symptoms Are Associated with Impairments in Common Executive Function. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2022.838645

Leuchter, M.K., **Rosenberg, B.M.**, Schapira, G., Wong, N.R., Leuchter, A.F., McGlade, A.L., Krantz, D.E., Ginder, N.D., Lee, J.C., … & Iacoboni, M. (2022). Treatment of Spider Phobia Using Repeated Exposures and Adjunctive Transcranial Magnetic Stimulation (TMS): A Randomized Controlled Pilot and Feasibility Study. *Frontiers in Psychiatry*. https://doi.org/10.3389/fpsyt.2022.823158

Rosenberg, B.M., Taschereau-Dumouchel, V., Young, K.S., Lau, H., Zinbarg, R.E., Nusslock, R., & Craske, M.G. (2022). A multivoxel pattern analysis of anhedonia during fear extinction – implications for safety learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2021.12.008

Rosenberg, B.M.*, Kodish, T.*, Cohen, Z.D., Gong-Guy, E., & Craske, M.G. (2022). Design and Implementation of a Novel Peer-to-Peer Coaching Program to Support Digital Mental Health. *JMIR Mental Health*, *9*(1):e32430. https://doi.org/10.2196/32430

Peterson, E.C., **Rosenberg, B.M.**, Hough, C.M., Sandman, C.F., Neilson, C., Miklowitz, D.J., Kaiser, R.H. (2021). Behavioral Mediators of Stress-Mood Relationships Over Time. *Journal of Affective Disorders*, *294*, 94-102. https://doi.org/10.1016/j.jad.2021.06.079

Oathes, D.J., Zimmerman, J., Duprat, R, Japp, S.S., Scully, M., **Rosenberg, B.M.**, Flounders, M.W., Long, H., Elliott, M., … & Linn, K. (2021). Resting fMRI guided TMS results in subcortical and brain network modulation indexed by interleaved TMS/fMRI. *Experimental Brain Research, 239*(4), 1165-1178. https://doi.org/10.1007/s00221-021-06036-5

Treanor, M., **Rosenberg, B.M.**, & Craske, M.G. (2020). Pavlovian Learning Processes in Pediatric Anxiety Disorders: A Critical Review. *Biological Psychiatry, 89*(7), 690-696. https://doi.org/10.1016/j.biopsych.2020.09.008

Rosenberg, B.M., Mennigen, E., Monti, M.M., & Kaiser, R.H. (2020). Functional segregation of human brain networks across the lifespan: an exploratory analysis of static and dynamic resting-state functional connectivity. *Frontiers in Neuroscience*, *14*, 1275. https://doi.org/10.3389/fnins.2020.561594

McTeague, L.M., **Rosenberg, B.M.**, Huemer, J., Carreon, D.M., Jiang, Y., Yee, A.E., Eickhoff, S.B., & Etkin, A. (2020). Identification of Common Neural Circuit Disruptions in Emotional Processing Across Psychiatric Disorders. *American Journal of Psychiatry*, *177*(5), 411-421. https://doi.org/10.1176/appi.ajp.2019.18111271

Fonzo*, G.A., Goodkind*, M.S., Oathes, D.J., Zaiko, Y.V., Harvey, M., Peng, K.K., Weiss, M.E., Thompson, A.L., Zack, S.E., Mills-Finnerty, C.E., **Rosenberg, B.M.**, … & Etkin, A. (2017). Selective Effects of Psychotherapy on Frontopolar Cortical Function in Post-Traumatic Stress Disorder. *American Journal of Psychiatry*, *174*(12), 1175-1184. https://doi.org/10.1176/appi.ajp.2017.16091073

Shou, H., Yang, Z., Satterthwaite, T.D., Cook, P.A., Bruce, S.E., Shinohara, R.T., **Rosenberg, B.M.**, & Sheline, Y.I. (2017). Cognitive behavioral therapy increases amygdala connectivity with the cognitive control network in both MDD and PTSD. *NeuroImage: Clinical, 14,* 464-470. https://doi.org/10.1016/j.nicl.2017.01.030

GENERAL INTRODUCTION

Anxiety disorders represent a highly prevalent and disabling category of psychological diagnoses (Yang et al., 2021). Psychological treatments for anxiety are widely studied and supported (Bandelow et al., 2018; Carpenter et al., 2018; Van Dis et al., 2020), with exposure therapy considered a gold-standard (Rauch et al., 2021; Abramowitz et al., 2019). However, despite the considerable effectiveness of exposure therapy, many anxious individuals do not experience clinically significant gains from treatment (Loerinc et al., 2015). It is critical for scientists to develop an understanding of the psychological dimensions that are associated with differences in the core mechanisms of exposure therapy. Such insights may fuel future innovations in the treatment of anxiety disorders.

The following sections outline the current state of the literature. Pavlovian learning is a long-studied and well-validated area of research that has implications across many areas of psychological science. Studies of Pavlovian fear learning, and particularly fear extinction, have aided the field's understanding of the mechanisms underlying successful exposure therapy (Craske et al., 2008; Craske et al., 2012; Craske et al., 2014; Pittig et al., 2016; Scheveneels et al., 2016). Pavlovian learning is likewise associated with specific neurocircuitry that is altered in anxiety disorders. Anxiety disorders are highly comorbid with depressive disorders, and co-occurring symptoms are associated with greater severity of functional impairment. Many depressed individuals experience anhedonia, a symptom dimension characterized by low positive affect, which is associated with broad deficits in learning across a range of psychological paradigms. There is a growing literature supporting high positive affect as a correlate of Pavlovian fear learning and memory retention (Zbozinek & Craske, 2017b), which is of particular relevance to anxiety disorders. Likewise, anhedonia (i.e., low positive affect) is

1

associated with widespread differences in neural functioning, including some of the same circuits implicated in Pavlovian fear learning. These areas collectively motivate three empirical studies aimed at understanding how anhedonia relates to mechanisms of fear extinction and exposure therapy.

Pavlovian Fear Learning

The classic Pavlovian paradigm involves learning that an inherently valenced stimulus (unconditioned stimulus; US) has been repeatedly paired with a neutral stimulus (conditioned stimulus; CS), such that the CS becomes a meaningful signal that reliably provokes a conditioned response (CR). Within the Pavlovian literature, one particularly interesting area of study concerns the acquisition of learned fears. The standard paradigm involves learning that a CS (CS+) predicts the delivery of an aversive US (e.g., shock), such that individuals begin to show increased physiological arousal (i.e., activation of the sympathetic nervous system, CR) in response to the CS prior to US delivery. A different CS is typically included as a control stimulus (CS-), such that it never predicts the US and therefore corresponds with minimal physiological arousal.

Following fear acquisition, many experiments then involve extinction, in which the CS is no longer paired with a US (CS-NoUS). Through this process, an individual learns that a previously threatening CS has become safe (extinguished CS; CS+E). Successful fear extinction is thought to rely on prediction-error signaling during the unexpected omission of an anticipated US – for the strongest CS-US predictions, omission of the US results in the most substantial learning and subsequent updating of CS-US predictions (Rescorla & Wagner, 1972). Of note, contemporary models highlight inhibitory learning as a central mechanism of this process. Rather than "unlearning" the original CS-US association, extinction involves strengthening the

CS-NoUS association. In retaining the original CS-US association, an individual can more readily reacquire fear for a CS+E than for a neutral CS (Bouton, 2002). Likewise, retention of the extinguished CS-US relationship enables the individual to recognize contextual factors that govern where and when a specific cue is more likely to be threatening, thereby adapting its behavior according to the context (Bouton, 2002; Bouton, 2004; Bouton & Moody, 2004).

Pavlovian Learning and the Treatment of Anxiety Disorders

Pavlovian learning processes are not confined to the laboratory. Existing literature highlights a critical role for Pavlovian learning in supporting the development and maintenance of anxiety disorders (Mineka & Oehlberg, 2008; Mineka & Zinbarg, 2006; Treanor et al., 2020), a category of highly prevalent and impairing diagnoses (Kessler et al., 2010). Although trait anxiety does not appear to correspond with aberrant fear learning (Lonsdorf & Merz, 2017; Torrents-Rodas et al., 2013), individuals diagnosed with anxiety disorders demonstrate stronger CRs during fear acquisition and extinction, compared with healthy control subjects (Duits et al., 2015; Lissek et al., 2005).

Importantly, predominant models of exposure therapy are built upon principles of Pavlovian learning, particularly extinction, in order to facilitate a client's learning that a previously threatening cue is now safe (Craske et al., 2008; Craske et al., 2012; Craske et al., 2014). For example, an individual may have previously learned an association between dogs (CS) and biting (US), such that this individual will experience physiological arousal (CR) around dogs or in contexts likely to contain dogs. As the CR is itself aversive, many individuals learn to avoid interacting with dogs altogether, thereby reducing physiological arousal and fostering a pleasant experience of relief (Willems & Vervliet, 2021). Through repeated avoidance over time, the individual is unlikely to learn a CS-noUS relationship (e.g., "not all dogs bite"), such that the

3

individual will continue to perceive the CS as threatening (e.g., "dogs are still scary") (Lovibond et al., 2009). As a result, the individual is likely to remain avoidant of dogs, such that they will continue to endorse high CS-US predictions and may develop a clinical phobia.

Exposure therapy emphasizes a process similar to fear extinction (Pittig et al., 2016; Scheveneels et al., 2016), wherein the phobic patient repeatedly interacts with a CS in the absence of a US (e.g., approaching dogs without getting bitten). In line with the findings of Rescorla and Wagner discussed above, this violation of expectations results in the strengthening the CS-NoUS relationship, such that the individual learns that the CS is safe (e.g., "most dogs do not bite") (Craske et al., 2014). Patients may continue to endorse strong CS-US predictions (e.g., "some dogs bite"), but by the end of successful exposure therapy, the CS-NoUS relationship has become most prominent. As this learning process is critical to successful exposure therapy, individual differences in extinction learning and associated neurocircuitry may inform why treatment is more effective for some patients than others.

Neurocircuitry of Fear Learning

Neuroscience has played an important role in characterizing the basic circuitry supporting fear learning processes in humans and in generating novel approaches to treatment of anxiety disorders (Milad et al., 2014). The extant literature has highlighted roles for the amygdala, dorsal anterior cingulate cortex (dACC), anterior insular cortex (ACC), and ventromedial prefrontal cortex (vmPFC) during fear acquisition (Battaglia et al., 2020; Etkin & Wager, 2007; Fullana et al., 2016; Greco & Liberzon, 2016; Milad et al., 2014; Sehlmeyer et al., 2009), although there are inconsistencies in the precise role of the amygdala in such studies (Fullana et al., 2016). Existing research has further highlighted activation of the medial prefrontal cortex, amygdala, and hippocampus during fear extinction (Greco & Liberzon, 2016; Milad et al., 2014) and the

4

vmPFC during fear extinction recall (Greco & Liberzon, 2016). Likewise, the extent to which an individual experiences threat-induced physiological arousal is correlated with activation of the amygdala during fear extinction and the dorsal ACC during fear extinction recall (Marin et al., 2020).

These basic neural associations are implicated in clinical studies of anxiety, as well. For instance, diagnosis of an anxiety disorder is associated with reduced activation of the vmPFC during fear extinction recall (Marin et al., 2020). Fear extinction studies are particularly implicated in the translational neuroscience literature of anxiety (Milad & Quirk, 2012), with safety signals tending to correspond with reduced activation of the vmPFC (Fullana et al., 2020). Greater vmPFC activation and weaker amygdala or insular activation during fear extinction has similarly been shown to predict poorer treatment responses in social anxiety disorder (Ball et al., 2017).

Similar neurocircuitry is involved in studies of threat responding and defensive behaviors, specifically implicating a circuit containing the vmPFC, amygdala, ACC, periaqueductal grey, and insula (Mobbs et al., 2007; Mobbs et al., 2009). The vmPFC has been highlighted as one region that, in concert with other regions such as the posterior cingulate cortex and hippocampus, is responsible for long-term decision-making processes in evaluating the relative costs and benefits of approach/avoidance behaviors (Qi et al., 2018; Mobbs et al., 2020). In contrast, "reactive" fear circuits, including regions such as the midcingulate cortex and periaqueductal grey, are thought to be responsible for in-the-moment decisions as the US is most proximal (Qi et al., 2018).

Although the extant literature is limited, successful treatment with exposure therapy appears to be associated with changes in the functioning of many of these same brain regions,

including reduced functional activation of the dorsolateral prefrontal cortex, amygdala, insular cortex, and vmPFC (Hauner et al., 2012) as well as the subgenual ACC (Helpman et al., 2016). Exposure therapy is also associated with changes in functional connectivity between the inferior frontal gyrus with the amygdala, insular cortex, and ACC (Kircher et al., 2013), between the amygdala and ACC (Lueken et al., 2013; Sandman et al., 2020), and between the frontopolar cortex and the vmPFC (Fonzo et al., 2017). It is thought that these neural changes are representative of strengthening top-down regulation of fear neurocircuitry during the course of successful therapy.

Comorbidity of Anxiety and Depression

Thus, the behavioral and neural mechanisms of fear learning are well-understood, and insights from the fear learning literature have been crucial to the development of effective treatments for anxiety disorders. However, anxiety disorders are a heterogeneous category of diagnoses including specific fears (e.g., phobias), widespread fears (e.g., generalized anxiety disorder), and fears about the physiological experience of fear (e.g., panic disorder). Likewise, anxiety disorders frequently occur in the context of co-occurring diagnoses, which are often linked to greater symptom severity and poorer treatment outcomes (Kessler et al., 2010; Kroenke et al., 2007).

Comorbidity among anxiety and depressive disorders is particularly common and is associated with greater impairment than either diagnosis occurring alone (Belzer & Schneier, 2004; Brown et al., 1996; Kessler et al., 2008; Pollack, 2005; Breteler et al., 2021). Prior studies have shown a bidirectional pathway between these diagnoses, wherein anxiety and depression are risk factors for the future onset of one another (Jacobsen & Newman, 2017). Comorbidity may be partially attributable to overlapping diagnostic criteria between depression and anxiety

6

disorders, particularly generalized anxiety disorder (Zbozinek et al., 2012), suggesting the presence of overarching dimensions common to both diagnostic categories. As a result, researchers have begun applying dimensional models of psychopathology to disentangle to the overlapping and unique features of seemingly different pathology categories (Cuthbert, 2014; Cuthbert & Insel, 2013; Kotov et al., 2017). It is thought that, by deconstructing these seemingly distinct entities, it may be possible to create personalized, process-targeted treatments that yield superior treatment outcomes (Brown & Barlow, 2009; Insel, 2014; Thompson-Hollands et al., 2014).

To account for the considerable overlap among diagnoses, dimensional models of psychopathology utilize a hierarchical structure, wherein broad factors may be shared across diagnostic categories and then contribute to unique clinical phenotypes. The Hierarchical Taxonomy of Psychopathology (HiTOP) is one such model, which aims to improve diagnostic classification by grouping clinical syndromes on the basis of covarying symptom profiles (i.e., likely pathways of comorbidity) (Kotov et al., 2007; Krueger et al., 2008). Anxiety and depression are closely related in this model, although some anxiety disorders (e.g., generalized anxiety disorder, posttraumatic stress disorder) are considered more related to depression than others (e.g., panic disorder, specific phobias).

Given the high comorbidity of anxiety and depression, hierarchical structures have been developed to specifically address their overlapping characteristics. For instance, the Tri-Level Model of depression and anxiety describes a single broad factor of symptoms characteristic of both anxiety and depression ("General Distress"), as well as intermediate factors specific to depression ("Anhedonia-Apprehension") and anxiety ("Fears") (Naragon-Gainey et al., 2016; Prenoveau et al., 2010). The Anhedonia-Apprehension factor of this model primarily includes

symptoms relating to diminished positive affect, as well as some apprehension symptoms that are unique to this factor beyond General Distress and Fears. The Fears factor primarily includes symptoms relating to anxious arousal, worry, and specific fears that are unique to this factor beyond General Distress and Anhedonia. Such a model enables researchers to evaluate the effects of a specific dimension (e.g., Anhedonia-Apprehension) over and above the effects of other dimensions that are relevant to the symptomatology of anxiety and depression (e.g., Fears and General Distress). This level of specificity can, in turn, inform efforts to reduce diagnostic heterogeneity and to develop personalized treatments.

Positive Emotions and Learning

Positive emotions have been associated with differences in learning across a range of psychological paradigms. For example, positive mood has been shown to deepen encoding of newly learned associations and enhance long-term retention of learned information (Federmeier et al., 2001; Hänze & Hesse, 2008; Isen, 1987). Positive mood has been further shown to deepen mental rehearsal of stored information, thereby enhancing consolidation and retrieval of long-term memories (Craik, 2002; Craik & Lockhart, 1972). Research has often focused on how positive emotions specifically relate to learning about rewards. For example, individual differences in anhedonia (characterized by low positive affect and the inability to experience pleasure) are associated with a variety of deficits in reward processes (Borsini et al., 2020; Huys et al., 2013; Slaney et al., 2022).

However, reward processes are also relevant to fear learning. For instance, studies of counterconditioning have shown that associations with a positive US can transform aversive CS-US relationships into neutral or positive ones (Keller et al., 2020). Relatedly, the presence of competing rewards has been shown to reduce avoidance behaviors despite having no observed

8

impact on fear itself (Pittig, 2019; Pittig & Dehler, 2019). Unfortunately, studies of fear acquisition and extinction are lacking among individuals with depression or depression-specific symptoms (e.g., anhedonia), particularly in comparing these individuals to those with a comorbid or independent anxiety disorders (Pittig et al., 2018). This is particularly notable, considering the impact of stress and threats in reducing sensitivity to rewards (Berghorst et al., 2013; Bogdan & Pizzagalli, 2006; Kumar et al., 2014).

Importantly, individual differences in positive affect are known to moderate fear learning processes, such that high positive affect reduces reacquisition (Zbozinek & Craske, 2017a) and reinstatement (Zbozinek et al., 2015) of fear following extinction learning. Indeed, given the numerous influences of positive affect on fear learning, interventions to increase positive affect have been suggested as potential strategies for enhancing exposure therapy treatment outcomes (Craske et al., 2016; Zbozinek & Craske, 2017b). In support of this notion, positive affect has also been shown to moderate the relationship between chronic stress and social anxiety disorder (Sewart et al., 2019), whereas anhedonia has been shown to predict poorer treatment outcomes for individuals diagnosed with social anxiety disorder (Craske et al., 2014). Thus, despite the apparent roles of reward sensitivity and positive affect during fear learning, more research is needed to evaluate the precise role of anhedonia in this process.

Neurocircuitry of Anhedonia in Relation to Fear Extinction

Of note, efforts to differentiate anxiety and depression have highlighted distinct neural pathways supporting threat and reward processing (Dillon et al., 2013). Prior studies of anhedonia have tended to focus on reductions in sensitivity to reward (i.e., reward consumption, or "liking") and motivation to pursue rewards (i.e., reward motivation, or "wanting"), with associated brain regions including the ventral striatum, amygdala, and hippocampus, medial

PFC, vmPFC, orbitofrontal cortex, and ACC (Berridge & Robinson, 2003; Treadway & Zald, 2011).

However, the neural correlates of reward sensitivity are known to interact with threats in one's environment. For example, fear extinction has been shown to rely on dopaminergic "reward" pathways during prediction-error signaling (i.e., following the non-occurrence of an expected aversive US) (Salinas-Hernández et al., 2019; Papalini et al., 2020), and this signaling has been shown to support the long-term consolidation and retrieval of fear extinction memories (Kalish et al., 2019). Early work has directly evaluated associations between Anhedonia-Apprehension and brain activation during fear extinction, highlighting several brain regions including the insular cortex, dorsal ACC, and amygdala (Young et al., 2021). Likewise, in a study pairing monetary incentives with a CS of varying threat, shifting from approach to avoidance of the CS appeared to correspond with changing activation in the dorsal ACC, dorsomedial PFC, vmPFC and dorsolateral PFC (Schlund et al., 2016). Although neural studies of counterconditioning are lacking in human subjects, rodent studies have implicated activity in regions such as the habenula, thalamus, amygdala, insular cortex, hippocampus, and nucleus accumbens (Keller et al., 2020). Collectively, these studies suggest that, despite notable differences in the neurocircuitry of threat and reward systems, the neural correlates of anhedonia are nonetheless related to processes relevant for fear extinction.

Summary

In sum, Pavlovian experiments are useful paradigms for understanding the processes that underlie the development, maintenance, and treatment of anxiety disorders. The existing literature has emphasized factors specific to anxiety disorders that contribute to disruptions in

fear extinction, and these studies have had a substantial influence on the creation and advancement of psychological treatments.

However, anxiety disorders are not a monolithic entity – rather, they are a heterogeneous category of diagnoses that often co-occur with other conditions, particularly depression and associated symptom dimensions. Existing studies have implicated positive affect as a moderator of fear extinction processes, suggesting that processes specific to depression may be relevant to disruptions in Pavlovian learning. Anhedonia symptoms are also associated with brain regions and networks that are recruited during fear acquisition and extinction. However, the complex relationships among anhedonia, extinction learning, and neural functioning remain underexamined. Such relationships may have important implications regarding the successful treatment of anxiety disorders.

This dissertation introduces three studies of anhedonia and fear extinction aiming to encompass basic, translational, and clinical research. Study 1 involves a data-driven approach to functional neuroimaging in order to identify patterns of brain activity that are associated with Anhedonia-Apprehension, over and above Fears and General Distress, during fear extinction and fear extinction recall tasks. Study 2 builds upon Study 1 to examine if anhedonia-specific patterns of brain activity further relate to behavioral and physiological patterns characteristic of aberrant extinction learning. Study 3 brings these questions into a clinical context, evaluating the extent to which anhedonia and its biological correlates are associated with aberrant prediction error during exposure therapy for social anxiety disorder.

STUDY 1

A MULTIVOXEL PATTERN ANALYSIS OF ANHEDONIA DURING FEAR EXTINCTION
– IMPLICATIONS FOR SAFETY LEARNING

From: Rosenberg, B.M., Taschereau-Dumouchel, V., Young, K.S., Lau, H., Zinbarg, R.E., Nusslock, R., & Craske, M.G. (2022). A multivoxel pattern analysis of anhedonia during fear extinction – implications for safety learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.

INTRODUCTION

Anhedonia, the loss of interest or pleasure in activities, is a symptom dimension commonly associated with major depression but also relevant to anxiety disorders. Extant research has focused largely on reward-related processes in relation to anhedonia, such as reductions in sensitivity to reward (i.e., reward consumption, or "liking") (Berridge & Robinson, 2003; Treadway & Zald, 2011; Huys et al., 2013; Thomsen, 2015), motivation to pursue rewards (i.e., reward anticipation, or "wanting"), and dopaminergic prediction error signaling associated with impairments in capacity to update behavior following reinforcement learning (Kumar et al., 2008; Pizzagalli et al., 2008; Gradin et al., 2011; Pizzagalli, 2014). The current study extends beyond reward processes to address threat-related processes in relation to anhedonia.

A common paradigm for measuring threat-related processes is Pavlovian fear learning. Activation of the insular cortex, dorsal anterior cingulate cortex (dACC), amygdala (Young et al., 2021) and other regions such as the ventromedial prefrontal cortex (vmPFC), has been consistently highlighted in neuroimaging studies of fear learning (Etkin & Wager, 2007; Sehlmeyer et al., 2009; Milad et al., 2014; Fullana et al., 2016; Greco & Liberzon, 2016; Battaglia et al., 2020), although there are inconsistencies regarding the precise role of the amygdala in human studies (Fullana et al., 2016; Young et al., 2021). Behavioral and neural

aberrations in Pavlovian fear acquisition and particularly fear extinction have been observed in individuals at risk for and with anxiety disorders, including perturbations in the insular cortex, dACC, amygdala, and vmPFC (Graham & Milad, 2011; Craske et al., 2017; Craske et al., 2018). Beyond the typical threat neurocircuitry, fear extinction has been shown to rely on dopaminergic "reward" pathways in 1) signaling the unexpected omission of an aversive unconditional stimulus (Salinas-Hernández et al., 2018; Papalini et al., 2020), or relief, which may itself be considered a type of reward (Carver, 2009), and 2) supporting the long-term consolidation of extinction memories (Kalisch et al., 2019). Since anhedonia has been associated with reductions in dopaminergic prediction error signaling, classically evaluated within reward learning paradigms (Huys et al., 2013; Rizvi et al., 2016), there is reason to hypothesize that anhedonia influences reward pathways involved in fear extinction. In partial support, behavioral studies show an association between low positive affect (a central feature of anhedonia) and less stable long-term fear extinction, as measured by stronger reacquisition (Zbozinek & Craske, 2017a) and reinstatement (Zbozinek et al., 2015) of conditioned fear.

We previously found more direct support for the role of anhedonia in neural responses during fear extinction (Young et al., 2021). Specifically, we utilized a dimensional model of symptoms of anxiety and depression (tri-level model) within a regions-of-interest analytical framework and found that the dimension of Anhedonia-Apprehension, but not dimensions of General Distress or Fears, was associated with increased activation of several brain regions during extinction learning, including the insular cortex, dorsal anterior cingulate cortex (dACC), and amygdala (Young et al., 2021). Notably, these regions overlap with the Salience network, where aberrations as a function of anhedonia have been found in studies of reward consumption, anticipation, and decision making (Treadway & Zald, 2011; Zhang et al., 2013).

Neural processes associated with anhedonia extend beyond the Salience network to regions of the Limbic (e.g., ventral striatum and hippocampus) and Cognitive Control (e.g., orbitofrontal cortex, dorsolateral prefrontal cortex) networks (Drevets, 2007; Zhang et al., 2013). Regions of the Default-Mode network (e.g., medial prefrontal cortex, posterior cingulate cortex) are thought to play a central role in the self-referential processes characteristic of depressive disorders (Sheline et al., 2009; Hamilton et al., 2015; Zhou et al., 2020) and may relate to anhedonia symptoms, as well. Given the wide range of brain systems associated with anhedonia, it is conceivable that the influences of anhedonia upon fear extinction extend beyond regions of the traditional "fear network." The present study built upon the prior study (Young et al., 2021) by 1) analyzing patterns of brain activity during fear extinction and recall to *predict* individual differences in anhedonia, and 2) addressing the breadth of brain activity associated with anhedonia during fear learning and specifically fear extinction, beyond the "fear network."

One approach to neuroimaging data, multivoxel pattern analysis (MVPA), is particularly well-suited to research in novel areas and may aid efforts to uncover associations between anhedonia and fear extinction. MVPA uses machine learning to "decode" patterns of brain activity that are consistently associated with a specific psychological construct. Unique strengths of this approach include its 1) emphasis on distributed patterns of brain activity rather than evaluating individual brain areas separately, 2) ability to directly test these patterns by predicting symptoms in an external validation sample, and 3) flexibility to detect unexpected associations by combining the predictive strengths of different features (which individually may not be strong enough to reach significance). MVPA has been widely employed in clinical neuroscience (Dwyer et al., 2018; Zhang et al., 2020), including studies identifying patterns of brain activity associated with disruptions in Pavlovian fear learning (Hennings et al., 2020), indices of

14

subjective fear and physiological arousal (Taschereau-Dumouchel et al., 2020), and anxious compared with non-anxious subjects during Pavlovian fear learning (Wen et al., 2021).

The present study aimed to uncover patterns of brain activity associated with anhedonia by decoding anhedonia symptoms using extinction (n=254) and extinction recall (n=249) fMRI data collected across two study sites. These tasks were selected due to prior evidence of neural associations with anhedonia within this dataset (Young et al., 2021), as well as known associations between anhedonia and prediction error signaling, a process central to the extinction of learned fear. We hypothesized that the decoders would train successfully during both task phases, and that the decoded patterns of brain activity would generalize to an external validation sample (i.e., data that were not included decoder training). We further hypothesized that successful decoders would be specific to Anhedonia-Apprehension, over and above other Tri-Level transdiagnostic symptom factors (i.e., General Distress or Fears). Exploratory analyses repeated the decoding approach by training and validating the decoder 1) between study sites (i.e., training within data from one site and generalizing to the other), as well as 2) within individual brain networks, highlighting candidate brain circuits that may be central to decoder results and warrant further research.

METHODS AND MATERIALS:

Participants

As described previously (Young et al., 2021), participants were recruited for the Brain, Motivation and Personality Development (BrainMAPD) study at the University of California, Los Angeles and Northwestern University, which investigated depression and anxiety in late adolescence and early adulthood. Participants were 272 individuals aged 18-19 years (182 female, mean age=19.16 years, SD=0.52). Recruitment was based upon self-reported scores of

15

trait Neuroticism (Eysenck & Eysenck, 1975), and Reward Sensitivity (Carver & White, 1994).

Oversampling on these dimensions was employed to ensure that the sample included individuals

at risk for the onset of depression and anxiety (see Supplement). Exclusion criteria were: lack of

right-handed dominance, not fluent in English, traumatic brain injury, MRI contraindications,

pregnancy, color blindness, lifetime psychotic symptoms, bipolar I disorder, clinically significant

substance use disorder in the past 6 months, and antipsychotic medication usage.

Of this group n=254 (UCLA=116, NU=138) are included for fear extinction and n=249

(UCLA=116, NU=133) are included for extinction recall (see Supplement for exclusion details).

Of the 254 participants, 250 participants completed SCID-5 interviews, of whom 79 participants

(31.60%) met criteria for a current anxiety disorder, but no depressive disorder; 19 (7.60%) met

criteria for current anxiety and depressive disorders; and three (1.20%) met criteria for a

depressive disorder but no anxiety disorder. 20 participants (8.00%) reported current use of at

least one psychotropic medication (see Supplement for details). All participants provided written,

informed consent. Participant demographics and tri-level symptoms are summarized in Table 1.

Tri-level measures of general distress, fear and anhedonia-apprehension

Immediately prior to MRI scans, participants completed questionnaire measures of

anxiety and depression to generate hierarchical Tri-Level model factor scores for General

Distress, Fear, and Anhedonia-Apprehension (see Supplement for details).

Fear Acquisition, Extinction, Extinction Recall

The two-day procedure for fear acquisition, extinction, and extinction recall was based on

the validated paradigm developed by Milad and colleagues (Milad, Quirk, et al., 2007; Milad,

Wright, et al., 2007). As described previously (Young et al., 2021), this slow event-related fMRI

paradigm consisted of four phases: habituation, acquisition, extinction (all conducted on day one)

16

and extinction recall (conducted on day two, 1-7 days later) (see Supplement for details). Images were office or conference rooms (context) with different colored lights (red/yellow/blue) as CS stimuli (color order and context images were counterbalanced across participants). During all task phases, inter-trial intervals varied from 12-18sec (mean 15sec) and included a jitter of 125ms per trial to reduce slice timing bias. The task was programmed in E-Prime (version 2.0 SP1) and presented to participants using a mirror and projector system.

fMRI Acquisition and Analysis

We used identical Siemens Prisma 3 Tesla MRI scanners at the UCLA Ahmanson-Lovelace Brain Mapping Center and the Northwestern University Center for Translational Imaging. High resolution structural (T1-weighted) images and blood oxygenation level-dependent (BOLD, T2*-weighted) functional images were acquired and preprocessing procedures applied (see Supplement for details).

As has been done in prior fMRI studies of fear extinction and extinction recall (Wen et al., 2021; Marin et al., 2017; Marin et al., 2020), analyses specifically focused on the end of fear extinction (i.e., the final four trials CS+E minus the final four trials CS-) and the beginning of extinction recall (i.e., the first four trials CS+E minus the first four trials CS-). Functional images were masked using a standard MNI template (Mazziotta et al., 2001). MVPA was implemented in the SciKit-Learn toolbox (Pedregosa et al., 2011) using the ElasticNetCV function (see Supplement and Table S1 for parameters evaluated during the training stage). Subjects were randomized into the training and testing datasets, yielding a training sample of n=127 subjects (UCLA=58, NU=69) and a testing sample of n=127 subjects (UCLA=58, NU=69) for the extinction task. Five subjects were not included in the extinction recall analysis, yielding a training sample of n=127 subjects (UCLA=58, NU=69) and a testing sample of n=122 subjects

(UCLA=58, NU=64). Within the testing sample, whole-brain decoders yielded brain-predicted Anhedonia-Apprehension values for each participant. The coefficient of determination (R^2) and correlation coefficient (r) were calculated between Anhedonia-Apprehension and brain-predicted Anhedonia-Apprehension values to determine successful prediction of scores in the external validation sample. To determine the R^2 cutoff score corresponding with statistical significance (*p*<.05), Anhedonia-Apprehension scores were permuted 10,000 times and R^2 was computed for each permutation.

Motion Outliers

To account for confounds due to motion, analyses tested the association between the percent of fMRI volumes censored due to motion (see Supplement) and Anhedonia-Apprehension as well as brain-predicted Anhedonia-Apprehension values in the training sample (covarying for Fears, General Distress, and Site). The percent of volumes censored due to motion was also included as a covariate in specificity analyses (see below).

Specificity Analysis: Associations over and above General Distress, Fears, Site, and Motion

Within the external validation sample, the correlation coefficient (r) was first computed separately between brain-predicted Anhedonia-Apprehension and each of the Tri-Level factors (Anhedonia-Apprehension, Fears, and General Distress), covarying for Site and Motion. The correlation coefficient was then calculated between brain-predicted Anhedonia-Apprehension and Anhedonia-Apprehension, covarying for Fears, General Distress, Site, and Motion.

Exploratory Analyses: Network-by-Network Effects

To explore localization of the decoder effects, we used the brain atlas developed by Schaefer and colleagues to divide the brain into 100 parcels, grouped into seven functional brain networks (Cognitive Control, Dorsal Attention, Default-Mode, Limbic, Salience, Somatomotor,

Visual) (Yeo et al., 2011; Schaefer et al., 2018). We then re-ran the decoding procedure seven times, masking within each network. Among significant networks, we additionally re-ran the decoding procedure masking within each individual region of the network.

Exploratory Analyses: Testing Between-Site Decoding of Anhedonia-Apprehension

To explore the robustness of decoder effects, we re-ran significant decoders utilizing a between-sites external validation approach (see Supplement). Decoder training was completed within the NU cohort (n=138) and validation was completed within the UCLA cohort (n=116). R^2 and r were calculated between Anhedonia-Apprehension and brain-predicted Anhedonia-Apprehension values to determine successful prediction of data within the external validation sample. We then re-ran the decoding procedure seven times, masking within each network, as described above.

RESULTS

Whole-Brain Decoder Effects

Permutation testing yielded a significance cutoff of R^2=.0186 for external validation (corresponding with two-tailed *p*<.05). Initial training of the whole-brain decoder during fear extinction was successful, such that the decoder predicted Anhedonia-Apprehension values (R^2=.168). The extinction decoder significantly predicted Anhedonia-Apprehension values in the external validation sample (R^2=.047; r=.276, *p*=.002) (Fig. 1).

Initial training of the whole-brain decoder during extinction recall was successful, such that the decoder predicted Anhedonia-Apprehension values (R^2=.336). However, the extinction recall decoder did not significantly predict Anhedonia-Apprehension values in the external validation sample (R^2<.001, r=-.063, *p*=.492). Therefore, the extinction recall decoder was not evaluated in subsequent analyses.

Motion Outliers

Covarying for Fears, General Distress, and Site, there was a significant association between percent of volumes censored and Anhedonia-Apprehension (t(249)=-2.059, *p*=.041, r=0.130), such that individuals with greater anhedonia tended to exhibit less movement in the MRI scanner. Covarying for Fears, General Distress, and Site, there was no association between percent of volumes censored and brain-predicted Anhedonia-Apprehension values within the training sample (t(122)=-1.277, *p*=.204, r=0.115) or the testing sample (t(122)=0.016, *p*=.987, r=.002).

Specificity Analysis: Associations over and above General Distress, Fears, Site, and Motion

Covarying for Site and Motion, brain-predicted Anhedonia-Apprehension was significantly associated with Anhedonia-Apprehension (t(123)=3.192, *p*=.002, r=0.274), but not Fears (t(123)=0.660, *p*=.511, r=0.060) or General Distress (t(123)=0.338, *p*=.736, r=0.031). Covarying for Fears, General Distress, Site, and Motion, brain-predicted Anhedonia-Apprehension was significantly associated with Anhedonia-Apprehension (t(121)=3.209, *p*=.002, r=0.277).

Exploratory Analyses: Network-by-Network Effects

Exploratory analyses demonstrated that the following network-masked decoders significantly predicted Anhedonia-Apprehension within the external validation sample: Cognitive Control (R^2=.020; r=.245, *p*=.006), Default-Mode (R^2=.040; r=.263, *p*=.003), Limbic (R^2=.029; r=.217, *p*=.014), and Visual (R^2=.022; r=.181, *p*=.042) (see Fig. 2, Fig. S1, and Table 2 for details). The Salience decoder met criteria for significance using r, but not R^2, as the metric of external validation (R^2=.004; r=.229, *p*=.010). Exploratory analyses within these networks

further demonstrated that several region-specific decoders could significantly predict Anhedonia-Apprehension (see Table 3 for details).

Exploratory Analyses: Decoding of Anhedonia-Apprehension with a Between-Sites Approach

Initial training of the whole-brain decoder during fear extinction was successful (R^2=.420; see Supplement for details). The between-sites decoder met criteria for significance using r, but not R^2, as the metric of external validation (R^2=-.024; r=.190, *p*=.040). Specificity analyses revealed that, utilizing the between-sites approach, the Dorsal Attention (R^2=-.025; r=.186, *p*=.045) and Visual (R^2=-.015; r=.236, *p*=.011) network-decoders met criteria for significance using r, but not R^2, as the metric of external validation (see Supplement for details).

DISCUSSION:

The present study utilized MVPA to characterize unique patterns of functional brain activity during fear extinction and extinction recall associated with anhedonia symptoms. We found anhedonia-specific whole-brain patterns of functional activity during fear extinction that generalized to an external validation sample. Importantly, these patterns were significantly associated with the dimension of Anhedonia-Apprehension over and above other symptom dimensions of General Distress and Fears.

Within individual networks and regions, the patterns of activity appeared complex. Although plotting the decoder weights can aid in the interpretation of which regions and networks are implicated in the whole-brain decoder, these results should be interpreted with caution. For example, high beta weights could indicate voxels that cancel out noise, rather than increased activation. Similarly, if two voxels provide an equivalent amount of information, the decoder may arbitrarily select one voxel and omit the other (for additional information on interpreting decoder results, see Kriegeskorte et al., 2019). Hence, exploration of specific brain

21

networks implicated in the anhedonia decoder is highly tentative. With that caveat in mind, we identified activity within the Cognitive Control, Default-Mode, Limbic, Salience, and Visual networks that generalized across the training and external validation samples.

The Anhedonia-Apprehension decoder appeared to involve predominantly positive beta weights among regions of the Salience network. Regions of the Salience network, such as the insular cortex, dACC, and amygdala, overlap with the "fear network" (Fullana et al., 2016; Young et al., 2021). One interpretation of heightened activation in the Salience network is persistent attentional salience of extinguished stimuli, perhaps representing strength of CS-US associations (i.e., weakened extinction). Additional research is needed to explore this possibility particularly considering the limited interpretability of directional findings in MVPA analyses. Likewise, it has been suggested that the Salience network integrates information from both the Default-Mode and Cognitive Control networks in directing external and internal attention (Sridharan et al., 2008; Menon & Uddin, 2010; Goulden et al., 2014), and that this process is altered in major depression (Hamilton et al., 2011; Kaiser et al., 2015). Additional research may explore the extent to which persistent activation among regions of the Cognitive Control or Default-Mode networks, in coordination with the Salience network, relates to deficits in extinction associated with anhedonia.

Another potential pattern was for the Anhedonia-Apprehension decoder to involve predominantly heightened activity within the Cognitive Control network. Prior studies of major depression have highlighted aberrant activity in this network, particularly the dorsolateral prefrontal cortex (DLPFC), in studies of attentional bias and emotion regulation (Gotlib & Hamilton, 2008; Koenigs & Grafman, 2009; Disner et al., 2011). Heightened DLPFC activation has been implicated as compensation for diminished reward processing as a function of

depression (Zhang et al., 2013) and transdiagnostically (Brolsma et al., 2021), and thus offers another pathway for the relationship between anhedonia and extinction.

Furthermore, the Anhedonia-Apprehension decoder appeared to involve predominantly positive beta weights among regions of the Default-Mode network. Hyperactivation and hyperconnectivity of the Default-Mode network have been implicated in studies of depression, particularly during unconstrained rest or during tasks involving internally directed attention, such as autobiographical memory and rumination (Sheline et al., 2009; Hamilton et al., 2015; Kaiser et al., 2015; Zhou et al., 2020). The present results support the potential applicability of the Default-Mode network within studies of anhedonia or extinction learning, although additional research is needed to elucidate these associations more precisely.

Given the predominantly positive beta weights evident across the Anhedonia-Apprehension decoder, it is possible that the findings of this study represent hyperactivity among specific brain systems, such as the Salience, Default-Mode, and Cognitive Control networks, as noted above. Such an interpretation is supported by the finding that these networks were not only implicated in the whole-brain decoder, but each decoded Anhedonia-Apprehension separately. However, negative beta weights were also evident across these systems and may constrain this interpretation. In addition, it is further possible that Anhedonia-Apprehension is positively associated with non-specific patterns of brain activity during fear extinction, and that these non-specific patterns happen to encompass regions of the Salience, Default-Mode, and Cognitive Control networks. Such an interpretation would imply that Anhedonia-Apprehension is associated with generally increased activity across the brain during fear extinction, and that network-by-network interpretations have limited utility. Additional research is needed to evaluate the extent to which Anhedonia-Apprehension is associated with elevated brain activity

during extinction, and if broad differences in brain activity may account for patterns in the specific regions and networks evaluated in this study.

The current findings highlight the role of anhedonia in relation to fear learning constructs that have been traditionally considered primarily within the context of anxiety disorders. Although anhedonia has been considered mostly within the context of depression, it is transdiagnostic and associated with several anxiety disorders including social anxiety (Kashdan, 2007), obsessive-compulsive disorder (Abramovitch et al., 2014), and posttraumatic stress (Nawjin et al., 2015). Greater recognition of the role of anhedonia in anxiety disorders and fear learning processes is consistent with dimensional models of psychopathology that cut across conventional diagnostic categories (Brown & Barlow, 2009; Prenoveau et al., 2010; Insel, 2014; Thompson-Hollands et al., 2014; Naragon-Gainey et al., 2016; Kotov et al., 2017; Kotov et al., 2018; Krueger et al., 2018) and have direct implications for optimal care (Ruggero et al., 2019; Hopwood et al., 2020). Pending replication of the current findings, the role of anhedonia in fear learning could be leveraged in the development of personalized, process-targeted treatments. For example, studies of fear extinction have provided a foundation for contemporary models of exposure therapy (Pittig et al., 2016; Scheveneels et al., 2016; Fullana et al., 2020), which emphasize prediction error (and other features, such as contextual modulation) for optimizing exposure therapy effectiveness (Craske et al., 2008; Craske et al., 2012; Craske et al., 2014; Craske et al., 2018). Given the potential interference with prediction error posed by anhedonia, novel exposure protocols may incorporate interventions to increase positive affect already shown to augment extinction (Craske et al., 2016; Craske et al., 2019), for anxious individuals with anhedonic symptoms.

Additionally, neuromodulation targeting the control, default-mode, or salience networks may augment exposure therapy for individuals with anhedonia. For example, preliminary studies of transcranial magnetic stimulation have targeted the dorsolateral prefrontal cortex to augment the effects of exposure therapy for posttraumatic stress disorder (Osuch et al., 2009; Karsen et al., 2014; Fryml et al., 2019). Combining brain stimulation and exposure therapy may prove particularly useful for anxious patients who present with elevated anhedonia or a comorbid depressive disorder.

Decoder cross-validation was associated with relatively small coefficient of determination (R^2) values in the present study. Despite the potential advantages of R^2 in prediction studies (Poldrack et al., 2020), the application of R^2 in MVPA studies may also be limited due to scaling issues. For example, the present study collected data at two different fMRI scanners, which could affect the R^2 metric. For this reason, we have also reported correlation coefficients (r), which are relatively independent of the scale used and tended to indicate stronger associations between anhedonia and brain activation.

The current study involves several strengths, namely the 1) comparatively large sample size of both the training and external validation datasets, 2) emphasis on dimensional psychopathology, 3) test of effects in an external validation sample, and 4) exploration of effects using a between-sites approach. The narrow age range of participants could also be considered a strength of the present study, as the reported effects are unlikely to be explained by variations in the age of participants. However, the narrow age range may also reduce generalizability of these results to other developmental stages. Additional limitations include the 1) small number of experimental trials analyzed in the extinction and extinction recall tasks, 2) exploratory nature of network- and region-specific analyses, including some cases in which r but not R^2 met criteria

for statistical significance, 3) comparatively small size of decoder prediction values, and 4) tentative interpretability of directional results. Future research is needed to replicate the findings of the present study, to explore avenues for strengthening decoder predictions, and to evaluate the directionality of network- and region-specific results.

In sum, the present study suggests that patterns of brain activity during extinction learning are predictive of anhedonia symptoms. Extinction is a fear learning process traditionally considered in relation to anxiety symptoms but rarely in relation to transdiagnostic symptom dimensions, such as anhedonia. The patterns of brain activity identified in this study may be characteristic of anhedonia-specific deficits during fear learning and warrant additional research.

TABLES AND FIGURES:

Table 1. Demographics and Symptom Dimensions

Characteristics	UCLA (n = 116)	Northwestern (n = 138)	Statistic	p Value
Gender, n (%)			$\chi^2_2 = 0.84$	.656
Female, cisgender	78 (67.24%)	92 (66.67%)		
Male, cisgender	38 (32.76%)	45 (32.61%)		
Male, transgender	0 (0.0%)	1 (0.72%)		
Age, Years, Mean (SD)	19.09 (0.52)	19.25 (0.52)	$t_{253} = 2.46^a$	.014
Ethnicity, n (%)			$\chi^2_1 = 1.23$	.268
Not Hispanic/Latino	82 (70.69%)	106 (76.81%)		
Hispanic/Latino	34 (29.31%)	32 (23.19%)		
Race, n (%)			$\chi^2_5 = 20.67^a$	.001
Asian	45 (38.79%)	27 (19.57%)		
Black	5 (4.31%)	14 (10.14%)		
Multiracial	3 (2.59%)	17 (12.32%)		
Native American	1 (0.86%)	3 (2.17%)		
White	61 (52.59%)	77 (55.80%)		
Declined to report	1 (0.86%)	0 (0.0%)		
Current Psychotropic Medication Use, n (%)	2 (1.72%)	18 (13.04%)	$\chi^2_1 = 11.13^a$	.001
Symptom Dimension Scores, Mean (SD)				
General distress	−0.032 (0.94)	0.128 (0.89)	$t_{253} = 1.39$	.17
Fears	0.080 (0.93)	−0.115 (0.79)	$t_{253} = 1.81$	.07
Anhedonia-apprehension	0.112 (0.84)	−0.094 (0.94)	$t_{253} = 1.82$	.07

Demographic factors and symptom dimension scores of participants compared across scanning site. The racial identity of individuals across sites was significantly different, with a higher proportion of Asian participants at UCLA and a higher proportion of Black and Multiracial participants at Northwestern University.

UCLA, University of California Los Angeles.

[a]Denotes statistical significance ($p < .05$).

Table 2. Network-by-Network Decoder Results

Network Mask	Decoder Training	Testing in External Validation Sample		
	R^2	R^2	r	p
Cognitive Control	0.045	0.020[a]	0.245[a]	.006
Default Mode	0.088	0.040[a]	0.263[a]	.003
Dorsal Attention	0.271	0.008	0.140	.116
Limbic	0.183	0.029[a]	0.217[a]	.014
Salience	0.020	0.004	0.229[a]	.010
Somatomotor	Failed training	N/A	N/A	N/A
Visual	0.155	0.022[a]	0.181[a]	.042

Exploratory network-by-network decoder results implicated in the whole-brain decoder.

N/A, not applicable.

[a]Denotes statistical significance ($R^2 > 0.0186$ or $p < .05$).

Table 3. Region-by-Region Decoder Results

Regional Masks	Schaefer Atlas ROI Index	Testing in External Validation Sample		
		R^2	r	p
Cognitive Control Network				
L dorsolateral PFC, lateral	35	0.016	0.267[a]	.002
R dorsolateral PFC, anterior	83	0.016	0.215[a]	.015
R dorsolateral PFC, lateral	84	0.017	0.249[a]	.005
R dorsolateral PFC, dorsal	85	0.026[a]	0.227[a]	.010
R frontal eye field	86	0.034[a]	0.250[a]	.005
R medial posterior PFC/frontal eye field	88	0.018	0.222[a]	.012
Default Mode Network				
L medial temporal gyrus	38	0.013	0.177[a]	.046
L medial temporal gyrus	39	0.057[a]	0.299[a]	.001
L angular gyrus	41	0.028[a]	0.210[a]	.018
L pars orbitalis	42	0.014	0.290[a]	.001
L pars orbitalis	43	0.044[a]	0.259[a]	.003
L dorsal anterior cingulate cortex	44	0.028[a]	0.215[a]	.015
L anterior PFC	45	0.028[a]	0.293[a]	.001
L dorsolateral PFC, dorsal	46	0.025[a]	0.200[a]	.024
L premotor/supplementary motor area	47	0.023[a]	0.224[a]	.011
L frontal eye field	48	0.039[a]	0.295[a]	.001
L ventral posterior cingulate cortex	49	-0.007	0.225[a]	.011
L ventral posterior cingulate cortex	50	0.021[a]	0.229[a]	.010
R superior temporal gyrus	93	0.011	0.237[a]	.007
R pars orbitalis	94	0.026[a]	0.285[a]	.001
R Broca's triangle	95	0.036[a]	0.274[a]	.002
R anterior PFC	96	0.032[a]	0.222[a]	.012
R dorsolateral PFC, dorsal	97	0.020[a]	0.178[a]	.045
R frontal eye field	98	0.047[a]	0.295[a]	.001
R ventral posterior cingulate cortex	100	0.001	0.194[a]	.029
Limbic Network				
L orbitofrontal cortex	31	0.014	0.217[a]	.014
L temporal pole	32	0.037[a]	0.256[a]	.004
R orbitofrontal cortex	79	0.010	0.211[a]	.018
Salience Network				
L insula/frontal operculum	25	-0.001	0.288[a]	.001
L insula/frontal operculum	26	0.001	0.225[a]	.011
L anterior lateral PFC	27	0.031[a]	0.269[a]	.002
L dorsal anterior cingulate cortex	28	0.024[a]	0.221[a]	.013
L premotor/supplementary motor area	30	0.007	0.191[a]	.032
R insula/frontal operculum	76	0.009	0.255[a]	.004
Visual Network				
L visual association area	2	0.018	0.175[a]	.049
L visual association area	3	0.021[a]	0.272[a]	.002
L visual association area	7	0.013	0.214[a]	.016
L visual association area	8	0.032[a]	0.209[a]	.018
R fusiform gyrus	51	0.022[a]	0.227[a]	.010
R fusiform gyrus	52	0.023[a]	0.194[a]	.029
R primary visual cortex	55	0.016	0.194[a]	.029
R visual association area	57	0.028[a]	0.197[a]	.026

Significant exploratory ROI-by-ROI decoder results within networks implicated in the whole-brain decoder.

L, left hemisphere, PFC, prefrontal cortex; R, right hemisphere; ROI, region of interest.

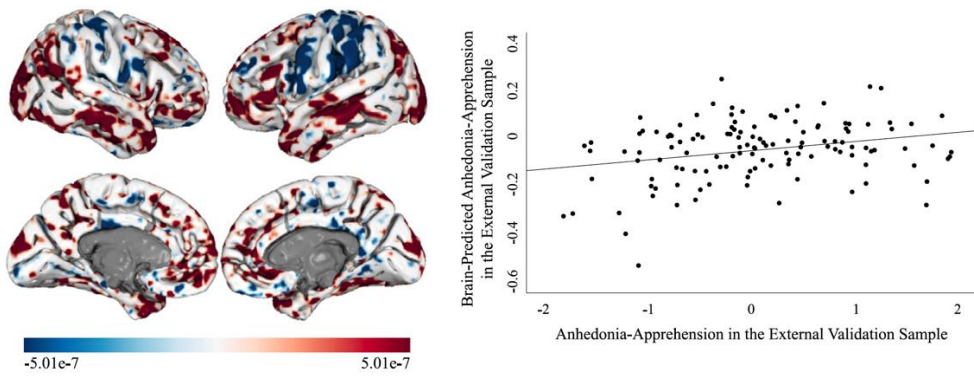[a]Denotes statistical significance ($R^2 > 0.0186$ or $p < .05$).

Figure 1. Whole-brain decoder results. Weights of the whole-brain decoder are presented for illustration purposes only. They should not be interpreted as indicating involvement of a specific brain region (as is the case in mass univariate analyses) (1). Predicted anhedonia apprehension values were significantly associated with anhedonia apprehension in the external validation sample.
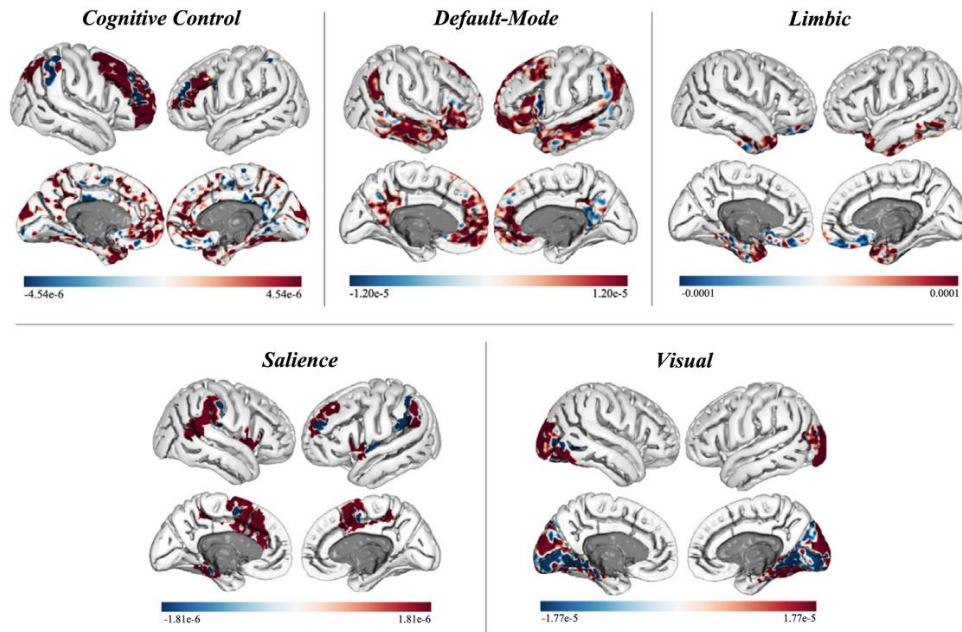

Figure 2. Network-by-network decoder results. Decoder plots for individual network decoders that were significantly associated with anhedonia-apprehension in the external validation sample.

SUPPLEMENT:

Participants

Participants were sampled such that those who fell in the high, mid, and low (i.e., tertiles) on each scale were represented in the study sample. Specifically, participants were oversampled from the two diagonals of the bivariate space defined by the EPQ-N and BAS scales, meaning that the sample comprised individuals with scores that were high on each scale, low on each, mid-range on each, or high on one measure and low on the other.

20 participants (8.20%) reported current use of at least one psychotropic medication, including the following: Adderall (4), Alprazolam (1), Amitriptyline (1), Atomoxetine (1), Citalopram (2), Escitalopram (1), Fluoxetine (3), Lamotrigine (3), Lisdexamfetamine (2), Methylphenidate (2), Nortriptyline (1), Sertraline (4).

Tri-Level Model: Anhedonia-Apprehension

Based on the original construction and replication of the tri-level model, these items consisted of selected questions from the Fear Survey Schedule-II (Geer, 1965), Albany Panic and Phobia Questionnaire (Rapee et al., 1994), Self-Consciousness subscale of the Social Phobia Scale (Zinbarg & Barlow, 1996; Mattick & Clarke, 1998), Inventory to Diagnose Depression (Zimmerman et al., 1986), Mood and Anxiety Symptom Questionnaire (Watson et al., 1995), Penn State Worry Questionnaire (Meyer et al., 1990), and Obsessive Compulsive-Inventory Revised (Foa et al., 2002).

The Anhedonia-Apprehension factor is largely driven by positive affect items (e.g., reverse-scored items such as: "felt like I was having a lot of fun", "felt really happy") which have a standardized loading average magnitude of .71 whereas the strongest standardized loading of an apprehension item (e.g., "feeling discouraged about the future", "feeling pessimistic about

the future") has a magnitude of only .28). The General Distress factor includes most questionnaire items and is most strongly driven by with depression and worry items. The Fears factor includes items assessing social anxiety, specific fears, obsessive-compulsive symptoms, anxious arousal/somatic sensations, and interoceptive/agoraphobia fears.

Fear Acquisition, Extinction, Extinction Recall

During habituation, participants viewed each of three CS images for four six-second trials, to reduce novelty. During acquisition, participants viewed images of two CS+ stimuli and one CS- stimulus. During each trial, participants first viewed the context image (3sec), followed by the CS embedded in the context (6sec). There were 8 trials of each CS+ (16 trials total) and 16 trials of the CS-. Using a 62.5% reinforcement rate, each CS+ was immediately followed by a mild electric shock applied to the left bicep on 5 out of 8 trials. During extinction, participants viewed 16 trials of one CS+ (the "extinguished" CS+ now termed the CS+E) and 16 trials of the CS-, none of which were followed by shock. During extinction recall, participants viewed 8 trials of the CS+E, 8 trials of the CS+ that was not presented during extinction (the "unextinguished" CS+ now termed the CS+U) and 16 trials of the CS-.

Unconditional Stimulus

Shocks were delivered using a DS7a constant current high voltage stimulator (Digitimer Ltd, England) at UCLA and a STMISOC constant voltage stimulator (Biopac Systems Inc, USA) at Northwestern. Shock levels were determined during a "work-up" procedure conducted on Day 1 before scanning. In this procedure, participants were presented with shocks of increasing intensity and were asked to rate each on a pain scale of 1-10 (1 = "not at all painful", 10 = "most pain imaginable"). Participants were informed we aimed to reach a level of shock that was

"uncomfortable but not painful" and "took some effort to tolerate" (i.e., a rating of 5-6 that they were willing to tolerate for the experiment).

fMRI Acquisition and Analysis

High resolution structural images (T1-weighted) were acquired using a magnetized prepared rapid acquisition gradient echo (MPRAGE) sequence using 0.8mm isotropic voxels, TR/TE/flip angle=2300ms/2.99ms/7°, FOV= 256mm2, 208 slices. Blood oxygenation level-dependent (BOLD, T2*-weighted) functional images were acquired parallel to the AC-PC line using Siemens AutoAlign function, using 2mm isotropic voxels, TR/TE/flip angle=2000ms/25ms/80°, FOV=208mm2, 64 slices, multiband acceleration factor=2, sequential slice acquisition, 380 volumes (per task phase).

Raw dicom files were converted to NIFTI format using dcm2nii (MRIcroN, http://www.cabiatl.com/mricro/mricron/dcm2nii.html). Data were processed and analyzed using FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Structural data was corrected for spatial intensity variations (bias field correction) using FAST (FMRIB's Automated Segmentation Tool) (Zhang et al., 2001) and brain extraction was performed using optiBET (optimized brain extraction) (Lutkenhoff et al., 2014).

Functional data was first assessed for outlier volumes (75th percentile +1.5 time interquartile range) based on framewise displacement (average of rotation and translation parameter differences, using weighted scaling 43 as implemented in the fslmotionoutliers function). Outlier volumes were censored in first level analyses by including a regressor with a single time-point corresponding to each outlying volume. Functional data were brain extracted using BET (Brain Extraction Tool) (Smith, 2002) and bias field corrected using N4BiasFieldCorrection, run twice (ANTS registration suite) (Avants et al., 2009).

Table S1. Reasons for participant exclusion across task phases.

	Extinction	Recall
Total n included	**254**	**249**
Reported falling asleep during Acquisition or Extinction	10	10
Technical failure	6	6
Incomplete Symptom Data	2	2
No scan or incomplete scan data	n/a	5

fMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00. Registration to high resolution structural space images was carried out using FLIRT (Jenkinson & Smith, 2001; Jenkinson et al., 2002). Registration from high resolution structural to standard space was then further refined using FNIRT (nonlinear registration) (Andersson et al., 2007). The following pre-statistics processing was applied: motion correction using MCFLIRT (Jeninson et al., 2002), slice-timing correction using Fourier-space time-series phase-shifting, spatial smoothing using a Gaussian kernel of FWHM 4.0mm, grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and high-pass temporal filtering (0.01Hz) to remove low frequency artifacts.

First-level analyses of neural activation included regressors of interest (context, CS+, CS- and shock) and temporal derivatives, six motion regressors and additional regressors to censor outlying volumes. Time-series statistical analysis was carried out using FILM with local autocorrelation correction (Smith et al., 2004).

Decoder Training Parameters

Several decoders were tested within the training dataset using 5-fold cross-validation in SciKit-Learn and varying the hyperparameters according to the table below. For both fear extinction and extinction recall, decoder training was superior when the L1/L2 ratio was highly

similar to ridge regression (i.e., L1/L2 Ratio = 0.00001 or L1/L2 Ratio = 0.0001 respectively). The tables below include statistics regarding the performance of various L1/L2 regularization methods during the training phase.

Table S2: Decoder training parameters using cross-validation with a held-out sample

Extinction			Extinction Recall		
Method	**L1/L2 Ratio**	**R^2**	**Method**	**L1/L2 Ratio**	**R^2**
Ridge	0	<0.001	Ridge	0	<0.001
Elastic Net	***0.00001***	***0.168***	Elastic Net	0.00001	0.273
Elastic Net	0.0001	0.121	***Elastic Net***	***0.0001***	***0.336***
Elastic Net	0.001	0.091	Elastic Net	0.001	0.189
Elastic Net	0.01	0.077	Elastic Net	0.01	0.046
Elastic Net	0.1	0.060	Elastic Net	0.1	0.013
Elastic Net	0.25	0.078	Elastic Net	0.25	0.004
Elastic Net	0.5	0.018	Elastic Net	0.5	<0.001
Elastic Net	0.75	0.024	Elastic Net	0.75	<0.001
Elastic Net	0.9	0.030	Elastic Net	0.9	<0.001
Elastic Net	0.99	<.0001	Elastic Net	0.99	<0.001
Elastic Net	0.999	<.0001	Elastic Net	0.999	<0.001
Elastic Net	0.9999	<.0001	Elastic Net	0.9999	<0.001
Lasso	1	<.0001	Lasso	1	<0.001

In exploratory analyses, the extinction decoder was also evaluated during the training stage when utilizing a between-sites approach (i.e., training data = NU, external validation = UCLA). Decoder training was superior when the L1/L2 ratio was similar to ridge regression (i.e., L1/L2 Ratio = 0.01).

Table S3: Decoder training parameters utilizing external validation approach

	Extinction	
Method	**L1/L2 Ratio**	**R^2**
Ridge	0	<0.001
Elastic Net	0.00001	0.190
Elastic Net	0.0001	0.360
Elastic Net	0.001	0.408
Elastic Net	***0.01***	***0.420***
Elastic Net	0.1	0.360
Elastic Net	0.25	0.350
Elastic Net	0.5	0.334
Elastic Net	0.75	0.346
Elastic Net	0.9	0.283
Elastic Net	0.99	0.251
Elastic Net	0.999	0.251
Elastic Net	0.9999	0.251
Lasso	1	0.251

In exploring the networks implicated in the between-sites decoding approach, network-predicted Anhedonia-Apprehension was not significantly associated with Anhedonia-Apprehension in the external validation sample.

Table S4: External Validation Decoder Results by Network

Network Mask	**Training Sample (Northwestern)**	**External Validation Sample (UCLA)**		
	R^2	R^2	r	p
Control	0.204	-0.074	0.074	0.429
Default-Mode	0.199	-.038	0.159	0.089
Dorsal Attention	0.200	-0.025	0.186*	0.045
Limbic	0.145	-0.039	0.143	0.125
Salience	0.258	-0.030	0.173	0.063
Somatomotor	0.345	-0.071	0.069	0.464
Visual	0.103	-0.015	0.236*	0.011

Network- and Region-Specific Masks

Fig. S1 – Individual network masks that were associated with anhedonia during specificity

analyses in the external validation sample (red = cognitive control, orange = default-mode, blue

= limbic, green = salience, yellow = visual).
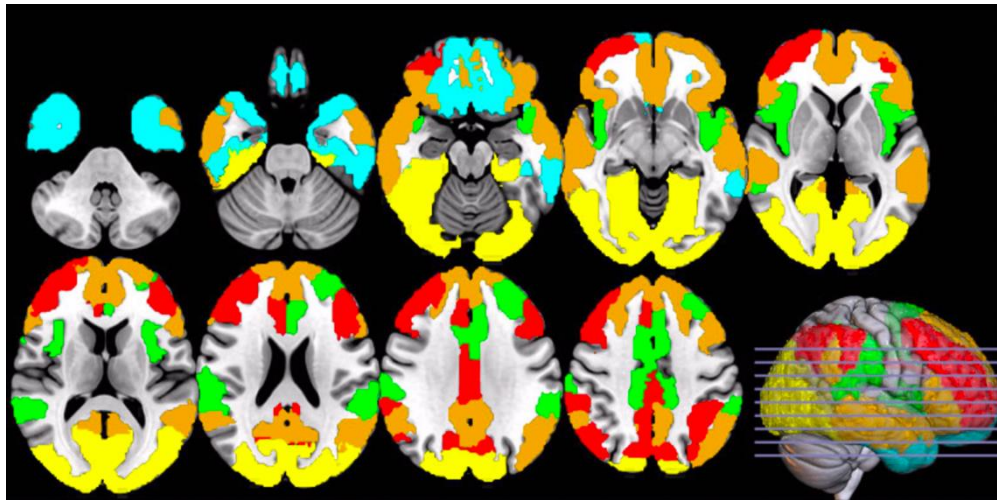
Coefficient of Determination (R²)

The coefficient of determination (R2) is computed according to the following formula:

$$R^2 = 1 - \frac{\text{Sum of Squares (residal)}}{\text{Sum of Squares (total)}}$$

Therefore, in some instances, it may be possible for a decoder to yield a negative R^2 value (i.e., if

Sum of Squares $_{residual}$ > Sum of Squares $_{total}$). Whereas a positive R^2 indicates that the model

performs *better* than chance level, a negative R^2 would indicate that the model performs *worse*

than chance level. For additional information, see (Poldrack et al., 2020).

STUDY 2: ANHEDONIA IS ASSOCIATED WITH LESS ACQUISITION AND

SUBSEQUENT OVER-GENERALIZATION OF NEGATIVE PREDICTIONS

INTRODUCTION

Pavlovian fear acquisition is a widely used paradigm in which an aversive stimulus (unconditional stimulus; US) is repeatedly paired with a neutral stimulus (conditional stimulus; CS). Individuals with anxiety disorders have been shown to exhibit heightened conditional physiological responses to CSs that predict a US (CS+) and to CSs that do not predict a US (CS-) during Pavlovian fear acquisition and extinction paradigms (Duits et al., 2015; Lissek et al., 2005). Anxious individuals have also been shown to exhibit less discrimination between stimuli according to their similarity to a CS, indicative of heightened fear generalization (Lissek et al., 2010; Lissek et al., 2014; Dunsmoor & Paz, 2015). Aberrations in Pavlovian learning, particularly fear acquisition and generalization, are thought to underlie the etiology and maintenance of anxiety disorders. Likewise, Pavlovian fear extinction is considered a central mechanism of exposure therapy (Craske et al., 2014), the gold-standard psychotherapy for anxiety disorders (Rauch et al., 2021; Abramowitz et al., 2019).

A growing body of research has highlighted a potential role for low positive affect, a core feature of anhedonia, in relation to Pavlovian fear learning. For example, reward processes, which tend to be dampened among anhedonic individuals, are thought to influence fear learning and related approach/avoidance decisions via a range of mechanisms including dopaminergic prediction error signaling (Papalini et al., 2020), relief-pleasantness (i.e., positive feeling when an expected negative event does not occur, see Willems & Vervliet, 2021), or competing incentives (i.e., when a negative outcome is also associated with a positive outcome (Pittig, 2019; Pittig & Dehler, 2019). In contrast, high positive affect has been shown to increase

attention toward approach-motivating positive stimuli (Gable & Harmon-Jones, 2008; Huntsinger et al., 2014) and is associated with the formation of stable, long-term extinction memories (Zbozinek et al., 2015; Zbozinek & Craske, 2017a). Nonetheless, despite the importance of reward processes and positive affect during Pavlovian learning, associations between anhedonia and Pavlovian fear learning remain understudied (Pittig et al., 2018).

Prior studies have evaluated the associations between depression and Pavlovian fear, but these studies have not specifically looked at anhedonia and have yielded mixed results. For example, depression has been associated with enhanced acquisition of conditional skin conductance responses during Pavlovian fear acquisition (Nissen et al., 2010), impaired learning of US probabilities during acquisition (Wurst et al., 2021), enhanced extinction of startle responses (Kuhn et al., 2014). Likewise, one study found no association between depression and impaired fear generalization (Wurst et al., 2021). Focusing analyses on anhedonia, a specific symptom dimension of depression, may clarify discrepant results and inform future studies in this area.

Studies of brain activity during fear learning processes may offer a unique lens into anhedonia and its relation to Pavlovian fear learning, with existing evidence tending to suggest that anhedonia is associated with aberrant activity of the neural circuits central to Pavlovian fear learning. In particular, studies of anhedonia have highlighted distinct patterns of brain activity during Pavlovian fear extinction, including activity in regions such as the ventromedial prefrontal cortex (vmPFC) and insula (Young et al., 2021; Rosenberg et al., 2022). Prior research has reliably demonstrated that activation of the vmPFC is associated with successful fear acquisition, extinction, and extinction recall (Milad et al., 2014; Fullana et al., 2016; Greco & Liberzon, 2016). Likewise, activity in the vmPFC and insula are associated with successful

discrimination of generalization stimuli (Dunsmoor et al., 2011; Greenberg et al., 2013a; Greenberg et al., 2013b; Lissek et al., 2014). Furthermore, although associations with anhedonia were not been specifically evaluated, less discriminant vmPFC responses have also been shown to correlate with trait levels of both anxiety and depression (Greenberg et al., 2013b). Given that neural, subjective, and physiological measures do not always concord (Lougheed et al., 2021), it is important to assess whether neural patterns associated with anhedonia translate to these other domains and explain discrepant results among depressed samples.

The aim of this study was to test if symptoms of anhedonia, or anhedonia-related patterns of brain activity, are associated with physiological and self-report indices of Pavlovian fear learning. We hypothesized that higher anhedonia would be associated with 1) greater conditional fear for the CS+ during fear acquisition, extinction, and extinction recall (measured by skin conductance response (SCR) and self-reported US contingency ratings), and 2) overgeneralization of fear for stimuli that resemble a CS+ (measured by SCR and self-reported US expectancy ratings). In addition to these hypotheses, to detect potential concordance with a neural index of anhedonia-apprehension, exploratory analyses evaluated these associations using anhedonia values that were predicted from brain activity during a fear extinction task (Rosenberg et al., 2022).

MATERIALS AND METHODS:

This study used participants from the same sample described by Rosenberg et al., 2022 (N=254) to generate measures of Anhedonia-Apprehension, Fears, and General Distress. The decoder described by Rosenberg et al. resulted in predicted Anhedonia-Apprehension values for each subject in the training sample (n=127). These values were used in separate analyses of

brain-predicted Anhedonia-Apprehension in relation to measures of Pavlovian fear learning (see below).

Fear Acquisition, Extinction, Extinction Recall

The two-day procedure for fear acquisition, extinction, and extinction recall was identical to the fMRI procedure described by Rosenberg et al., 2022.

Fear Generalization

The fear generalization procedure, performed at a prior experimental visit from fMRI scanning visits (mean = 13.91 days before), was based on the validated paradigm developed by Lissek and colleagues (Lissek et al., 2014). Briefly, participants underwent a fear acquisition procedure involving one CS+ and one CS- (a small ring and a large ring, counter-balanced) presented 12 times. The CS+ was paired with a US (electric shock) during nine trials for a reinforcement rate of 75%. Following acquisition, participants completed the generalization task, which incorporated eight generalization stimuli (rings of varying sizes between the CS+ and CS-; GS). The CS+ was presented four times and was paired with a US twice, for a reinforcement rate of 50%. The CS- was presented four times, and the eight GSs were each presented twice each. The CS- and GS stimuli were never paired with a US. During both the acquisition and generalization phases of the task, all stimuli were presented in a quasi-random order (i.e., no more than two trials of the same type in a row). All CSs and GSs were presented for 8.5 seconds with an intertrial interval of 17-22 seconds.

Self-Report Ratings

Following fear acquisition, extinction, and extinction recall, participants provided self-report ratings of US contingency for the unextinguished CS+ (CS+U), CS+E, and CS- on a 1-3 scale corresponding with "high risk," "moderate risk," and "no risk." Ratings were completed for

n=254 subjects during acquisition, n=254 subjects during extinction, and n=247 subjects during extinction recall.

Throughout the fear generalization task, participants provided US expectancy ratings on a 0-100 scale for every trial during CS presentation to assess their perceived likelihood of a US. During acquisition, US expectancy ratings were evaluated for every trial and were not averaged for either stimulus type. During generalization, US expectancy ratings were evaluated for every trial and were averaged for each stimulus type (four trials for CS+ or CS-, two trials for all GSs). Ratings were completed in the fear generalization task for n=211 subjects during the acquisition phase and n=208 subjects during the generalization phase.

Galvanic Skin Response

Due to its reliable differentiation of threat and safety cues, galvanic skin response is one of the predominant approaches for measuring physiological arousal in fear conditioning studies (Kreibig, 2010). Galvanic skin response was recorded throughout fear acquisition, extinction, and extinction recall in the MRI scanner as well as during the independent behavioral test of fear generalization.

Signals were acquired using a GSR100c amplifier (Biopac Systems Inc., USA) and digitized using AcqKnowledge Data Acquisition and Analysis Software (Biopac Systems Inc., USA). Data were sampled at a rate of 1kHz, with a gain of 5 μS/V and further processed using the software ANSLAB. Data were visually inspected, movement artifacts were edited out (on a trial-by-trial basis) and data that still had poor quality signal following this step (i.e., technical issues with data collection, lack of variance in acquired data or excessive motion artifacts that could not be edited out) were removed.

Skin conductance response (SCR) to the CSs were calculated by subtracting pre-CS baseline skin conductance level (SCL; -2 to 0s before CS onset) from the maximum CS SCL (occurring between 0 to 6s after CS onset). Data were normalized using the natural logarithm of 2+SCR. After data exclusion based on signal quality and motion artifacts, there were n=204 for fear acquisition, n=202 for fear extinction, n=180 for extinction recall, and n=214 for fear generalization.

Analyses

Fear Acquisition, Extinction, Extinction Recall (in-scanner tasks)

Since the self-report ratings of US contingency during fear acquisition were rated from 1-3, we defined "acquirers" (i.e., individuals who successfully acquired an awareness of the association between the CS and shock) as individuals who rated the CS+E as either "moderate risk" or "high risk," and "non-acquirers" as individuals who rated the CS+E as "no risk." Likewise, since the self-report ratings of US contingency during fear extinction and extinction recall were rated from 1-3, we defined "extinguishers" (i.e., individuals who successfully learned that the CS no longer predicted shock) as individuals who rated the CS+E as "no risk" and "non-extinguishers" as individuals who rated the CS+E as either "moderate risk" or "high risk." We then computed the partial correlation between extinction group and Anhedonia-Apprehension values (controlling for Fears, General Distress and Site) during acquisition, extinction, and extinction recall. Within the decoder-training sample, analyses were repeated using brain-predicted Anhedonia-Apprehension.

The relationships between Anhedonia-Apprehension and SCR were evaluated using linear multilevel modeling (MLM) in Stata 17.0. All models included random effects of the intercept and fixed slopes for each subject. Restricted maximum likelihood was utilized for

estimation of degrees-of-freedom (Kenward & Roger, 1997). We computed the association

between Anhedonia-Apprehension (controlling for Fears, General Distress, and Site) and SCR to

the CS+E. All measures of SCR were computed as reactivity to the CS+E relative to the CS-

(i.e., CS+E minus CS-). Likewise, all measures of US expectancy were computed as expectancy

for the CS+E relative to the CS- (i.e., CS+E minus CS-).

<u>Fear Generalization</u>

The relationships between Anhedonia-Apprehension and SCR or US Expectancy were

evaluated using linear multilevel modeling (MLM) in Stata 17.0. All models included random

effects of the intercept and fixed slopes for each subject. Restricted maximum likelihood was

utilized for estimation of degrees-of-freedom (Kenward & Roger, 1997). For the acquisition

phase, we computed the interaction between Anhedonia-Apprehension and Stimulus Type

(controlling for Fears, General Distress and Site) in association with US expectancy ratings or

SCR for the CS+. For the generalization phase, we computed the interaction between

Anhedonia-Apprehension and Stimulus Type (controlling for Fears, General Distress and Site) in

association with US expectancy ratings or SCR for all stimulus types. All measures of SCR were

computed as reactivity to the CS+ relative to the CS- (i.e., CS+ minus CS-). Likewise, all

measures of US expectancy were computed as expectancy for the CS+ relative to the CS- (i.e.,

CS+ minus CS-).

<u>Exploratory Analyses</u>

Within the decoder-training sample, all analyses were repeated using brain-predicted

Anhedonia-Apprehension (although analyses did not include the fear acquisition task occurring

directly prior to the extinction task on which this decoder was trained). Furthermore, exploratory

analyses repeated all procedures described above (i.e., analyzing SCR, US contingency, and US

expectancy during fear acquisition, extinction, extinction recall, and fear generalization) to evaluate the effects of other tri-level factors (i.e., Fears or General Distress) in relation to Pavlovian fear learning processes.

RESULTS:

Group Definitions

During fear acquisition, n=236 were defined as "acquirers" and n=18 were defined as "non-acquirers." During fear extinction, n=211 were defined as "extinguishers" and n=43 were defined as "non-extinguishers." During extinction recall, n=225 were defined as "extinguishers" and n=23 were defined as "non-extinguishers."

Associations with Anhedonia-Apprehension

Table 1 summarizes the various associations between Anhedonia-Apprehension and tasks in the present study.

<u>Fear Acquisition, Extinction, and Extinction Recall (in-scanner tasks)</u>

Acquisition

Acquisition group (acquirers vs nonacquirers) was not significantly associated with Anhedonia-Apprehension (t(249)=-.908, *p*=.365). The interaction between Anhedonia-Apprehension and Trial in predicting SCR to the CS+E was not significant (b=.01, 95% CI:[-.01,.02], t(1426)=1.08, *p*=.281).

Extinction

Extinction group (extinguishers vs nonextiguishers) was not significantly associated with Anhedonia-Apprehension (t(249)=-.712, *p*=.477). The interaction between Anhedonia-Apprehension and Trial in predicting SCR to the CS+E was not significant (b=0.00, 95% CI:[-.00,.00], t(3023.06)=.41, *p*=.684).

Extinction Recall

Extinction recall group (extinguishers vs nonextiguishers) was not significantly associated with Anhedonia-Apprehension (t(243)=-1.427, *p*=.155). The interaction between Anhedonia-Apprehension and Trial in predicting SCR to the CS+E was not significant (b=-0.01, 95% CI:[-.03,.01], t(1250.91)=-.99, *p*=.321).

Fear Generalization

Acquisition Phase

There was a significant interaction between Anhedonia-Apprehension and Stimulus Type in predicting US expectancy to the CS+ (b=-.49, 95% CI:[-.82,-.16], t(2319)=-2.90, p=.004) (Fig. 1a). The interaction between Anhedonia-Apprehension and Stimulus Type in predicting SCR to the CS+ was not significant (b=.00, 95% CI:[-.01,.01], t(2319)=-.32, *p*=.752).

Generalization Phase

There was a significant interaction between Anhedonia-Apprehension and Stimulus Type in predicting self-reported US expectancy, such that participants high on Anhedonia-Apprehension tended to have a flatter generalization curve (i.e., higher US expectancies for the CS- and GSs resembling the CS-) (b=.53, 95% CI:[.22,.83], t(1870)=3.38, *p*=.001) (Fig. 2). There was not a significant interaction between Anhedonia-Apprehension and Stimulus Type in predicting SCR to the stimuli (b=.00, 95% CI:[-.01,.00], t(1886)=-.54, *p*=.591).

Associations with Brain-Predicted Anhedonia-Apprehension

Fear Acquisition, Extinction, and Extinction Recall (in-scanner tasks)

Acquisition

As the brain-predicted Anhedonia-Apprehension were generated during the extinction phase (which directly follows this phase of the task), associations with brain-predicted Anhedonia-Apprehension were not evaluated.

Extinction

Extinction group was marginally associated with brain-predicted Anhedonia-Apprehension (t(122)=-1.719, *p*=.088), such that non-extinguishers tended to exhibit higher brain-predicted Anhedonia-Apprehension. There was not a significant interaction between brain-predicted Anhedonia-Apprehension and Trial in predicting SCR (b=0.01, 95% CI:[-.03,.05], t(1478.29)=.45, *p*=.651).

Extinction Recall

Extinction group was not significantly associated with brain-predicted Anhedonia-Apprehension (t(121)=-.925, *p*=.357). There was not a significant interaction between brain-predicted Anhedonia-Apprehension and Trial in predicting SCR (b=-.04, 95% CI:[-.22,.15], t(632.84)=-.38, *p*=.704).

Fear Generalization

Acquisition Phase

There was a significant interaction between brain-predicted Anhedonia-Apprehension and Stimulus Type in predicting US expectancy to the CS+E, such that individuals with greater brain-predicted anhedonia tended to rate lower expectancy for the CS+E compared to the CS- by the end of the acquisition phase (b=-3.40, 95% CI:[-6.08,-.72], t(1186)=-2.48, p=.013) (Fig. 1b). There was not a significant interaction between brain-predicted Anhedonia-Apprehension and Stimulus Type in predicting SCR to the CS+E (b=.03, 95% CI:[-.05,.11], t(1186)=.67, *p*=.501).

Generalization Phase

There was not a significant interaction between brain-predicted Anhedonia-Apprehension and stimulus type in predicting US expectancy (b=1.98, 95% CI:[-.57,4.52], t(943)=1.52, *p*=.128). There was not a significant interaction between brain-predicted Anhedonia-Apprehension and stimulus type in predicting SCR (b=.02, 95% CI:[-.03,.06], t(959)=.77, *p*=.444).

Exploratory Analyses: Associations with Fears or General Distress:

Fear Acquisition, Extinction, and Extinction Recall (in-scanner tasks)

Fear Acquisition

Acquisition group was not significantly associated with Fears (t(249)=.541, *p*=.589) or General Distress (t(249)=.243, *p*=.808). There was not a significant interaction between Fears and Trial (b=.00, 95% CI:[-.01,.02], t(1426)=.41, *p*=.682) in predicting SCR. There was a marginal interaction between General Distress and Trial (b=.01, 95% CI:[.00,.10], t(1426)=1.64, *p*=.101), such that individuals with the highest General Distress tended to show the greatest increase in SCR to the CS+E compared to the CS- throughout acquisition.

Extinction

Extinction group was not significantly associated with Fears (t(249)=.918, *p*=.359) or General Distress (t(249)=.811, *p*=.418). There was not a significant interaction between Fears and Trial (b=.00, 95% CI:[.00,.01], t(3023.93)=.54, *p*=.587) or between General Distress and Trial (b=.00, 95% CI:[-.01,.00], t(3025.46)=-.84, *p*=.402) in predicting SCR to the CS+E.

Extinction Recall

Extinction group was not significantly associated with Fears (t(243)=-.999, *p*=.319) or General Distress (t(243)=-1.186, *p*=.237). There was not a significant interaction between Fears

and Trial (b=.00, 95% CI:[-.02,.02], t(1254.25)=1.19, *p*=.721) or between General Distress and

Trial (b=-.01, 95% CI:[-.02,.01], t(1254.72)=-.50, *p*=.620) in predicting SCR.

<u>Fear Generalization</u>

Acquisition Phase

There was not a significant interaction between Fears and Stimulus Type in predicting

US expectancy to the CS+E (b=-.10, 95% CI:[-.45,.24], t(2319)=-.589, *p*=.556). There was a

marginally significant interaction between General Distress and stimulus type in predicting US

expectancy to the CS+E (b=-.29, 95% CI:[-.62,.03], t(2319)=-1.77, p=.077), such that individuals

with greater General Distress tended to show lower expectancy for the CS+E compared to the

CS- at the end of acquisition.

There was not a significant interaction between Fears and Stimulus Type (b=.00, 95%

CI:[-.01,.01], t(2319)=-.68, *p*=.496) or between General Distress and Stimulus Type (b=.00, 95%

CI:[-.01,.01], t(2319)=-.34, *p*=.737) in predicting SCR to the CS+E.

Generalization Phase

There was not a significant interaction between Fears and Stimulus Type in predicting

self-reported US expectancy during fear generalization (b=-.13, 95% CI:[-.45,.19], t(1870)=-.80,

p=.422). However, there was a marginal main effect of Fears, such that individuals with greater

Fears tended to exhibit greater US expectancy ratings across all stimuli (b=2.47, 95% CI:[-

.17,5.10], t(203)=1.84, *p*=.067) (Fig. 3a). There was not a significant interaction between

General Distress and Stimulus Type in predicting SCR during fear generalization (b=.04, 95%

CI:[-.26,.34], t(1870)=.26, *p*=.792). However, there was a significant main effect of General

Distress, such that individuals with greater General Distress tended to exhibit greater US

expectancy ratings across all stimuli (b=2.48, 95% CI:[.04,4.92], t(203)=2.00, *p*=.046) (Fig. 3b).

There was not a significant interaction between Fears and Stimulus Type in predicting SCR during fear generalization (b=.00, 95% CI:[.00,.01], t(1886)=.73, *p*=.467). There was not a significant interaction between General Distress and Stimulus Type in predicting SCR during fear generalization (b=.00, 95% CI:[-.01,.01], t(1886.02)=-.68, *p*=.496).

DISCUSSION:

The present study evaluated the extent to which symptom-derived anhedonia, as well as anhedonia predicted from brain activity during a fear extinction task, related to subjective and skin conductance measures of Pavlovian fear learning across two sets of tasks: (1) fear acquisition, extinction, and extinction recall collected during an fMRI scan, and (2) fear acquisition and generalization during a separate laboratory visit. Results indicated that there were no significant associations between anhedonia and fear learning during the first set of tasks. However, results showed that anhedonia, over and above other symptom dimensions, was associated with less acquisition of US expectancy and subsequently greater generalization of US expectancy during the second task.

Overgeneralization of fear is associated with a range of anxiety disorders, including panic disorder (Lissek et al., 2010), generalized anxiety disorder (Lissek et al., 2014; Tinoco-González et al., 2015), social anxiety disorder (Ahrens et al., 2016), and posttraumatic stress disorder (Kaczkurkin et al., 2017), and is additionally associated with maladaptive instrumental-avoidance behaviors (van Meurs et al., 2014). The present study found that anhedonia symptoms, which are commonly found in depression but also associated with anxiety disorders (e.g., Kashdan, 2007; Abramovitch et al., 2014; Nawjin et al., 2015), were associated with greater generalization over and above other symptom dimensions characteristic of anxiety and depression. In contrast, neither general (i.e., "General Distress") nor anxiety-specific (i.e.,

49

"Fears") symptoms were associated with aberrances in any phases of the fear learning tasks. These findings differ from prior research which has reported that anxiety disorders are associated with differential SCR during fear acquisition and extinction (Duits et al., 2015). The lack of significant findings for either general distress (which includes symptoms relevant to both anxiety and depression) or fears (more specific to anxiety) in the present study may be attributable to several possibilities. For example, although direct comparisons of individuals with anxiety disorders versus healthy controls tend to show consistent fear learning effects, these effect sizes tend to be modest (Duits et al., 2015), and studies of trait anxiety have not shown similar effects (e.g., Torrents-Rodas et al., 2013). It is possible that significant effects would emerge for group comparisons focused on the extremes of general distress or fears symptom dimensions. Furthermore, measurement of SCR in an MRI scanner may have presented methodological issues that interfered with the detection of SCR differences during the present study (Gray et al., 2009; Lonsdorf et al., 2017). Additional research is needed to further clarify the extent to which symptoms dimensions other than anhedonia are associated with aberrances in fear acquisition or generalization.

The initial learning of CS-US associations is a crucial step informing subsequent fear generalization (Lissek et al., 2008). In the present study, symptom derived anhedonia was associated with less discrimination of the CS+ from the CS- during fear acquisition as measured by US expectancy, which may have contributed to generalization effects. If anhedonic individuals form the initial CS-US predictions less strongly compared with other individuals, they may subsequently recall the CS-US association less accurately, resulting in greater uncertainty across the generalization stimuli. Indeed, it has been suggested that overgeneralization of fear among individuals with an anxiety disorder occurs because anxious

individuals tend to extend fear associations to entire classes of stimuli, rather than learning highly specific CS-US associations (Lissek, 2012). Results from the present study suggest that anhedonia symptoms may account for overgeneralization effects commonly found among anxiety disorders, particularly regarding self-reported expectancy of a negative US. These results were specific to anhedonia over and above fears and general distress, which did not show similar effects during fear generalization.

Results from the present study did not highlight psychophysiological indices of fear generalization. Whereas the present study used galvanic skin response as the primary metric of fear physiology, many prior studies have evaluated fear generalization using fear-potentiated startle (i.e., electromyography) (e.g., Lissek et al., 2008; Lissek et al., 2010; Lissek et al., 2014; Tinoco-González et al., 2015; for a review, see Dymond et al., 2015). It is possible that anhedonia symptoms would have been associated with fear-potentiated startle during the fear generalization task. Evidence suggests that depression is associated with generally reduced startle responses to fearful stimuli (Kaviani et al., 2004; Mneimne et al., 2008; Taylor-Clift et al., 2011; Yancey et al., 2015), although this association has not been studied in the context of fear generalization. It is additionally possible that anhedonia symptoms are specifically associated with overgeneralization of self-reported US expectancy, whereas other symptoms associated with anxiety disorders are associated with overgeneralization of fear-potentiated startle responses. Future research is needed to directly test the potential coherence or divergence of these fear generalization indices according to orthogonal symptom dimensions, such as those described by the Tri-Level Model (Naragon-Gainey et al., 2016).

Unlike the main Anhedonia-Apprehension measure, brain-predicted Anhedonia-Apprehension was not significantly associated with fear generalization. Effects for brain-

predicted Anhedonia-Apprehension were nonetheless in the same direction as the main measure of Anhedonia-Apprehension. The lack of statistically significant replication may be attributable to the relatively small R^2 values derived in Study 1 ($R^2 = .168$ in training; $R^2 = .047$ in external validation), such that brain-predicted Anhedonia-Apprehension corresponds with Anhedonia-Apprehension above a chance level but with imperfect correspondence. Thus, brain-predicted Anhedonia-Apprehension may still associate with similar measures as Anhedonia-Apprehension, but with less specificity.

If anhedonia is associated with disruptions in Pavlovian fear learning, it is possible that anhedonia is further associated with other processes that rely on these mechanisms, such as exposure therapy (Pittig et al., 2016; Scheveneels et al., 2016; Fullana et al., 2020). Future studies should evaluate the relationship between anhedonia and mechanisms of exposure therapy. In addition, whereas low positive affect (i.e., anhedonia) may impair fear acquisition and lead to subsequent overgeneralization of fear associations, high positive affect has been shown to enhance long-term retention of fear extinction memories (Zbozinek et al., 2015; Zbozinek & Craske, 2017a). It is therefore possible that high positive affect is associated with the formation of more specific CS-US relationships during fear acquisition that do not generalize to other stimuli and therefore are more readily extinguished. Interventions designed to increase positive affect (e.g., Craske et al., 2019; for a review see Sandman et al., 2022) may improve the efficacy of exposure therapy (Craske et al., 2016; Zbozinek & Craske, 2017b) and warrant additional research.

Anhedonia has frequently been considered in relation to reward learning mechanisms (e.g., Huys et al., 2013; Rizvi et al., 2016). It is possible that anhedonia is associated with general disruptions to associative learning, detected here in a more specific assay of Pavlovian fear

generalization. For example, it is possible that anhedonia might be associated with aberrances in reward acquisition and subsequent generalization to perceptually similar stimuli. Future studies of anhedonia may incorporate reward and fear tasks to demonstrate the specificity or breadth of learning effects.

The present study should be interpreted in light of several limitations. First, the significant association between anhedonia and fear acquisition was only detected during one of the two sets of tasks, underscoring the need for replication. Importantly, the self-report aspects of the in-scanner tasks (acquisition, extinction, extinction recall) were performed retrospectively and involved a small range of responses (1-3), whereas the fear generalization task employed in-the-moment assessments of US expectancy across a wide range of responses (1-100). This difference may have made it more difficult to detect self-report effects in the fear conditioning tasks. Second, this study recruited individuals across a wide range of anhedonia symptoms. Future research is needed to evaluate associations between anhedonia and fear generalization among subjects specifically recruited for high anhedonia symptoms. Indeed, studies of trait anxiety have found no associations with aberrant fear learning (Lonsdorf & Merz, 2017; Torrents-Rodas et al., 2013), whereas clinical studies have reliably shown differences in extinction among individuals with anxiety disorders (Lissek et al., 2005; Graham & Milad, 2011; Duits et al., 2015; Marin et al., 2020). A similar pattern may be evident for studies of anhedonia. Finally, as noted in Rosenberg et al., 2022, this study involved a narrow age range of participants and warrants replication among a sample that encompasses the full lifespan.

In conclusion, the present study found that anhedonia is associated with overgeneralization of self-reported US expectancies during acquisition and generalization phases of a fear generalization task. Future research is needed to explore if anhedonia is also associated

with 1) fear-potentiated startle as an index of fear generalization and 2) distinct functioning of the neural circuits that support threat discrimination. Likewise, future research should evaluate if interventions to increase positive affect can improve the efficacy exposure therapy, particularly among individuals with elevated anhedonia symptoms.

TABLES AND FIGURES:

Table 1: Associations between Tri-Level symptoms and behavioral measures.

Fear Conditioning Tasks		Tri-Level Factor			
		Anhedonia-Apprehension	Brain-Predicted Anhedonia-Apprehension	Fears	General Distress
Acquisition	**Contingency Ratings (1-3)**	NS	N/A	NS	NS
	SCR *(CS+E > CS-)*	NS	N/A	NS	*Interaction (p=.101)*
Extinction	**Contingency Ratings (1-3)**	NS	*Main Effect (p=.088)*	NS	NS
	SCR *(CS+E > CS-)*	NS	NS	NS	NS
Extinction Recall	**Contingency Ratings (1-3)**	NS	NS	NS	NS
	SCR *(CS+E > CS-)*	NS	NS	NS	NS
Fear Generalization Tasks		**Tri-Level Factor**			
		Anhedonia-Apprehension	Brain-Predicted Anhedonia-Apprehension	Fears	General Distress
Acquisition	**US Expectancy Ratings (1-100)** *(CS+E > CS-)*	**Interaction (*p* =.004)**	**Interaction (*p* =.013)**	NS	*Interaction (p=.077)*
	SCR *(CS+E > CS-)*	NS	NS	NS	NS
Generalization	**US Expectancy Ratings (1-100)** *(CS+E > CS-)*	**Interaction (*p* =.001)**	NS	*Main Effect (p=.067)*	**Main Effect (*p* =.046)**
	SCR *(CS+E > CS-)*	NS	NS	NS	NS

Fig. 1: Anhedonia is associated with less acquisition of fear (measured by US expectancy) (green = high anhedonia, red = mean anhedonia, blue = low anhedonia). **Fig. 1a** depicts the effect of anhedonia, and **Fig. 1b** depicts the effect of brain-predicted anhedonia.
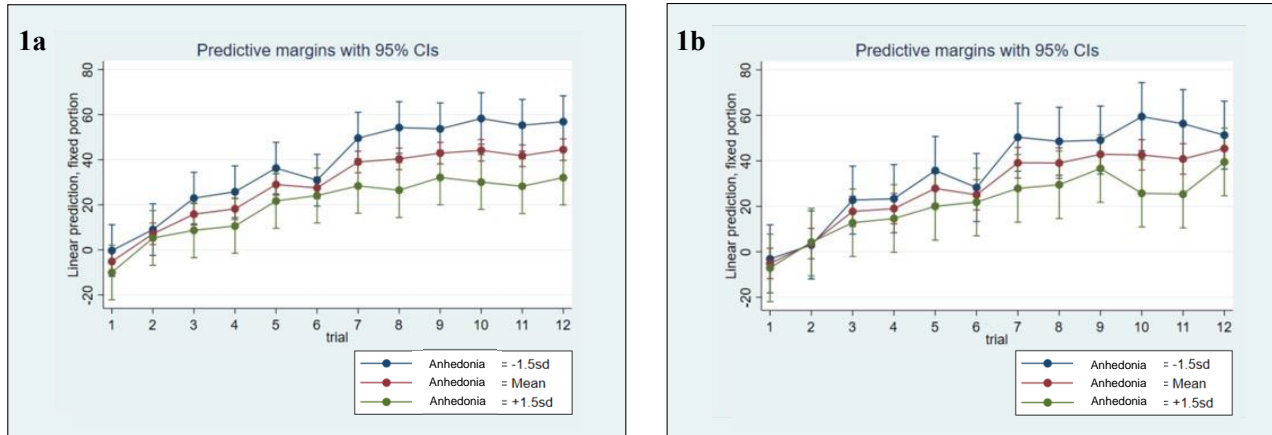

Fig. 2: Anhedonia is associated with overgeneralization of fear (measured by US expectancy) (green = high anhedonia, red = mean anhedonia, blue = low anhedonia).
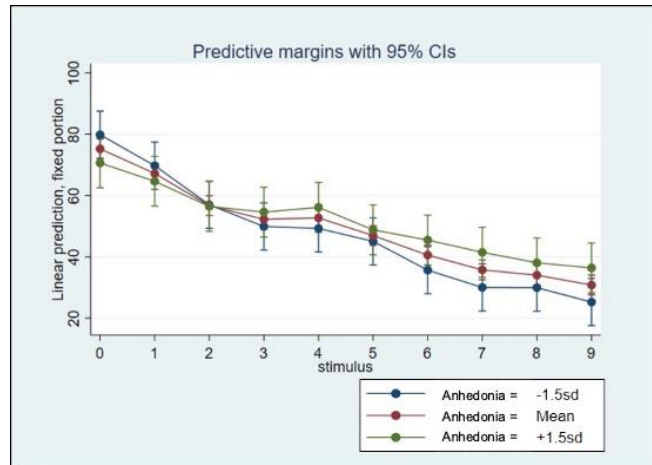

56

Fig. 3: Exploratory analyses of Fears and General Distress. **Fig. 3a** depicts a marginal association between Fears and heightened fear (US expectancy) across generalization stimuli (green = high fears, red = mean fears, blue = low fears). **Fig. 3b** depicts a significant association between General Distress and heightened fear (US expectancy) across generalization stimuli (green = high fears, red = mean fears, blue = low fears).
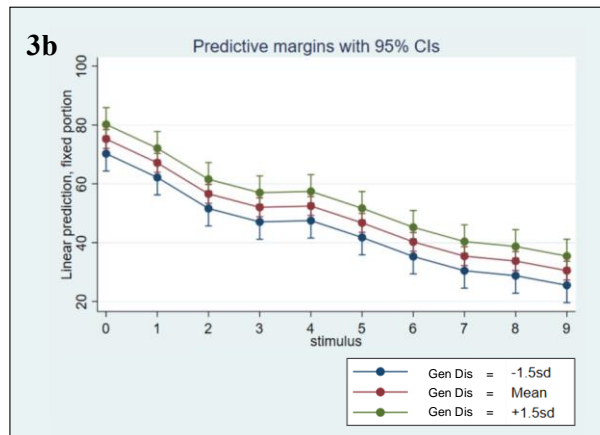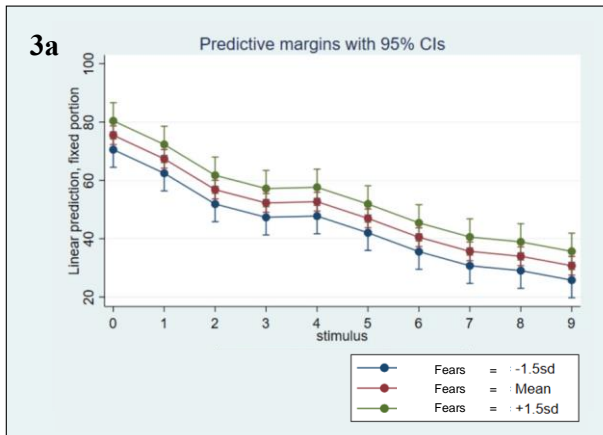
SUPPLEMENT

Fig. S1: On average, participants appeared to acquire fear (SCR) during the in-scanner fear acquisition task.
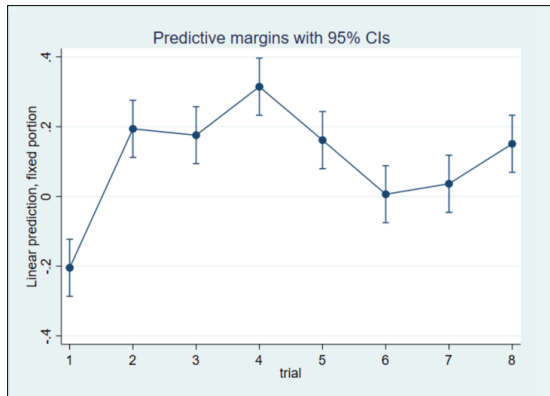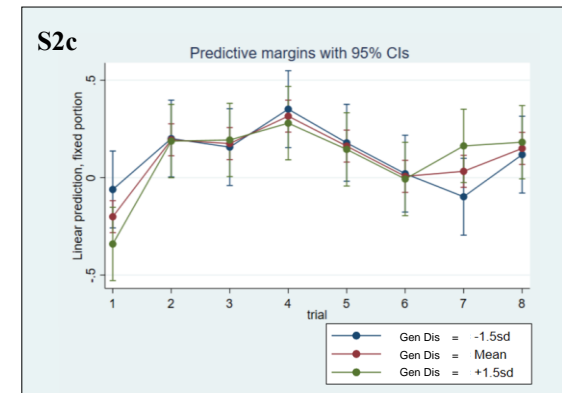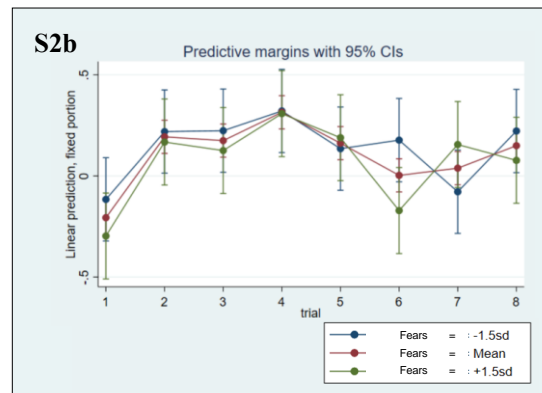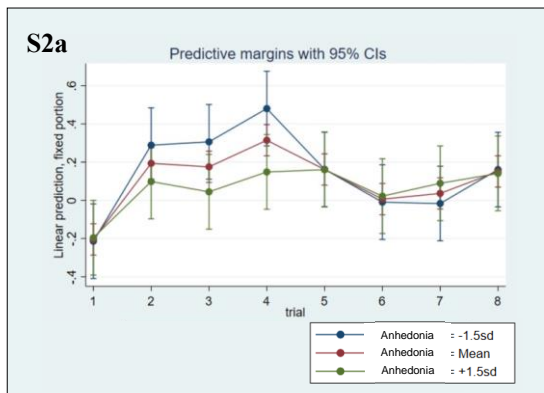

Fig. S2: Participants appeared to acquire fear (SCR) similarly across levels of Anhedonia (Fig. S2a), Fears (Fig. S2b), or General Distress (Fig. S2c) during the in-scanner fear acquisition task.

STUDY 3:

POSITIVE AFFECT MODERATES EXPECTANCY VIOLATION DURING EXPOSURE

THERAPY FOR SOCIAL ANXIETY

<u>INTRODUCTION:</u>

Exposure therapy is considered a gold-standard treatment for anxiety disorders (Rauch et al., 2021; Abramowitz et al., 2019). However, an estimated 50% of individuals do not achieve clinically significant gains (Loerinc et al., 2015). Depressive disorders are highly comorbid with anxiety disorders and are associated with increased severity of anxiety (Breteler et al., 2021). Comorbidity with depression may undermine the efficacy of exposure therapy for anxiety disorders, although evidence is mixed (see Abramowitz & Landy, 2013; Rozen & Aderka, 2020). The goal of the current study is to evaluate whether depressive symptoms are associated with differences in the learning mechanisms underlying exposure therapy for social anxiety.

Exposure therapy involves repeated confrontation with feared stimuli (objects, situations, interoceptive cues, or memories), often in the absence of the feared outcome. Procedurally, this is analogous to Pavlovian extinction, in which the conditional stimulus (CS) that was previously paired with an aversive outcome (unconditional stimulus, US) is repeatedly presented without the US. Extinction learning occurs as a function of prediction error: a mismatch between high US expectancy (either explicit or implicit) and low actual rate of US occurrence (Rescorla & Wagner, 1972). Whereas earlier models emphasized *unlearning* of the excitatory CS-US association (Rescorla & Wagner, 1972), the inhibitory retrieval model of extinction posits that the original CS-US association acquired during fear conditioning is not erased, but rather is left intact as new, secondary learning about the CS-US association develops – specifically, that the CS no longer predicts the US (Bouton, 1993; Craske, Treanor et al., in press). Consequently, the

CS possesses two meanings following extinction: the original excitatory meaning (CS-US) and a new meaning (CS-noUS).

The inhibitory retrieval model of exposure therapy draws from the basic science of extinction to emphasize both prediction error and inhibitory learning (see Craske et al., 2018). Prediction error is maximized through expectancy violation, wherein anxious patients learn that feared stimuli (CS) do not reliably predict feared outcomes (US), thereby strengthening CS-noUS associations that compete with original CS-US associations (Craske et al., 2018). This process is thought to prompt the updating of future predictions, such that patients ultimately reduce their expectation of a CS-US pairing and extinguish their fear, moderated by contextual specificity.

Prediction error or expectancy violation has been shown to rely, in part, on dopaminergic reward mechanisms (e.g., Bayer et al., 2005; Waelti et al., 2001). Such dopaminergic reliance has been demonstrated during prediction error within fear extinction, measured at the point of non-occurrence of an expected aversive US (Salinas-Hernández et al., 2018; Papalini et al., 2020). US omission has been posited to induce a state of relief-pleasantness that is hypothesized to be central to the formation of CS-noUS associations as they begin to compete with CS-US associations during extinction (Vervliet et al., 2017). Relief-pleasantness has been shown to be greatest for strong CS-US predictions (Willems & Vervliet, 2021). In other words, individuals experience maximal relief-pleasantness when they are most certain that a CS will lead to a US, and no US occurs. As such, individual differences in relief-pleasantness may minimize or magnify the positivity of realizing that a feared outcome did not occur, and thereby moderate the effectiveness of exposure therapy.

Relief-pleasantness is closely related to the construct of reward responsiveness. Reward hyporesponsivity is characteristic of anhedonic depression (Admon & Pizzagalli, 2015; Treadway & Zald, 2011), and evidence is accruing to suggest a link between anhedonia and extinction. Specifically, low positive affect (a core feature of anhedonia) may be related to impairments in extinction learning and retention (Geschwind et al., 2015; Zbozinek et al., 2015; Zbozinek et al., 2015; Zbozinek & Craske, 2017a). Likewise, anhedonia has been shown to correlate with neural functioning during extinction (Young et al., 2021; Rosenberg et al., in press) in ways that suggest impairments in the brain systems responsible for threat processing and US expectancy updating. Aberrances in dopaminergic functioning have been associated with anhedonia (e.g., Dunlop & Nemeroff, 2007; Treadway & Zald, 2011), with known perturbations in prediction error signaling and learning (e.g., Gradin et al., 2011; Vrieze et al., 2013; Pizzagalli, 2014; Greenberg et al., 2015) tending to correspond with reduced updating of behaviors following reinforcement learning (Pizzagalli, 2014). One potential explanation is the tendency for depressed individuals to overlook positive information that disconfirms negative expectations (e.g., Liknaitzky et al., 2017; Everaert et al., 2018; Kube & Glombiewski, 2021; Kube et al., 2021: Kube & Rozenkrantz, 2021). Interestingly, evidence that depressed individuals update their beliefs from *moderate* expectancy violations moreso than from maximal expectancy violations (Kube et al., 2021). This suggests a non-linear relationship between US expectancy and expectancy violation with positive feedback (Kube et al., 2020), which contrasts traditional linear models of expectancy violation (e.g., Rescorla & Wagner, 1972). The extent to which these effects are driven by anhedonic features of depression or depression more broadly is yet to be determined.

In summary, relief-pleasantness is theorized to be central to the formation of CS-noUS associations during extinction (Willems & Vervliet, 2021), and maximal prediction error corresponds with maximal relief-pleasantness. If depressed individuals (and particularly those with elevated anhedonia) show reduced sensitivity to the most relieving outcomes, then weaker CS-noUS associations may evolve during extinction/exposure for such individuals. Therefore, it is conceivable that one pathway through which depression becomes associated with anxiety is by anhedonic reductions in capacity to update predictions following expectancy violation.

The present study was conducted as part of a randomized-controlled trial comparing within-session fear reduction and inhibitory retrieval learning models of exposure therapy [ClinicalTrials.gov ID: NCT04048824]. The goal was to examine if a core feature of anhedonia (i.e., low positive affect) is associated with learning mechanisms of exposure therapy among individuals with social anxiety disorder – a condition highly co-occurring with anhedonia symptoms (Kashdan, 2007). Within the inhibitory retrieval learning treatment condition, we hypothesized that lower trait positive affect would be associated with lower expectancy violation during *in vivo* exposures, specifically when the expectancy of a feared outcome is at its highest. To investigate the specificity of effects to inhibitory retrieval learning relative to overall distress, we conducted secondary analyses within the fear reduction treatment condition to evaluate the degree to which anhedonia influenced fear reduction.

MATERIALS AND METHODS

Trial Design

This study was a parallel, two-arm, randomized (1:1), open label clinical superiority trial comparing inhibitory and fear reduction models of exposure therapy among individuals with social anxiety disorder or panic disorder. All participants signed consent. Aside from COVID-19

restrictions (see Supplement), no changes occurred after trial commencement (ClinicalTrials.gov

ID: NCT04048824). This study adhered to Consolidated Standards of Reporting Trials

(CONSORT) reporting guidelines. No serious adverse events were reported.

Participants

A brief study description was distributed via lab websites, social media,

ClinicalTrials.gov, and campus and local area flyers. The inclusion criteria for participation were

age 18-65, English-speaking, seeking treatment for social anxiety disorder or panic disorder, and

either stabilized on psychotropic medications or medication-free. Exclusion criteria were serious

medical conditions, pregnancy, history of suicidal ideation or self-harm in the past year, history

of suicide attempts in the last 10 years, history of bipolar disorder, psychosis, intellectual

disability or organic brain damage, alcohol or substance use disorder within last 6 months, and

concurrent therapy focused on anxiety.

The present report specifically focuses on individuals recruited for treatment of social

anxiety disorder. To determine the presence of social anxiety symptoms, potential participants

were initially screened using the Mini-Social Phobia Inventory (Mini-SPIN; Connor et al., 2001).

Screened participants then underwent the Structured Clinical Interview for Diagnosis (SCID-V;

First, 2014) – a semi-structured diagnostic interview, conducted by reliability-certified

interviewers with diagnostic consensus led by experienced clinicians (MT or MGC). For

diagnosed disorders, clinical severity was rated from 0 to 8, and participants were eligible for

inclusion if they received a clinical severity rating of 4 or higher for social anxiety disorder.

Participants were included in analyses only if they completed at least one exposure

session (i.e., completion of session two).

Interventions

Treatment entailed nine weekly individual, manualized therapy sessions. The first session lasted 90 minutes and involved psychoeducation regarding the exposure therapy approach used in this study. Sessions two through nine lasted 60 minutes and involved *in vivo* exposures and assignment of homework exercises. Therapy was provided by highly trained doctoral students in the UCLA Department of Psychology and supervised by licensed clinical psychologists (MT and MGC). Sessions took place in the UCLA Psychology Clinic or remotely via an encrypted video-conferencing platform due to the COVID-19 pandemic. Participants were encouraged not to seek other forms of treatment (psychological or psychiatric) and to maintain their current medication regimen (if applicable) until completion of the nine-week treatment period.

In the inhibitory retrieval learning condition, participants first defined a nexus of factors influencing the likelihood of CS-US relationships (e.g., contexts or cues that make the US more or less likely to occur; Craske, Treanor et al., in press). Among socially anxious participants, the US (i.e., social rejection) was typically operationalized according to behavioral markers of rejection or judgement (e.g., explicit statements or facial expressions, or other pre-defined behaviors). Exposure was therapist-directed during each session; participants first rated their expectancy of a US (PreExp), followed by an exposure activity designed to violate expectancies (see Craske, Treanor et al., in press, for details of the inhibitory retrieval model of exposure), consolidation exercises, and finally a rating of expectancy for the US if they were to hypothetically repeat this exposure (PostExp). Participants were instructed to complete homework exposures to test the same CS-US association between sessions, although exposures varied contextual elements to maximize extinction generalization.

In the fear reduction condition, participants first constructed a fear hierarchy in ascending order of subjective fear ratings. During therapist-directed exposure sessions, participants first

rated Fear (PreFear), followed by an exposure activity repeated the number of times necessary for fear to reduce by 50% or a maximum of 45 mins, and finally a rating of Fear following the last exposure (PostFear). Participants were instructed to complete the exact same exposures as homework between sessions.

Randomization and Masking

Participants were randomized 1:1 to inhibitory retrieval learning vs. fear reduction via computer-generated allocation using permuted block randomization. Randomization was managed by investigator AS. Only evaluators were masked to group allocation. Study coordinators enrolled and assigned participants according to the randomization table.

Measures

Prior to beginning treatment, participants completed a two-day baseline assessment including the SCID-5 (First, 2014), behavioral assessments (see Supplement for details), and self-report questionnaires. The primary outcome measure was the Liebowitz Social Anxiety Scale (LSAS; Liebowitz; 1987) to measure social anxiety symptoms. Positive and negative affect were measured using the Positive and Negative Affect Schedule (PANAS-41; Watson et al., 1988), with ratings anchored over the preceding month. The primary metric of anhedonia was derived from the Positive Affect subscale (PANAS-P), which comprised items from the "General Dimension Positive Affect," "Basic Positive Emotion," and "Serenity" scales (24 items, see Supplement). The primary metric of negative affect was derived from the Negative Affect subscale (PANAS-N), which comprised items from the "General Dimension Negative Affect" and "Basic Negative Emotion" scales (17 items, see Supplement). These scales were repeated at mid-treatment (week 5, following the fifth treatment session) and at post (week 10) along with behavioral measures and diagnostic evaluation.

For every exposure (during and between therapy sessions), participants rated either their expected likelihood of the US occurring (Inhibitory Retrieval Learning condition) or subjective fear (Fear Reduction condition) on a scale from 0 to 100. These values served as expectancy (PreExp or PostExp) or Fear (PreFear or PostFear) ratings in subsequent analyses.

Statistical Methods

The relationship between positive affect and exposure therapy mechanisms was evaluated using linear multilevel modeling (MLM) in Stata 17.0. MLM includes participants with missing data and uses all available data, which is recommended for clinical trial data (Hamer & Simpson, 2009). Restricted maximum likelihood was utilized and is recommended for small-sample studies (Kenward & Roger, 1997). All models included random effects of the intercept and fixed slopes for each subject. Analyses were conducted separately within the Inhibitory Retrieval Learning and Fear Reduction conditions. As participants completed symptom measures at baseline (before session 1), mid-treatment (before session 5), and post-treatment (after session 9), we utilized US expectancy and Fear ratings from therapy sessions two, five, and nine in analyses (note: session one does not involve exposure). Finally, the intensity of exposures is designed to change over time during exposure therapy (i.e., PreExp or PreFear frequently increases) due to the progressive nature of exposures throughout treatment. Therefore, all analyses controlled for time.

Within the Inhibitory Retrieval Learning condition, we calculated change in expectancy (ΔExp) as the PreExp - PostExp difference score. For sessions in which participants completed multiple exposure activities, we selected the exposure exercise with the maximum difference score and the highest PreExp. Although there is no current standard for measurement of expectancy violation, a range of options exist (see Supplement). Of note, as the expected

outcome frequently does not occur during exposure therapy, differing levels of PreExp are likely to determine the relative size of expectancy violation when calculated as the PreExp - PostExp difference score. Therefore, to account for relative differences in PreExp, we computed two MLMs using fixed slopes to test potential effects: (1) Main effect of anhedonia on ΔExp; (2) Interaction effect of anhedonia x PreExp on PostExp.

Within the Fear Reduction condition, we calculated change in Fear (ΔFear) as the PreFear - PostFear difference score. For sessions in which participants completed multiple exposure activities, we determined PreFear as the Fear value prior to first exposure and PostFear as the Fear value following the final exposure. We computed two MLMs to test potential effects: (1) Main effect of anhedonia on ΔFear; (2) Interaction effect of anhedonia x PreFear on PostFear.

Primary analyses utilized PANAS-P values for each timepoint. Follow-up analyses were conducted to disentangle effects according to within-subject or between-subject variance in PANAS-P over time. To this end, we calculated each subject's average PANAS-P over time as well as mean-centered PANAS-P values at each time-point. Follow-up analyses evaluated within-subject and between-subject PANAS-P as covariates for each model.

In addition, specificity analyses evaluated associations over and above PANAS-N. Interaction effects evaluated associations over and above PANAS-N interactions, the recommended approach when the moderator (PANAS-P) and covariate (PANAS-N) variables are correlated with one another (Yzerbyt et al., 2004). Lastly, specificity analyses evaluated associations between PANAS-N with ΔExp or ΔFear over and above PANAS-P values. Interaction effects evaluated associations over and above PANAS-P interactions.

Post hoc power analysis using the *ipdpower* command in Stata 17.0 (Kontopantelis et al., 2016) determined that, to achieve 80% power with n=26 subjects, the present study would need an effect size of b=-10.20 for the interaction effect of anhedonia x PreExp on PostExp.

RESULTS

Participants

The final sample consisted of N=64 participants randomized to either the inhibitory retrieval learning condition (n=32) or the fear reduction condition (n=32). Participant characteristics are summarized in Table 1.

Inhibitory Retrieval Learning

PANAS-P was not associated with ΔExp throughout treatment (b=.19, 95% CI:[-.22,.60], t(64.21)=.92, *p*=.361). This effect was not significant after adding PANAS-N as a covariate (b=.17, 95% CI:[-.25,.60], t(62.35)=.81, *p*=.423). Follow-up analyses indicated that neither within-subject variance in PANAS-P (b=.51, 95% CI:[-.12,1.14], t(42.29)=1.63, *p*=.111) nor between-subject variance in PANAS-P (b=-.03, 95% CI:[-.52,.50], t(31.80)=-.01, *p*=.980) was associated with ΔExp. These effects were not significant after adding PANAS-N as a covariate (b=.49, 95% CI:[-.15,1.13], t(41.66)=1.55, *p*=.129; b=-.10, 95% CI:[-.55,.49], t(31.26)=-.10, *p*=.922).

There was a significant PANAS-P x PreExp interaction, such that higher PANAS-P decreased the effect of PreExp on PostExp (b=-.02, 95% CI:[-.03,.00], t(60.15)=-2.01, *p*=.049) (Fig. 1a). This effect was not significant after adding PANAS-N as a covariate (b=-.01, 95% CI:[-.03,.01], t(51.69)=-.83, *p*=.409). Follow-up analyses indicated that within-subject variance in PANAS-P significantly moderated the association between PreExp and PostExp (b=-.06, 95% CI:[-.10,-.01], t(49.32)=-2.64, *p*=.011) (Fig 1b). Specifically, when PreExp was greatest (i.e., 80-

100), individuals who reported relatively low positive affect (i.e., below their within-subject mean level of positive affect) tended to show the greatest PostExp (i.e., the smallest amount of expectancy updating). This effect remained significant after adding PANAS-N as a covariate (b=-.05, 95% CI:[-.09,.00], t(48.36)=-2.19, p=.033). In contrast, between-subject variance in PANAS-P did not significantly moderate the association between PreExp and PostExp (b=-.01, 95% CI:[-.02,.01], t(62.30)=-.60, p=.552). This effect was not significant after adding PANAS-N as a covariate (b=.00, 95% CI:[-.02,.03], t(51.77)=.27, p=.791).

Within-Session Fear Reduction

PANAS-P was not associated with ΔFear throughout treatment (b=.12, 95% CI:[-.18,.41], t(68.62)=.79, p=.433). This effect was not significant after adding PANAS-N as a covariate (b=-.16, 95% CI:[-.17,.48], t(60.45)=.95, p=.345). Follow-up analyses indicated that neither within-subject variance in PANAS-P (b=.03, 95% CI:[-.42,.47], t(44.89)=.19, p=.911) nor between-subject variance in PANAS-P (b=.18, 95% CI:[-.19,.55], t(33.67)=.97, p=.338) was associated with ΔFear. These effects were not significant after adding PANAS-N as a covariate (b=.13, 95% CI:[-.37,.62], t(40.53)=.51, p=.612; b=.17, 95% CI:[-.22,.57], t(31.95)=.88, p=.383).

PANAS-P did not moderate the relationship between PreFear and PostFear (b=.00, 95% CI:[-.02,.01], t(65.61)=-.25, p=.806) (Fig 2a). This effect was not significant after adding PANAS-N as a covariate (b=.00, 95% CI:[-.02,.02], t(61.76)=.12, p=.905). Follow-up analyses indicated that neither within-subject variance in PANAS-P (b=.00, 95% CI:[-.03,.03], t(55.45)=.04, p=.966) (Fig. 2b) nor between-subject variance in PANAS-P (b=.00, 95% CI:[-.02,.01], t(67.95)=-.41, p=.681) moderated the relationship between PreFear and PostFear. These effects were not significant after adding PANAS-N as a covariate (b=.01, 95% CI:[-.02,.04], t(48.98)=.52, p=.604; b=.00, 95% CI:[-.02,.02], t(61.95)=-.19, p=.850).

Associations with PANAS-N

PANAS-N was not associated with ΔExp throughout treatment (b=-.13, 95% CI:[-.59,.33], t(68.78)=-.57, *p*=.570). This effect was not significant after adding PANAS-P as a covariate (b=-.08, 95% CI:[-.56,.39], t(67.72)=-.35, *p*=.727). Follow-up analyses indicated that neither within-subject variance in PANAS-N (b=-.14, 95% CI:[-.80,.52], t(46.03)=-.43, *p*=.667) nor between-subject variance in PANAS-N (b=-.12, 95% CI:[-.74,.49], t(34.69)=-.40, *p*=.692) was associated with ΔExp. These effects were not significant after adding PANAS-P as a covariate (b=-.13, 95% CI:[-.79,.53], t(44.48)=-.39, *p*=.698; b=-.04, 95% CI:[-.69,.61], t(34.19)=-.12, *p*=.904).

There was a significant PANAS-N x PreExp interaction, such that higher PANAS-N increased the effect of PreExp on PostExp (b=.02, 95% CI:[.00,.04], t(63.53)=2.19, *p*=.033) (Fig. 3a). This effect was not significant after adding PANAS-P as a covariate (b=.01, 95% CI:[-.01,.04], t(53.52)=1.12, *p*=.270) (Fig. 3b). Follow-up analyses indicated that within-subject variance in PANAS-N did not moderate the association between PreExp and PostExp (b=.02, 95% CI:[-.01,.06], t(52.64)=1.37, *p*=.176). This effect was not significant after adding PANAS-P as a covariate (b=.02, 95% CI:[-.01,.06], t(52.64)=1.37, *p*=.176). Likewise, between-subject variance in PANAS-N did not significantly moderate the association between PreExp and PostExp (b=.02, 95% CI:[.00,.04], t(65.74)=1.76, *p*=.100). This effect was not significant after adding PANAS-P as a covariate (b=.01, 95% CI:[-.02,.04], t(58.44)=.77, *p*=.447).

PANAS-N was not associated with ΔFear throughout treatment (b=-.06, 95% CI:[-.42,.29], t(66.89)=-.36, *p*=.717). This effect was not significant after adding PANAS-P as a covariate (b=-.02, 95% CI:[-.39,.34], t(64.64)=-.13, *p*=.897). Follow-up analyses indicated that neither within-subject variance in PANAS-N (b=-.18, 95% CI:[-.45,.08], t(53.27)=-1.41, *p*=.164)

nor between-subject variance in PANAS-N (b=-.01, 95% CI:[-.46,.49], t(32.62)=.06, *p*=.955) was associated with ΔFear. These effects were not significant after adding PANAS-P as a covariate (b=-.02, 95% CI:[-.42,.38], t(47.76)=-.10, *p*=.919; b=.06, 95% CI:[-.44,.56], t(30.25)=.25, *p*=.803).

PANAS-N did not moderate the relationship between PreFear and PostFear (b=.02, 95% CI:[.00,.05], t(62.67)=1.69, *p*=.096). This effect was not significant after adding PANAS-P as a covariate (b=.02, 95% CI:[-.01,.04], t(60.07)=1.31, *p*=.194). Follow-up analyses indicated that neither within-subject variance in PANAS-N (b=.01, 95% CI:[-.01,.02], t(52.81)=.81, *p*=.419) nor between-subject variance in PANAS-N (b=.02, 95% CI:[-.01,.04], t(63.06)=1.13, *p*=.264) moderated the relationship between PreFear and PostFear. These effects were not significant after adding PANAS-P as a covariate (b=.01, 95% CI:[-.02,.03], t(47.79)=.57, *p*=.573; b=.01, 95% CI:[-.02,.04], t(61.18)=.74, *p*=.461).

DISCUSSION

This study evaluated the association between positive affect, a core feature of anhedonia, and learning mechanisms central to the effectiveness of exposure therapy. We found that within-subject variability in positive affect moderated the relationship between pre-exposure expectancy and post-exposure expectancy in the Inhibitory Retrieval Learning condition. Specifically, for exposures in which participants had the highest expectancy of the US, participants experiencing low positive affect were less likely to update their expectations following non-occurrence of the US. This interaction was significant over and above the effects of negative affect. Positive affect had no observable effects upon fear ratings within the Fear Reduction condition.

Anhedonia is associated with aberrances in dopaminergic prediction error signaling, which is central to inhibitory retrieval learning and may relate to experiences of relief during the

omission of an expected aversive US. Experiences of relief are thought to be strongest during CS-noUS occurrences, particularly when US expectancy is highest. If the experience of relief is central to extinction learning, and if relief-pleasantness is greatest for individuals with high compared with low positive affect, variability in relief-pleasantness may explain differences in extinction learning during exposure therapy. In support of this notion, the present study found that individuals experiencing low positive affect exhibited less updating of expectancy specifically when expectations of a negative outcome were highest (i.e., when relief-pleasantness is expected to be maximized). This effect was specifically observed when subjects experienced low positive affect relative to their respective baselines. However, relief-pleasantness has been associated with increased avoidance of feared stimuli, as the non-occurrence of a US may be experienced as highly reinforcing, thereby reducing likelihood of approaching a feared stimulus at all (i.e., during noCS-noUS predictions) (San Martín et al., 2020; Vervliet et al., 2017). Therefore, although relative elevations in positive affect may augment inhibitory retrieval learning in extinction/exposure paradigms, such elevations may also correspond with increased urges to avoid feared stimuli. An alternative mechanism is that increased positive affect promotes semantic learning, which is associated with deeper encoding of information and may therefore enhance extinction learning (Zbozinek & Craske, 2017b). This "semantic learning" hypothesis (Zbozinek & Craske, 2017b) would apply to learning more generally, whereas the relief-pleasantness hypothesis would apply specifically to extinction of aversive CS/US associations. Future work should disentangle these hypotheses experimentally to determine which mechanisms are responsible for the observed effect.

Similar effects were observed for negative affect, although these effects were marginally significant, and no interactions were significant over and above the effects of positive affect.

72

Although our hypotheses specifically focused on associations between positive affect and dopaminergic prediction error as one potential mechanism for disrupted learning, it is possible that negative affect disrupts learning through similar means. For example, cognitive immunization theory posits that some beliefs become stronger when challenged, and this framework has been applied to explain why positive disconfirming information is frequently interpreted with caution among depressed subjects (Kube & Glombiewski, 2021; Kube & Rozenkrantz, 2021; Kube et al., 2020; Kube et al., 2021). In such cases, the violation of expectancy is often treated as an exception, or else the credibility of the experience is otherwise viewed with skepticism. Such a result is common during exposure therapy, as socially anxious subjects frequently believe that a negative US (e.g., social rejection) occurred even if there is no overt marker of this occurrence. It is possible that the cognitive immunization effects derive from associations with positive affect. In support of this notion, individuals with depression commonly engage in maladaptive positive emotion regulation strategies such as dampening, which has been shown to blunt positive affect (Raes et al., 2009; Kiken & Shook, 2014; Burr et al., 2017; Dunn et al., 2018). For example, the dampening appraisal "This is too good to be true" appears on the Responses to Positive Affect scale (Feldman et al., 2008), which supports the interpretation that cognitive immunization effects have some association with positive affect.

Alternatively, it is possible that positive affect is associated with dopaminergic or relief-based mechanisms of expectancy violation, whereas negative affect is associated with cognitive immunization processes that also interfere with expectancy violation. In support of this notion, negative mood induction has been shown to increase cognitive immunization effects among depressed subjects (Kube & Glombiewski, 2021). Additional research is needed to test this possible interpretation. Future work should investigate positive emotion regulation and its

73

relation to positive and negative affect in the context of exposure therapy for anxiety disorders. Likewise, future research should evaluate the extent to which cognitive immunization and dopaminergic prediction error mechanisms of expectancy violation derive from variations in specifically positive or negative affect.

The association between positive affect and expectancy violation fits within a broad literature linking positive affect to desirable learning outcomes. Heightened levels of positive affect are associated with more stable long-term fear extinction (Zbozinek et al., 2015; Zbozinek & Craske, 2017a), and these findings have been central to the development of positive affect interventions to treat anhedonia and co-occurring conditions such as anxiety disorders (Craske et al., 2019). It will be necessary for future research to elucidate how relative variations in positive affect impact *both* approach and avoidance of feared stimuli, which can inform efforts towards applying positive affect interventions while minimizing of avoidance behaviors in such paradigms.

Major depressive disorder is highly comorbid with social anxiety disorder and associated with increased symptom severity (Breteler et al., 2021). However, existing evidence is mixed regarding the potential deleterious impacts of major depression on treatment of anxiety disorders (Abramowitz & Landy, 2013; Rozen & Aderka, 2020). The present study found that low positive affect within-subjects, but not between-subjects, predicted poorer expectancy updating after exposures. This result raises the possibility that relative decreases in positive affect account for associations between major depression and exposure therapy effectiveness, whereas presence of a depression diagnosis or chronic anhedonia symptoms may not themselves confer a poorer prognosis. Additional research is needed to investigate this potential interpretation, particularly

among a larger sample of anxious individuals presenting with a high rate of comorbid major depression.

The findings from the present study have several clinical implications that warrant further investigation. Recent approaches to the treatment of anhedonia incorporate strategies to increase positive affect (e.g., Craske et al., 2019; for a review see Sandman et al., 2022), which may be relevant to incorporate into exposure therapy. First, we found that lower positive affect within- but not between-subjects was associated with smaller expectancy violation. This might suggest that delivering a positive mood induction (e.g., watching a brief amusing video, pleasant mental imagery, or another enjoyable activity; Zbozinek et al., 2015) at the beginning of session, particularly when a client reports lower than their usual level of positive affect on weekly symptom questionnaires, could be beneficial for subsequent exposure. Second, it may be useful to guide attention to rewarding aspects of the exposure during consolidation of extinction memories. For example, during the post-exposure discussion, in addition to asking about whether the feared outcome occurred (Craske, et al., 2022), therapists could ask clients about what was rewarding, enjoyable, or went well during the exposure. Lastly, to the extent that dampening appraisals (e.g., "this is too good to be true," "my streak of luck is going to end soon") block extinction learning, therapist may provide psychoeducation about dampening appraisals and guide clients to mindfully disengage from these thoughts and instead refocus attention (on rewarding elements from the prior exposure or to gather more evidence via repeated exposures). Additional research is needed to directly test these suggestions within the context of exposure therapy as a potential strategy for augmenting long-term fear extinction.

The present study involves several strengths. First, by evaluating effects in both inhibitory retrieval learning and fear reduction treatment conditions, the study explores the

effects of positive affect that may be specific to expectancy violation mechanisms compared with overall distress during exposure therapy. Second, the use of MLM enables the inclusion of individuals with incomplete datasets, a common feature of psychotherapeutic trials. Third, the deconstruction of within-subject and between-subject effects within this analytical framework highlighted the particular importance of within-subject variation in positive affect during exposure therapy, which may explain the mixed literature regarding treatments of comorbid depression and anxiety.

Several limitations constrain the generalizability of this study. First, these findings should be considered preliminary due to the comparatively small sample size and low statistical power. Future research should replicate these findings within a larger sample of anxious individuals experiencing a range of anhedonic symptoms. Relatedly, the present study was conducted specifically among individuals seeking treatment for social anxiety disorder. Future research should examine associations between anhedonia and expectancy violation across a wider range of anxiety symptoms (e.g., panic disorder), comorbid anhedonia symptoms (e.g., major depressive disorder), or symptom severity (e.g., non-clinical samples). Furthermore, the present study utilized data from three time-points for each individual. Future research may leverage computational modeling approaches within behavioral assays of fear extinction, which can further elucidate associations between anhedonia and expectancy violation across numerous trials and timepoints. Finally, the present study focused on positive affect, a core feature of anhedonia that does not capture all aspects of anhedonia as a symptom dimension. Future research should utilize comprehensive assessments of anhedonia symptoms to determine precise effects of anhedonia within inhibitory retrieval learning paradigms.

In sum, the present study highlights the potential importance of positive affect in moderating the learning mechanisms of exposure therapy. Additional research is needed to replicate and extend these findings within a broader population of subjects presenting with a wider range of anxiety symptoms.

TABLES AND FIGURES:

Table 1: *Demographics.*

	Inhibitory Retrieval Learning	Habituation	Statistic	*p*-value
N	32	32		
Sex (N, %)			$X^2 = .78$	.376
Female	23 (71.88%)	26 (81.25%)		
Male	9 (29.12%)	6 (18.75%)		
Age (M, SD)	26.68 (9.72)	25.67 (10.34)	t = -.34	.732
Race (N, %)			$X^2 = 5.79$	.671
Alaskan Native/Native American/Indigenous	1 (3.13%)	0 (0%)		.
Asian	9 (28.13%)	11 (34.38%)		
Black	0 (0%)	3 (9.38%)		
Latino(a)/Hispanic (Non-White)	4 (12.50%)	3 (9.38%)		
Latino(a)/Hispanic (White)	2 (6.25%)	2 (6.25%)		
Middle Eastern	2 (6.25%)	1 (3.13%)		
White	8 (33.33%)	8 (33.33%)		
Multiracial	5 (15.63%)	4 (12.50%)		
Did not report	1 (3.13%)	0 (0%)		
Current psychotropic medication use (N, %)			$X^2 = .28$	.599
Yes	12 (37.50%)	10 (31.25%)		
No	20 (62.50%)	22 (68.75%)		

Fig. 1: Anhedonia moderates the association between PreExp and PostExp in the Inhibitory Retrieval Learning condition (blue = high anhedonia, red = mean anhedonia, green = low anhedonia). **Fig. 1a** depicts the significant effect of anhedonia. **Fig. 1b** depicts no significant effect of anhedonia over and above negative affect. **Fig. 1c** depicts the significant effect of within-subject levels of anhedonia over and above negative affect.
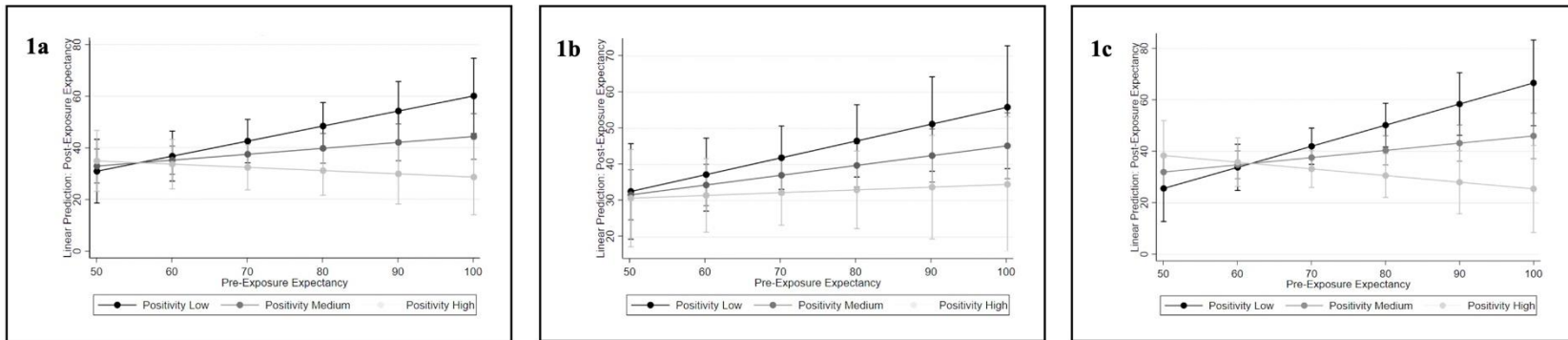

Fig. 2: Anhedonia does not moderate the association between PreFear and PostFear in the Fear Reduction condition (blue = high anhedonia, red = mean anhedonia, green = low anhedonia). **Fig. 2a** depicts no significant effect of anhedonia and **Fig. 2b** depicts no significant effect of within-subjects level of anhedonia.
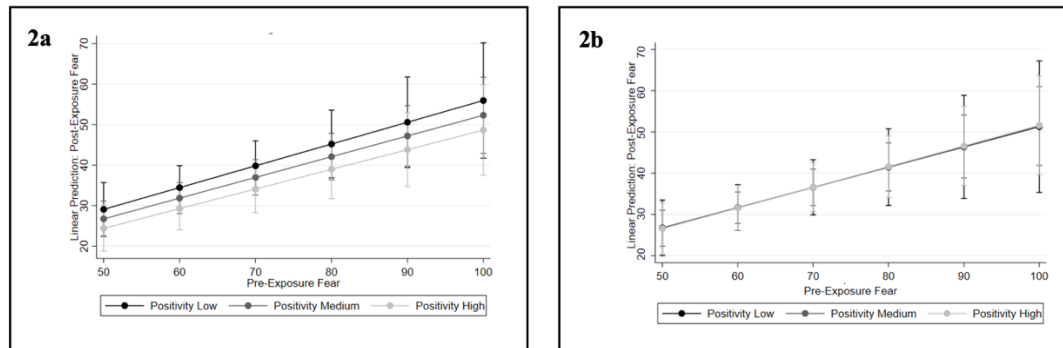
Fig. 3: Negative affect does not significantly moderate the association between PreExp and PostExp in the Inhibitory Retrieval Learning condition (blue = high negative affect, red = mean negative affect, green = low negative affect). **Fig. 3a** depicts the marginal effect of negative affect. **Fig. 3b** depicts no significant effect of negative affect after controlling for anhedonia.
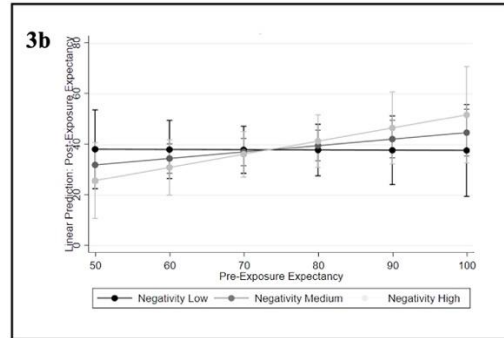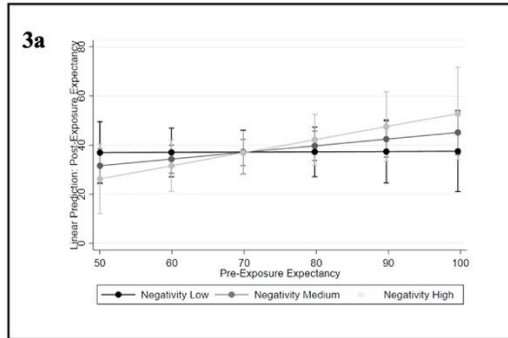
SUPPLEMENT:

Behavioral Tasks:

Several behavioral tasks were collected during the study but were not analyzed for the present report. These included:

- **Generalization of Learning:** This task is designed to measure generalization of acquired fear. It consists of the following types of stimuli: two CS+, one CS-, and eight generalization stimuli (GS's). This task consisted of three phases: acquisition, extinction and retest.

- **Public Speaking Task:** Participants completed a public speaking task in which they spoke for a short period of time in front of an audience. They were randomly assigned a topic from twelve possible topics and periodically reported their anxiety level throughout the task.

- **Implicit Pavlovian Task:** Participants completed a fifteen-minute computer task in which they were quickly presented a series of words. They were directed to press one of two buttons on a keyboard to indicate if it is a 'word' or a 'non-word', prompting the next word to flash on the screen. The second half of the task repeated the same procedure, but participants were presented two words in quick succession. Participants were instructed to focus their ratings on the second word only, as the first word is used as a distractor.

Measures of anhedonia and negative affect:

The following table lists the items that were included in the PANAS-P and PANAS-N scales

respectively:

PANAS-P	PANAS-N
Active	Afraid
Alert	Scared
Attentive	Frightened
Bold	Nervous
Cheerful	Jittery
Concentrating	Shaky
Confident	Hostile
Daring	Irritable
Delighted	Guilty
Determined	Ashamed
Energetic	Sad
Enthusiastic	Blue
Excited	Downhearted
Fearless	Alone
Happy	Lonely
Inspired	Distressed
Interested	Upset
Joyful	
Lively	
Proud	
Strong	

GENERAL DISCUSSION

The aim of this dissertation was to determine if anhedonia is associated with mechanisms of Pavlovian fear extinction and exposure therapy. Study 1 used a data-driven approach to highlight unique patterns of brain activity associated with anhedonia during fear extinction and extinction recall tasks. Study 2 evaluated the extent to which symptom-based anhedonia, as well as anhedonia-specific patterns of brain activity, were associated with physiological and behavioral indices of two sets of tasks: (1) fear acquisition, extinction, and extinction recall during an MRI scan, and (2) fear acquisition and generalization during a separate laboratory visit. Study 3 examined the clinical relevance of anhedonia symptoms in relation to the learning mechanisms of exposure therapy for social anxiety disorder.

Study 1

Study 1 aimed to characterize unique patterns of brain activity that were associated with anhedonia symptoms during fear extinction and extinction recall tasks. Prior work has shown that anhedonia is associated with increased activation of several brain regions during extinction learning, including the insular cortex, dorsal anterior cingulate cortex, and amygdala (Young et al., 2021). As anhedonia has been associated with different activation among a wide range of brain regions across a variety of tasks, we used data-driven methods to train a whole-brain decoder of anhedonia during fear extinction and extinction recall. We then used these decoders to predict anhedonia symptoms within an external validation dataset. In addition, we repeated analyses within individual brain networks and regions, highlighting candidate neurocircuitry that warrants additional research.

Results from Study 1 indicated that anhedonia was associated with unique patterns of brain activity during fear extinction that generalized to an external validation dataset. These

patterns were specific to anhedonia and did not characterize other symptom dimensions (i.e., fears or general distresss). Although interpretation of specific regions and networks should be interpreted with caution, these patterns included activity within the Cognitive Control, Default-Mode, Limbic, Salience, and Visual networks. By uncovering distinct neural processes associated with anhedonia during fear extinction, these results support the potential importance of understanding how anhedonia relates to Pavlovian processes. In particular, as contemporary models of exposure therapy are based in fear extinction and related processes (Pittig et al., 2016; Scheveneels et al., 2016; Fullana et al., 2020), such an understanding could inform innovations to augment exposure (Craske et al., 2019).

Study 2

Study 2 sought to directly build upon the results of Study 1. In Study 1, the anhedonia decoder involved regions such as the ventromedial prefrontal cortex and insular cortex, which are implicated in Pavlovian fear acquisition, extinction, extinction recall, and generalization processes. In this study, we examined if symptom-based anhedonia, as well as brain-predicted anhedonia, was further associated with self-report or psychophysiological indices of Pavlovian fear learning.

Results from Study 2 indicated that anhedonia, over and above other symptom dimensions characteristic of anxiety disorders, was associated with impaired fear acquisition and subsequently overgeneralization of acquired fear. Although overgeneralization of fear is associated with a range of anxiety disorders, results from this study support the notion that anhedonia symptoms, which are commonly found in depression but also associated with anxiety disorders (e.g., Kashdan, 2007; Abramovitch et al., 2014; Nawjin et al., 2015), may be central to these overgeneralization effects. If anhedonia is associated with disruptions in fear acquisition

84

and generalization, it is possible that anhedonia is further associated with other processes that rely on these mechanisms (Craske et al., 2008).

Study 3

Study 3 evaluated whether positive affect, a core feature of anhedonia, was associated with expectancy violation processes during a randomized controlled trial of exposure therapy. Individuals with social anxiety disorder completed exposure therapy that was designed to target either inhibitory retrieval learning or fear reduction. We measured both positive and negative affect during treatment to evaluate specificity of effects.

Results from Study 3 indicated that positive affect moderates expectancy violation during exposure therapy over and above negative affect. Specifically, individuals with low positive affect (i.e., elevated anhedonia symptoms) showed lower expectancy violation when CS-US predictions were the greatest (i.e., when a feared outcome was perceived as highly likely to occur). This result was specific to within-subject levels of positive affect (i.e., relative increases in poitivity compared to baseline), rather than between-subject levels of positive affect (i.e., an individual's level of positivity compared to other individuals). Furthermore, although negative affect showed a similar pattern of moderating expectancy violation, this effect was not significant over and above positive affect. Overall, our findings support the potential for positive affect interventions to augment the effectiveness of exposure therapy (Craske et al., 2019; Sandman & Craske, 2022), particularly among individuals experiencing relative increases in anhedonia symptoms. This area warrants additional research.

Strengths and Limitations

Strengths of this dissertation include the 1) wide range of Pavlovian fear learning paradigms (e.g., fear acquisition, extinction, extinction recall, generalization, *in vivo* exposure),

2) variety of psychobiological measurements (e.g., functional brain activity, psychophysiology, US contingency ratings, US expectancy ratings), 3) evaluation of anhedonia effects over and above other symptom dimensions (e.g., Fears, General Distress, Negative Affect), and 4) comparatively large sample sizes for Study 1 and Study 2. Findings from Study 1 and Study 2 provide mechanistic insights regarding associations between anhedonia and fear extinction or generalization. These insights directly inform the interpretation of Study 3, which has significant clinical value and may inform future innovations in the treatment of anxiety disorders.

The studies described in this dissertation should also be interpreted with caution, as anhedonia was associated with a variety of metrics across a range of different tasks. For example, anhedonia-specific patterns of brain activity were detected during fear extinction (Study 1), but anhedonia was not associated with physiological or self-reported indices of extinction (Study 2). Likewise, anhedonia effects during fear acquisition and generalization were evident using self-reported but not physiological indices of Pavlovian learning (Study 2). Although neural, physiological, and self-reported indices do not always concord, additional research is needed to clarify why anhedonia was associated with different metrics during different tasks. For example, it may be possible for prediction error learning to occur unconsciously (measured via neural or physiological indices) without conscious awareness (measured via self-reported US expectancy). Furthermore, anhedonia was associated with reduced acquisition of fear during only one of two behavioral tasks measuring fear acquisition. Although methodological differences during these tasks may explain this disagreement, additional research is needed to determine potential replication of acquisition effects. Finally, positive affect, a core feature of anhedonia, was associated with reduced prediction error during exposure therapy, but this study did not directly assess anhedonia symptoms (Study 3). A more

specific measurement of anhedonia symptoms may further strengthen the clinical application of these studies.

Conclusion

Collectively, this dissertation highlights an important role of anhedonia symptoms in association with Pavlovian fear learning processes that are central to the etiology, maintenance, and treatment of anxiety disorders. Anhedonia symptoms have received considerable attention in the literature on depression, but less is known about anhedonia in relation to anxiety. Taken together, these findings illuminate potential mechanisms of 1) how anhedonia symptoms become associated with anxiety disorders, including differences in the acquisition and generalization of learned fears, as well as 2) how these associations are maintained, including differences in the extinction of learned fears and the updating of predictions during exposure therapy.

The emerging role of anhedonia in relation to Pavlovian fear learning may have important implications for treatment. For example, interventions to increase positive affect may broadly boost the effectiveness of exposure-based treatments, which rely on Pavlovian learning as a core mechanism. Likewise, during the creation of individualized treatment plans, clinicians may assess anhedonia symptoms as a potential moderator of treatment effectiveness, even if anhedonia is not the primary treatment target. Furthermore, although anhedonia has most often been examined in the context of reward learning paradigms, the studies presented in this dissertation focus specifically on Pavlovian fear learning. It is possible that the findings from these studies are attributable to general disruptions in associative learning, which could influence transdiagnostic case conceptualizations and inform treatment of co-occurring mental health conditions (e.g., comorbid anxiety and depression). For example, anxious individuals with

elevated anhedonia symptoms may benefit from first engaging in treatments that directly target anhedonia before engaging in treatments that target fear processes. Additional research is needed to directly test these possibilities.

Overall, these studies converge on the potential importance of anhedonia during Pavlovian fear learning processes. The results from this dissertation support the notion that anhedonia is associated with disruptions to Pavlovian fear learning in ways that may impede the effectiveness of exposure therapy, the current gold standard for treatment of anxiety disorders. Additional research is needed to replicate these findings and to expand our understanding of how anhedonia relates to anxiety disorders. Future studies are needed to test adaptations of existing treatments based on these insights, such as the incorporation of positive affect interventions within treatments for anxiety disorders.

REFERENCES:

Abramovitch, A., Pizzagalli, D. A., Reuman, L., & Wilhelm, S. (2014). Anhedonia in obsessive-compulsive disorder: beyond comorbid depression. *Psychiatry research*, *216*(2), 223-229.

Abramowitz, J. S., & Landy, L. (2013). Treatment of Comorbid Depression. In *Handbook of Treating Variants and Complications in Anxiety Disorders* (pp. 243-254). Springer, New York, NY.

Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. (2019). *Exposure therapy for anxiety: Principles and practice*. Guilford Publications.

Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional reward processing in depression. *Current Opinion in Psychology*, *4*, 114-118.

Ahrens, L. M., Pauli, P., Reif, A., Mühlberger, A., Langs, G., Aalderink, T., & Wieser, M. J. (2016). Fear conditioning and stimulus generalization in patients with social anxiety disorder. *Journal of Anxiety Disorders*, *44*, 36-46.

Andersson, J. L., Jenkinson, M., & Smith, S. (2007). Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. *FMRIB Analysis Group of the University of Oxford*, *2*(1), e21.

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalization tools (ANTS). *Insight j*, *2*(365), 1-35.

Bandelow, B., Sagebiel, A., Belz, M., Görlich, Y., Michaelis, S., & Wedekind, D. (2018). Enduring effects of psychological treatments for anxiety disorders: meta-analysis of follow-up studies. *The British Journal of Psychiatry*, *212*(6), 333-338.

Battaglia, S., Garofalo, S., di Pellegrino, G., & Starita, F. (2020). Revaluing the role of vmPFC in the acquisition of Pavlovian threat conditioning in humans. *Journal of neuroscience*, *40*(44), 8491-8500.

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in neurosciences*, *26*(9), 507-513.

Breteler, J. K., Ikani, N., Becker, E. S., Spijker, J., & Hendriks, G. (2021). Comorbid depression and treatment of anxiety disorders, OCD, and PTSD: Diagnosis versus severity. *Journal of Affective Disorders*, *295*, 1005-1011.

Brolsma, S. C., Vassena, E., Vrijsen, J. N., Sescousse, G., Collard, R. M., van Eijndhoven, P. F., ... & Cools, R. (2021). Negative learning bias in depression revisited: Enhanced neural response to surprising reward across psychiatric disorders. *Biological psychiatry: cognitive neuroscience and neuroimaging*, *6*(3), 280-289.

Brown, T. A., & Barlow, D. H. (2009). A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological assessment*, *21*(3), 256.

Burr, L. A., Javiad, M., Jell, G., Werner-Seidler, A., & Dunn, B. D. (2017). Turning lemonade into lemons: Dampening appraisals reduce positive affect and increase negative affect during positive activity scheduling. *Behaviour Research and Therapy*, *91*, 91-101.

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depression and anxiety*, *35*(6), 502-514.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective

    responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality*

    and social psychology, *67*(2), 319-333.

Carver, C.S. (2009). Threat sensitivity, incentive sensitivity, and the experience of relief. *Journal*

    of personality. *77*(1), 125-38.

Connor, K. M., Kobak, K. A., Churchill, L. E., Katzelnick, D., & Davidson, J. R. (2001). Mini-

    SPIN: A brief screening assessment for generalized social anxiety disorder. *Depression and*

    anxiety, *14*(2), 137-140.

Craske, M. G., Farchione, T. J., Allen, L. B., Barrios, V., Stoyanova, M., & Rose, R. (2007).

    Cognitive behavioral therapy for panic disorder and comorbidity: more of the same or less of

    more?. *Behaviour Research and Therapy*, *45*(6), 1095-1109.

Craske, M. G., Hermans, D., & Vervliet, B. (2018). State-of-the-art and future directions for

    extinction as a translational model for fear and anxiety. Philosophical transactions of the

    Royal Society of London. *Series B, Biological sciences*, *373*(1742), 20170025.

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A.

    (2008). Optimizing inhibitory learning during exposure therapy. *Behaviour research and*

    therapy, *46*(1), 5-27.

Craske, M. G., Liao, B., Brown, L., & Vervliet, B. (2012). Role of inhibition in exposure

    therapy. *Journal of experimental psychopathology*, *3*(3), 322-345.

Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., & Dour, H. J. (2016). Treatment for

    anhedonia: a neuroscience driven approach. *Depression and anxiety*, *33*(10), 927-938.

Craske, M. G., Meuret, A. E., Ritz, T., Treanor, M., Dour, H., & Rosenfield, D. (2019). Positive affect treatment for depression and anxiety: A randomized clinical trial for a core feature of anhedonia. *Journal of Consulting and Clinical Psychology*, *87*(5), 457.

Craske, M. G., Stein, M. B., Eley, T. C., Milad, M. R., Holmes, A., Rapee, R. M., & Wittchen, H. U. (2017). Anxiety disorders. *Nature reviews. Disease primers*, *3*, 17024.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour research and therapy*, *58*, 10-23.

Craske, M. G., Treanor, M., Zbozinek, T. D., & Vervliet, B. (2022). Optimizing exposure therapy with an inhibitory retrieval approach and the OptEx Nexus. *Behaviour Research and Therapy*, 104069.

Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature reviews neuroscience*, *12*(8), 467-477.

Drevets, W. C. (2007). Orbitofrontal cortex function and structure in depression. *Annals of the New York Academy of Sciences*, *1121*(1), 499-527.

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., ... & Baas, J. M. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depression and anxiety*, *32*(4), 239-253.

Dunn, B. D., Burr, L. A., Smith, H. B., Hunt, A., Dadgostar, D., Dalglish, L., ... & Werner-Seidler, A. (2018). Turning gold into lead: Dampening appraisals reduce happiness and pleasantness and increase sadness during anticipation and recall of pleasant activities in the laboratory. *Behaviour Research and Therapy*, *107*, 19-33.

Dunsmoor, J. E., & Paz, R. (2015). Fear generalization and anxiety: behavioral and neural mechanisms. *Biological psychiatry*, *78*(5), 336-343.

Dunsmoor, J. E., Prince, S. E., Murty, V. P., Kragel, P. A., & LaBar, K. S. (2011). Neurobehavioral mechanisms of human fear generalization. *Neuroimage*, *55*(4), 1878-1888.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual review of clinical psychology*, *14*, 91-118.

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: systematic review and implications for anxiety disorder research. *Behavior therapy*, *46*(5), 561-582.

Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American journal of psychiatry*, *164*(10), 1476-1488.

Everaert, J., Bronstein, M. V., Cannon, T. D., & Joormann, J. (2018). Looking through tinted glasses: Depression and social anxiety are related to both interpretation biases and inflexible negative interpretations. *Clinical psychological science*, *6*(4), 517-528.

Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual of the Eysenck Personality Questionnaire (Junior and Adult). Hodder & Stoughton.

Feldman, G. C., Joormann, J., & Johnson, S. L. (2008). Responses to positive affect: A self-report measure of rumination and dampening. *Cognitive therapy and research*, *32*(4), 507–525.

First, M. B. (2014). Structured clinical interview for the DSM (SCID). *The encyclopedia of clinical psychology*, 1-6.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychological assessment*, *14*(4), 485.

Fryml, L. D., Pelic, C. G., Acierno, R., Tuerk, P., Yoder, M., Borckardt, J. J., ... & George, M. S. (2019). Exposure therapy and simultaneous repetitive transcranial magnetic stimulation: a controlled pilot trial for the treatment of posttraumatic stress disorder. *The journal of ECT*, *35*(1), 53-60.

Fullana, M. A., Dunsmoor, J. E., Schruers, K. R. J., Savage, H. S., Bach, D. R., & Harrison, B. J. (2020). Human fear conditioning: From neuroscience to the clinic. *Behaviour research and therapy*, *124*, 103528.

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Molecular psychiatry*, *21*(4), 500-508.

Gable, P. A., & Harmon-Jones, E. (2008). Approach-motivated positive affect reduces breadth of attention. *Psychological science*, *19*(5), 476-482.

Geer, J. H. (1965). The development of a scale to measure fear. *Behaviour research and therapy*, *3*(1), 45-53.

Gotlib, I. H., & Hamilton, J. P. (2008). Neuroimaging and depression: Current status and unresolved issues. *Current directions in psychological science*, *17*(2), 159-163.

Goulden, N., Khusnulina, A., Davis, N. J., Bracewell, R. M., Bokde, A. L., McNulty, J. P., & Mullins, P. G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. *Neuroimage*, *99*, 180-190.

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., … & Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, *134*(6), 1751-1764.

Graham, B. M., & Milad, M. R. (2011). The study of fear extinction: implications for anxiety disorders. *American journal of psychiatry*, *168*(12), 1255-1265.

Gray, M. A., Minati, L., Harrison, N. A., Gianaros, P. J., Napadow, V., & Critchley, H. D. (2009). Physiological recordings: basic concepts and implementation during functional magnetic resonance imaging. *Neuroimage*, *47*(3), 1105-1115.

Greco, J. A., & Liberzon, I. (2016). Neuroimaging of fear-associated learning. *Neuropsychopharmacology*, *41*(1), 320-334.

Greenberg, T., Carlson, J. M., Cha, J., Hajcak, G., & Mujica-Parodi, L. R. (2013). Neural reactivity tracks fear generalization gradients. *Biological psychology*, *92*(1), 2-8.

Greenberg, T., Carlson, J. M., Cha, J., Hajcak, G., & Mujica-Parodi, L. R. (2013). Ventromedial prefrontal cortex reactivity is altered in generalized anxiety disorder during fear generalization. *Depression and anxiety*, *30*(3), 242-250.

Greenberg, T., Chase, H. W., Almeida, J. R., Stiffler, R., Zevallos, C. R., Aslam, H. A., ... & Phillips, M. L. (2015). Moderation of the relationship between reward expectancy and prediction error-related ventral striatal reactivity by anhedonia in unmedicated major depressive disorder: findings from the EMBARC study. *American Journal of Psychiatry*, *172*(9), 881-891.

Hamer, R. M., & Simpson, P. M. (2009). Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials. *American journal of psychiatry*, *166*(6), 639-641.

Hamilton, J. P., Farmer, M., Fogelman, P., & Gotlib, I. H. (2015). Depressive rumination, the

   default-mode network, and the dark matter of clinical neuroscience. *Biological*

   psychiatry, *78*(4), 224-230.

Hamilton, J. P., Furman, D. J., Chang, C., Thomason, M. E., Dennis, E., & Gotlib, I. H. (2011).

   Default-mode and task-positive network activity in major depressive disorder: implications

   for adaptive and maladaptive rumination. *Biological psychiatry*, *70*(4), 327-333.

Hennings, A. C., McClay, M., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2020). Contextual

   reinstatement promotes extinction generalization in healthy adults but not PTSD.

   Neuropsychologia, *147*, 107573.

Hopwood, C. J., Bagby, R. M., Gralnick, T., Ro, E., Ruggero, C., Mullins-Sweatt, S., ... &

   Zimmermann, J. (2020). Integrating psychotherapy with the hierarchical taxonomy of

   psychopathology (HiTOP). *Journal of psychotherapy integration*, *30*(4), 477.

Huntsinger, J. R., Isbell, L. M., & Clore, G. L. (2014). The affective control of thought:

   malleable, not fixed. *Psychological review*, *121*(4), 600.

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto

   reinforcement learning: a behavioural meta-analysis. *Biology of mood & anxiety disorders*,

   3(1), 1-16.

Insel, T. R. (2014). The NIMH research domain criteria (RDoC) project: precision medicine for

   psychiatry. *American journal of psychiatry*, *171*(4), 395-397.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of

   brain images. *Medical image analysis*, *5*(2), 143-156.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the

robust and accurate linear registration and motion correction of brain

images. *Neuroimage*, *17*(2), 825-841.

Kaczkurkin, A. N., Burton, P. C., Chazin, S. M., Manbeck, A. B., Espensen-Sturges, T., Cooper,

S. E., … & Lissek, S. (2017). Neural substrates of overgeneralized conditioned fear in

PTSD. *American journal of psychiatry*, *174*(2), 125-134.

Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., & Pizzagalli, D. A. (2015). Large-scale

network dysfunction in major depressive disorder: a meta-analysis of resting-state functional

connectivity. *JAMA psychiatry*, *72*(6), 603-611.

Kalisch, R., Gerlicher, A. M., & Duvarci, S. (2019). A dopaminergic basis for fear extinction.

Trends in cognitive sciences, *23*(4), 274-277.

Karsen, E. F., Watts, B. V., & Holtzheimer, P. E. (2014). Review of the effectiveness of

transcranial magnetic stimulation for post-traumatic stress disorder. *Brain stimulation*, *7*(2),

151-157.

Kashdan, T. B. (2007). Social anxiety spectrum and diminished positive experiences: Theoretical

synthesis and meta-analysis. *Clinical psychology review*, *27*(3), 348-365.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted

maximum likelihood. *Biometrics*, 983-997.

Kiken, L. G., & Shook, N. J. (2014). Does mindfulness attenuate thoughts emphasizing

negativity, but not positivity?. *Journal of research in personality*, *53*, 22-30.

Koenigs, M., & Grafman, J. (2009). The functional neuroanatomy of depression: distinct roles

for ventromedial and dorsolateral prefrontal cortex. *Behavioural brain research*, *201*(2), 239-

243.

Kontopantelis, E., Springate, D. A., Parisi, R., & Reeves, D. (2016). Simulation-based power calculations for mixed effects modeling: ipdpower in Stata. *Journal of statistical software*, *74*(1), 1-25.

Kotov, R., Krueger, R. F., & Watson, D. (2018). A paradigm shift in psychiatric classification: the Hierarchical Taxonomy Of Psychopathology (HiTOP). *World psychiatry*, *17*(1), 24.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., ... & Eaton, N. R. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *Journal of abnormal psychology*, *126*(4), 454.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, *84*(3), 394-421.

Kriegeskorte, N., & Douglas, P. K. (2019). Interpreting encoding and decoding models. *Current opinion in neurobiology*, *55*, 167-179.

Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., ... & Zimmermann, J. (2018). Progress in achieving quantitative classification of psychopathology. *World psychiatry*, *17*(3), 282-293.

Kube, T., & Glombiewski, J. A. (2021). How Depressive Symptoms Hinder Positive Information Processing: An Experimental Study on the Interplay of Cognitive Immunisation and Negative Mood in the Context of Expectation Adjustment. *Cognitive therapy and research*, *45*(3), 517-528.

Kube, T., & Rozenkrantz, L. (2021). When beliefs face reality: An integrative review of belief updating in mental health and illness. *Perspectives on psychological science*, *16*(2), 247-274.

Kube, T., Kirchner, L., Lemmer, G., & Glombiewski, J. A. (2021). How the Discrepancy

Between Prior Expectations and New Information Influences Expectation Updating in

Depression—The Greater, the Better?. *Clinical psychological science*, 21677026211024644.

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted

cognitive processes in major depression: a predictive processing perspective. *Biological

psychiatry*, *87*(5), 388-398.

Kuhn, M., Höger, N., Feige, B., Blechert, J., Normann, C., & Nissen, C. (2014). Fear extinction

as a model for synaptic plasticity in major depressive disorder. *PloS one*, *9*(12), e115280.

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal

temporal difference reward-learning signals in major depression. *Brain*, *131*(8), 2084-2093.

Liebowitz, M. R. (1987). Social phobia. *Modern problems of pharmacopsychiatry*.

Liknaitzky, P., Smillie, L. D., & Allen, N. B. (2017). Out-of-the-blue: Depressive symptoms are

associated with deficits in processing inferential expectancy-violations using a novel

cognitive rigidity task. *Cognitive therapy and research*, *41*(5), 757-776.

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014).

Generalized anxiety disorder is associated with overgeneralization of classically conditioned

fear. *Biological psychiatry*, *75*(11), 909-915.

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine,

D. S. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behaviour

research and therapy*, *43*(11), 1391-1424.

Lissek, S., Rabin, S., Heller, R. E., Lukenbaugh, D., Geraci, M., Pine, D. S., & Grillon, C.

(2010). Overgeneralization of conditioned fear as a pathogenic marker of panic

disorder. *American Journal of Psychiatry*, *167*(1), 47-55.

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., ... & Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & biobehavioral reviews*, *77*, 247-285.

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans-Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience & biobehavioral reviews*, *80*, 703-728.

Lougheed, J. P., Vlisides-Henry, R. D., & Crowell, S. E. (2021). Advancing models and methods of emotional concordance. *Biological psychology*, *162*, 108112.

Lutkenhoff, E. S., Rosenberg, M., Chiang, J., Zhang, K., Pickard, J. D., Owen, A. M., & Monti, M. M. (2014). Optimized brain extraction for pathological brains (optiBET). *PloS one*, *9*(12), e115551.

Marin, M. F., Hammoud, M. Z., Klumpp, H., Simon, N. M., & Milad, M. R. (2020). Multimodal categorical and dimensional approaches to understanding threat conditioning and its extinction in individuals with anxiety disorders. *JAMA psychiatry*, *77*(6), 618-627.

Marin, M. F., Zsido, R. G., Song, H., Lasko, N. B., Killgore, W. D., Rauch, S. L., ... & Milad, M. R. (2017). Skin conductance responses and neural activations during fear conditioning and extinction recall across anxiety disorders. *JAMA psychiatry*, *74*(6), 622-631.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour research and therapy*, *36*(4), 455-470.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., ... & Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for

Brain Mapping (ICBM). *Philosophical transactions of the Royal Society of London. Series B: Biological sciences*, *356*(1412), 1293-1322.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain structure and function*, *214*(5-6), 655-667.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour research and therapy*, *28*(6), 487-495.

Milad, M. R., Quirk, G. J., Pitman, R. K., Orr, S. P., Fischl, B., & Rauch, S. L. (2007). A role for the human dorsal anterior cingulate cortex in fear expression. *Biological psychiatry*, *62*(10), 1191-1194.

Milad, M. R., Rosenbaum, B. L., & Simon, N. M. (2014). Neuroscience of fear extinction: implications for assessment and treatment of fear-based and anxiety related disorders. *Behaviour research and therapy*, *62*, 17-23.

Milad, M. R., Wright, C. I., Orr, S. P., Pitman, R. K., Quirk, G. J., & Rauch, S. L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological psychiatry*, *62*(5), 446-454.

Naragon-Gainey, K., Prenoveau, J. M., Brown, T. A., & Zinbarg, R. E. (2016). A comparison and integration of structural models of depression and anxiety in a clinical sample: Support for and validation of the tri-level model. *Journal of abnormal psychology*, *125*(7), 853.

Nawijn, L., van Zuiden, M., Frijling, J. L., Koch, S. B., Veltman, D. J., & Olff, M. (2015). Reward functioning in PTSD: a systematic review exploring the mechanisms underlying anhedonia. *Neuroscience & biobehavioral reviews*, *51*, 189-204.

Nissen, C., Holz, J., Blechert, J., Feige, B., Riemann, D., Voderholzer, U., & Normann, C. (2010). Learning as a model for neural plasticity in major depression. *Biological psychiatry*, *68*(6), 544-552.

Osuch, E. A., Benson, B. E., Luckenbaugh, D. A., Geraci, M., Post, R. M., & McCann, U. (2009). Repetitive TMS combined with exposure therapy for PTSD: a preliminary study. *Journal of anxiety disorders*, *23*(1), 54-59.

Papalini, S., Beckers, T., & Vervliet, B. (2020). Dopamine: from prediction error to psychotherapy. *Translational psychiatry*, *10*(1), 1-13.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *The journal of machine learning research*, *12*, 2825-2830.

Pittig, A. (2019). Incentive-based extinction of safety behaviors: Positive outcomes competing with aversive outcomes trigger fear-opposite action to prevent protection from fear extinction. *Behaviour research and therapy*, *121*, 103463.

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.

Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neuroscience & biobehavioral reviews*, *88*, 117-140.

Pittig, A., van den Berg, L., & Vervliet, B. (2016). The key role of extinction learning in anxiety disorders: behavioral strategies to enhance exposure-based treatments. *Current opinion in psychiatry*, *29*(1), 39-47.

Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: toward a synthesis and integrated model. *Annual review of clinical psychology*, *10*, 393-423.

Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008). Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *Journal of psychiatric research*, *43*(1), 76-87.

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of best practices for evidence for prediction: a review. *JAMA psychiatry*, *77*(5), 534-540.

Prenoveau, J. M., Zinbarg, R. E., Craske, M. G., Mineka, S., Griffith, J. W., & Epstein, A. M. (2010). Testing a hierarchical model of anxiety and depression in adolescents: A tri-level model. *Journal of anxiety disorders*, *24*(3), 334-344.

Raes, F., Daems, K., Feldman, G. C., Johnson, S. L., & Van Gucht, D. (2009). A psychometric evaluation of the Dutch version of the responses to positive affect questionnaire. *Psychologica Belgica*, *49*(4), 293.

Rapee, R. M., Craske, M. G., & Barlow, D. H. (1994). Assessment instrument for panic disorder that includes fear of sensation-producing activities: The Albany Panic and Phobia Questionnaire. *Anxiety*, *1*(3), 114-122.

Rauch, S. A., Eftekhari, A., & Ruzek, J. I. (2012). Review of exposure therapy: a gold standard for PTSD treatment. *Journal of rehabilitation research and development*, *49*(5), 679-688.

Rizvi, S. J., Pizzagalli, D. A., Sproule, B. A., & Kennedy, S. H. (2016). Assessing anhedonia in depression: Potentials and pitfalls. *Neuroscience & biobehavioral reviews*, *65*, 21-35.

Rømer Thomsen, K. (2015). Measuring anhedonia: impaired ability to pursue, experience, and learn about reward. *Frontiers in psychology*, *6*, 1409.

Rosenberg, B. M., Taschereau-Dumouchel, V., Lau, H., Young, K. S., Nusslock, R., Zinbarg, R. E., & Craske, M. G. (2022). A multivoxel pattern analysis of anhedonia during fear extinction–implications for safety learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.

Rozen, N., & Aderka, I. M. (2020). Do depressive symptoms affect the outcome of treatments for SAD? A meta analysis of randomized controlled trials. *Clinical psychology review*, *80*, 101874.

Ruggero, C. J., Kotov, R., Hopwood, C. J., First, M., Clark, L. A., Skodol, A. E., ... & Zimmermann, J. (2019). Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into clinical practice. *Journal of consulting and clinical psychology*, *87*(12), 1069.

Salinas-Hernández, X. I., Vogel, P., Betz, S., Kalisch, R., Sigurdsson, T., & Duvarci, S. (2018). Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive outcomes. *Elife*, *7*, e38818.

San Martín, C., Jacobs, B., & Vervliet, B. (2020). Further characterization of relief dynamics in the conditioning and generalization of avoidance: Effects of distress tolerance and intolerance of uncertainty. *Behaviour research and therapy*, *124*, 103526.

Sandman, C.F. & Craske, M.G. (2022). Psychological Treatments for Anhedonia. In D.A. Pizzagalli & A. Der-Avakian (Eds.). Anhedonia: Preclinical, translational, and clinical integration. *Current Topics in Behavioral Neurosciences*. Springer.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., ... & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral cortex*, 28(9), 3095-3114.

Scheveneels, S., Boddez, Y., Vervliet, B., & Hermans, D. (2016). The validity of laboratory-based treatment research: Bridging the gap between fear extinction and exposure treatment. *Behaviour research and therapy*, *86*, 87-94.

Sehlmeyer, C., Schöning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human fear conditioning and extinction in neuroimaging: a systematic review. *PloS one*, *4*(6), e5865.

Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., ... & Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences*, *106*(6), 1942-1947.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, *17*(3), 143-155.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*, S208-S219.

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences*, *105*(34), 12569-12574.

Taschereau-Dumouchel, V., Kawato, M., & Lau, H. (2019). Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Molecular psychiatry*, 1-13.

Thompson-Hollands, J., Sauer-Zavala, S., & Barlow, D. H. (2014). CBT and the future of personalized treatment: a proposal. *Depression and anxiety*, *31*(11), 909-911.

Tinoco-González, D., Fullana, M. A., Torrents-Rodas, D., Bonillo, A., Vervliet, B., Blasco, M.

J., ... & Torrubia, R. (2015). Conditioned fear acquisition and generalization in generalized

anxiety disorder. *Behavior therapy*, *46*(5), 627-639.

Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: lessons from

translational neuroscience. *Neuroscience & biobehavioral reviews*, *35*(3), 537-555.

Van Dis, E. A., Van Veen, S. C., Hagenaars, M. A., Batelaan, N. M., Bockting, C. L., Van Den

Heuvel, R. M., ... & Engelhard, I. M. (2020). Long-term outcomes of cognitive behavioral

therapy for anxiety-related disorders: a systematic review and meta-analysis. *JAMA*

psychiatry, *77*(3), 265-273.

van Meurs, B., Wiggert, N., Wicker, I., & Lissek, S. (2014). Maladaptive behavioral

consequences of conditioned fear-generalization: a pronounced, yet sparsely studied, feature

of anxiety pathology. *Behaviour research and therapy*, *57*, 29-37.

Vervliet, B., Lange, I., & Milad, M. R. (2017). Temporal dynamics of relief in avoidance

conditioning and fear extinction: Experimental validation and clinical relevance. *Behaviour*

research and therapy, *96*, 66-78.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A.

(1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of

anxiety and depression symptom scales. *Journal of abnormal psychology*, *104*(1), 3.

Wen, Z., Marin, M. F., Blackford, J. U., Chen, Z. S., & Milad, M. R. (2021). Fear-induced brain

activations distinguish anxious and trauma-exposed brains. *Translational psychiatry*, *11*(1),

1-10.

Willems, A. L., & Vervliet, B. (2021). When nothing matters: Assessing markers of expectancy

violation during omissions of threat. *Behaviour research and therapy*, *136*, 103764.

Wurst, C., Schiele, M. A., Stonawski, S., Weiß, C., Nitschke, F., Hommers, L., ... & Menke, A. (2021). Impaired fear learning and extinction, but not generalization, in anxious and non-anxious depression. *Journal of Psychiatric Research*, *135*, 294-301.

Yang, X., Fang, Y., Chen, H., Zhang, T., Yin, X., Man, J., ... & Lu, M. (2021). Global, regional and national burden of anxiety disorders from 1990 to 2019: results from the Global Burden of Disease Study 2019. *Epidemiology and psychiatric sciences, 30*.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Fischl, B. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, *106*(3), 1125-1165.

Young, K. S., Bookheimer, S. Y., Nusslock, R., Zinbarg, R. E., Damme, K. S. F., Chat, I. K., Kelley, N. J., Vinograd, M., Perez, M., Chen, K., Echiverri-Cohen, A., Craske, M.G. (2021). Dysregulation of threat neurocircuitry during fear extinction: the role of anhedonia. *Neuropsychopharmacology*, 1-8.

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of experimental social psychology*, 40(3), 424-431.

Zbozinek, T. D., & Craske, M. G. (2017a). Positive affect predicts less reacquisition of fear: relevance for long-term outcomes of exposure therapy. *Cognition and emotion*, *31*(4), 712-725.

Zbozinek, T. D., & Craske, M. G. (2017b). The role of positive affect in enhancing extinction learning and exposure therapy for anxiety disorders. *Journal of experimental psychopathology*, *8*(1), 13-39.

Zbozinek, T. D., Holmes, E. A., & Craske, M. G. (2015). The effect of positive mood induction on reducing reinstatement fear: Relevance for long term outcomes of exposure therapy. *Behaviour research and therapy*, *71*, 65-75.

Zhang, W. N., Chang, S. H., Guo, L. Y., Zhang, K. L., & Wang, J. (2013). The neural correlates of reward-related processing in major depressive disorder: a meta-analysis of functional magnetic resonance imaging studies. *Journal of affective disorders*, *151*(2), 531-539.

Zhang, X., Braun, U., Tost, H., & Bassett, D. S. (2020). Data-Driven Approaches to Neuroimaging Analysis to Enhance Psychiatric Diagnosis and Therapy. *Biological psychiatry: cognitive neuroscience and neuroimaging*, *5*(8), 780-790.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, *20*(1), 45-57.

Zhou, H. X., Chen, X., Shen, Y. Q., Li, L., Chen, N. X., Zhu, Z. C., ... & Yan, C. G. (2020). Rumination and the default mode network: Meta-analysis of brain imaging studies and implications for depression. *Neuroimage*, *206*, 116287.

Zimmerman, M., Coryell, W., Corenthal, C., & Wilson, S. (1986). A self-report scale to diagnose major depressive disorder. *Archives of general psychiatry*, *43*(11), 1076-1081.

Zinbarg, R. E., & Barlow, D. H. (1996). Structure of anxiety and the anxiety disorders: a hierarchical model. *Journal of abnormal psychology*, *105*(2), 181.