Clinical Psychological Science
1–33
DOI: 10.1177/21677026241240727

Mechanisms of Change in Exposure Therapy for Anxiety and Related Disorders: A Research Agenda

Kristen Benito[1,2], Andre Pittig[3], Jonathan Abramowitz[4], Joanna J. Arch[5], Denise Chavira[6], Rianne de Kleine[7], Alessandro S. De Nadai[8], Dirk Hermans[9], Stefan G. Hofmann[10], Jürgen Hoyer[11], Jonathan D. Huppert[12], Katharina Kircanski[13], Peter M. McEvoy[14,15], Heidi Meyer[16], Marie-H. Monfils[17], Santiago Papini[18], Winfried Rief[10], David Rosenfield[19], Eric A. Storch[20], Michael J. Telch[21], Michael W. Otto[22], Jasper A. J. Smits[21], and for the Exposure Therapy Consortium

[1]Psychiatry and Human Behavior, Warren Alpert Medical School, Brown University; [2]Pediatric Anxiety Research Center, Emma Pendleton Bradley Hospital, East Providence, Rhode Island; [3]Georg-Elias-Mueller-Institute of Psychology, Georg-August-Universität Göttingen; [4]Department of Psychology, The University of North Carolina at Chapel Hill; [5]Department of Psychology and Neuroscience, University of Colorado at Boulder; [6]Department of Psychology, University of California Los Angeles; [7]Institute of Psychology, Leiden University; [8]Department of Psychiatry, McLean Hospital and Harvard Medical School; [9]Center for the Psychology of Learning and Experimental Psychopathology, KU Leuven; [10]Department of Psychology, Philipps-Universität Marburg; [11]Clinical Psychology and Psychotherapy, Technische Universität Dresden; [12]Department of Psychology, The Hebrew University of Jerusalem; [13]Emotion and Development Branch, National Institute of Mental Health, Bethesda, Maryland; [14]Centre for Clinical Interventions, Northbridge, Australia; [15]School of Psychology, Curtin University; [16]Psychological and Brain Sciences, Boston University; [17]Research and Operations, The University of Texas at Austin; [18]Division of Research, Kaiser Permanente, Oakland, California; [19]Department of Psychology, Southern Methodist University; [20]Psychiatry and Behavioral Sciences, Baylor College of Medicine; [21]Department of Psychology, University of Texas at Austin; and [22]Department of Psychology, Boston University

Abstract

Anxiety and related disorders are a significant public-health burden with rising prevalence in the wake of the COVID-19 pandemic. As demand for effective anxiety treatment increases, so too does the need for strategies to bolster treatment outcomes. Research on the mechanisms of exposure therapy, the frontline behavioral treatment, will be critically important for optimizing clinical outcomes. We outline an initial agenda for future research on the mechanisms of change of exposure therapy, developed in collaboration with a large international team of researchers through the Exposure Therapy Consortium. Key questions and recommendations for future research focus on four priority areas: conceptualization, measurement, study design/analysis, and individual/contextual differences. Rising to the challenge of addressing these questions will require coordinated action and availability of centralized tools that can be used across trials, settings, and research groups.

Keywords

anxiety, treatment, exposure therapy, mechanisms, measurement

Received 9/6/23; Revision accepted 3/5/24

As the public-health impact of anxiety and its disorders continues to rise (U.S. Preventive Services Task Force, 2020), so does the demand for intervention science.

Corresponding Author:
Kristen Benito, Warren Alpert Medical School, Psychiatry and Human Behavior, Brown University
Email: Kristen_benito@brown.edu

Among important aims is improving the efficacy and uptake of evidence-based treatments for these disorders. To this end, we established the Exposure Therapy Consortium (ETC; https://exposure.la.utexas.edu/), an international collaboration among academic investigators that is designed to allow investigations of exposure-therapy mechanisms and outcomes to occur at scale and at relatively low cost/effort (Smits et al., in press). The scale of the ETC collaboration and potential for rapid replication of findings across sites can provide infrastructure to address research questions relevant to this first-line intervention more quickly and with more confidence.

Recognizing that mechanistic research facilitates the delivery and outcomes of interventions, we have begun to develop an agenda for research on exposure-therapy mechanisms of change. The process for developing this agenda involved a series of meetings organized and attended by a subset of this article's authors (K. Benito, A. Pittig, M. W. Otto, J. A. J. Smits) that focused on review of empirical articles and relevant theoretical accounts. Keeping in mind clinical application and acknowledging that agendas like these cannot be complete and will evolve over time, we generated a set of key questions assigned to one of four priority areas: (a) conceptualization, (b) measurement, (c) study design and data analysis, and (d) individual and contextual differences. We then reached out to colleagues who bring relevant experience and expertise and asked them to help shape this agenda. We describe this initial agenda, hoping that it can spur initiatives for larger-scale collaborative research. We provide a brief review of extant literature for background and justification for the proposed priority areas and key research questions.

Exposure-Therapy Efficacy

Exposure therapy involves repeated confrontation to feared cues or contexts without engaging in safety behaviors (i.e., avoidance, rituals). Feared cues and contexts can be external (e.g., agoraphobic situations, people, animals) or internal (e.g., emotions, bodily sensations, thoughts, images, memories), and exposure practice can vary in modality (e.g., in vivo, imaginal, virtual reality, interoceptive). Likewise, safety and defensive behaviors vary substantially (e.g., in relation to threat imminence; Hamm, 2020). Exposure is a primary component of established cognitive-behavioral interventions for anxiety, obsessive compulsive, and trauma and stressor-related disorders (Abramowitz et al., 2019). Exposure-based treatments have consistently fared well in clinical trials, outperforming placebo and active psychotherapy controls and rivaling or outperforming pharmacological interventions (Carpenter et al., 2018).

The observation that a significant minority of patients receiving exposure-based treatment do not respond, remain symptomatic, or experience return of fear has promoted research on combination and augmentation strategies. Approaches to developing and testing these strategies have varied and included adding a single anxiolytic strategy (e.g., benzodiazepine) and combining exposure therapy with strategies to further engage putative mechanisms (e.g., cognitive enhancers). These efforts have yielded mixed successes (Tuerk, 2014) and often have seen initial promising findings weaken or disappear in follow-up studies (Rosenfield et al., 2019). One reaction to this pattern is increasing emphasis on mechanisms of change, or renewed focus on understanding how exposure works. Exposure therapy benefits greatly from a rich body of experimental work with animals and humans (McNally, 2007), and studies have increasingly focused on translating mechanistic knowledge from these settings into clinical treatment. This has produced important advances, but gaps remain, and efforts to leverage mechanistic knowledge have not realized full potential for improving clinical outcomes.

Conceptual Considerations

In general, a mechanism of change can be defined as the processes through which treatment (e.g., exposure) produces the change (see Kazdin, 2007). Theories of the maintenance and recovery of fear, anxiety, and their disorders have identified multiple potential mechanisms of exposure-therapy efficacy (e.g., emotional-processing theory [EPT], Foa & Kozak, 1986; inhibitory learning [IL] model, Craske et al., 2008; acceptance and commitment [ACT] model, Twohig et al., 2015; self-efficacy theory, Bandura, 1988). Table 1 provides brief descriptions of common candidate mechanisms in research to date (for a recent review, see Knowles & Tolin, 2022). Keeping with the aims of this article, we do not provide a comprehensive review of these theories and mechanisms but instead share some conceptual observations that can provide direction for building on earlier work. Table 2 provides an overview of identified challenges related to this and other priority areas and related recommendations for a coordinated research agenda and action steps for individual researchers.

Overlap among theorized mechanisms

Most research to date has aimed to identify single theory-based mechanisms and/or compare theoretical approaches against one another. However, discussing

Table 1. Candidate Mechanisms of Exposure Therapy

Candidate mechanism	Brief description	Clinical example
Fear activation/emotional engagement	Elevation of distress, fear, or anxiety; typically at the start or peak of an individual exposure exercise	Initial exposure to a phobic stimulus (e.g., a spider, bodily sensations, thoughts, images, trauma memories) results in elevation of fear
Counterconditioning	Pairing a fear-eliciting stimulus with a positive stimulus to replace the fear response with a positive response	Receiving sweets whenever a phobic stimulus is presented
Within-session extinction	Decline of fear response within a single exposure exercise (often referred to as within-session "habituation")	A visit to a grocery store (e.g., agoraphobia) is continued until the patient experiences a decrease of physiological and/or subjective fear
Between-session extinction	Decline of peak fear response across multiple exposure exercises and/or sessions (often referred to as between-session "habituation")	Repeated exposure to the same grocery store until peak fear levels start to decline
Change in maladaptive fear cognition	Learning that a feared stimulus or situation is safer than previously believed	The visit to the grocery store is continued until patients learn that they will not experience a heart attack
Expectancy violation	Exposure exercises should violate threat expectancy (anticipation of a specific threat) through learning that the threat does not actually occur during exposure	A patient expects that a panic attack will result in a heart attack during exposure practice in a grocery store and experiences "surprise" when this does not actually occur
Expectancy change	Reduced threat expectancy across multiple exposure exercises and/or sessions	After exposure, a patient no longer expects that a heart attack will occur when having a panic attack in a grocery store
Distress tolerance	Repeated exposure results in higher distress tolerance (capacity to experience and withstand negative psychological and physiological experiences)	Repeated exposure to an aversive stimulus such as shortness of breath (in interoceptive exposure for panic sensations) increases a patient's ability to tolerate the discomfort associated with those sensations
Instrumental learning	Exposure practice changes behavioral responses from avoidance to approach behavior through positive and negative reinforcement	During exposure practice, a patient approaches a grocery store, which is followed by social reinforcement, feelings of pride, and/or absence of a feared outcome
Experiential avoidance	Exposure reduces avoidance/escape of aversive internal states and associated contexts, thus facilitating a more flexible and approach-oriented stance	Following exposure to crowds, patients are better able to attend social events (e.g., parties), thus better able to discern the behavioral choices they have in such situations rather than simply avoid them
Self-efficacy	Exposure strengthens specific beliefs related to the ability to cope with anxiety-provoking stimuli	After public speaking exposure practice, patients realize they are more capable of giving a speech and coping with distress during the speech than they had imagined
Metacognition	Exposure weakens beliefs about the benefits of dysfunctional strategies (e.g., repetitive negative thinking) to prevent threats and also weakens beliefs about repetitive negative thinking being uncontrollable and dangerous	Patients believe that their worry helps them to be prepared and to prevent bad things from happening but then worries that their habitual worrying itself is harmful and uncontrollable; these beliefs shift over time with exposure practice
Therapeutic relationship	Conducting therapist-guided exposure strengthens the therapeutic relationship, in turn improving treatment outcome	By repeatedly doing exposures together and having the therapist model them first, the patient feels more connected and trusting toward the therapist

Table 2. Priority Areas for Research on Mechanisms of Change in Exposure Therapy

Priority area	Current challenges	Research agenda *Areas of focus for the field*	Action steps *Concrete actions for individual researchers*
Conceptual considerations			
	• Focus on theoretical differences obscures similarities across overlapping constructs. • Different findings likely arise from emphasis on different stages (e.g., initial learning, consolidation, retrieval) or aspects (e.g., cognitive, emotional, behavioral) of the same underlying processes. • The distinctions between mechanisms of change versus (a) processes that moderate or facilitate them and (b) readout measures that signal them are often unclear. • There may be multiple readout measures for the same mechanism, which could differ across individuals.	• Work toward integrated mechanistic models that account for overlapping constructs and distinguish mechanisms versus their facilitators/moderators and readout measures. • As described in Development of consistent Terminology, Definitions and Methods section, test construct validity against conceptually overlapping measures. • Examine fine-grained temporal changes of constructs *in relation to one another*.	• Include substantive discussion of conceptual overlap in all articles (e.g., implications for study design, interpreting results). • To the fullest extent possible, measure an array of distinct, plausible mechanisms at multiple timepoints (see also Outcome Measurement/ Analysis in Clinical Trials section) in each study. • Design studies with the understanding that mechanisms and strategies that engage them (facilitators) may not always match as theorized. • Collaborate with investigators who have expertise in "competing" theories.
Measurement			
Psychometric properties of mechanistic measures	• The most commonly used mechanistic measures have very limited psychometric support. • Construct definitions, measurement procedures, and calculation methods are highly variable, making it difficult to compare findings across studies. • Measurement error from multiple sources is likely and can greatly reduce power.	• Work toward consistent operational definitions for mechanistic constructs. • Establish reliability and construct validity of mechanistic measures. • Enhance the ontological study of extinction with investigations of relationships between measures and the construct of extinction learning. • Focus efforts on practical measures to support forward translation. • Determine optimal measurement timing and calculation methods. • Establish best practices for reducing or accounting for measurement error introduced by therapy content.	• Include psychometric data (reliability, correlations with other measures, predictive validity) for mechanistic measures in publications and/or public repositories. • Create and publicly share detailed, reproducible, written procedures related to mechanism measurement (e.g., standard operating procedures), including rater training/reliability procedures. • Deploy *a priori* and/or post hoc methods to reduce the impact of data pollution (see Psychometric Properties of Measures section). • Increase measurement frequency to reduce error. • Follow recommendations for the reporting of mechanistic behavioral research (Birk et al., 2023).

(continued)

Table 2. *(continued)*

Priority area	Current challenges	Research agenda *Areas of focus for the field*	Action steps *Concrete actions for individual researchers*
Multimodal measurement	• Self-report, physiological, and behavioral measures typically do not converge; degree of convergence may vary by measure, context, and individual differences (e.g., age, insight).	• Design studies to explicitly test measure convergence and predictive validity. • Examine individual and contextual factors that influence measure convergence and predictive validity. • Establish best practices for multimodal measurement that minimizes burden with maximum potency.	• Incorporate multimodal measurement into clinical trials. • Include psychometric data (reliability, correlations with other measures, predictive validity) for multimodal measures in publications and/or public repositories.
Psychometric properties of laboratory models	• Commonly used stimuli have weak face validity. • Recent fear acquisition used in lab models may differ from long-standing, complex fear responses associated with clinical symptoms. • Limited data are available to support predictive validity with treatment outcomes.	• Develop and test more "life-like" stimuli. • Develop and test models that incorporate avoidance behavior. • Conduct larger tests of predictive validity in clinical trials. • Examine subgroup differences in validity (e.g., based on level of impairment, symptom duration, or exposure to chronic stressors).	• Develop close collaborations between clinical researchers and experimental researchers; doing so early and often will enhance validity of lab models.
Advances from models with non-human animals	• Early translational work with animals can inform novel approaches by identifying promising windows for measurement/intervention, novel mechanisms, and new measurement/treatment approaches.	• Innovative paradigms with nonhuman animals should continue to inform measurement strategies and initial trials with humans. • Human clinical studies are needed to examine promising interventions that target promising mechanisms from animal studies (e.g., retrieval-extinction targeting reconsolidation blockade).	• Develop close collaborations between nonhuman animal researchers and clinical researchers; do so early and often to enhance validity of animal models and speed translation into clinical trials.

(continued)

Table 2. *(continued)*

Priority area	Current challenges	Research agenda *Areas of focus for the field*	Action steps *Concrete actions for individual researchers*
Study design and analysis			
Experiment and clinical-trial design	• Laboratory experiments that manipulate mechanism using brief exposure protocols are efficient and well suited for isolating mechanisms of interest. • Full clinical trials are needed to test mechanisms in conjunction with efficacy and examine generalizability of earlier translational findings (e.g., to treatment-seeking clinical populations, multisession exposure). • Limited evidence is available about mechanism-relevant strategies that may have been used in original exposure efficacy trials. • It is difficult to design trials that compare mechanistic theories given conceptual overlap.	• Laboratory experiments should continue to be used for initial mechanism tests and to identify the most promising strategies for engaging mechanisms. • Rigorously designed clinical trials should specify, test, and confirm mechanisms and clinical efficacy before making recommendations for clinical practice. • Secondary analysis of exposure delivery in past trials is needed to inform design of "exposure as usual" comparison conditions for future trials. • Nontraditional (e.g., full factorial in multiphase optimization strategy framework) and adaptive designs (e.g., sequential multiple assignment randomized trial) can be used to test mechanisms and tailored treatment decisions related to mechanisms. • Innovative single-subject designs (e.g., multiple baseline, alternating treatment, knock-out) should be used to isolate the contribution of specific exposure strategies to mechanistic change (see also Machine-Learning Applications section).	• Avoid making recommendations for practice change without rigorous clinical-efficacy data (e.g., from masked randomized controlled trials with clinical, treatment-seeking populations receiving a full course of treatment). • Share detailed treatment manuals, fidelity measures, and related therapist-training procedures in supplemental materials and/or public repositories. • Design trials to test promising strategies under an integrated theoretical framework rather than to compare models against one another (see Conceptual Considerations section).
Measuring and accounting for exposure-delivery variation	• Exposure-delivery approach and dose vary widely, across and even within individual studies; measuring this is critical for understanding proximal "triggers" of mechanistic change, interpreting the validity of mechanistic measures (see Psychometric Properties of Measures section), and statistical modeling. • Adherence/fidelity measures are commonly used in clinical trials but rarely have psychometric support, lack predictive value, and fail to capture nuances relevant for mechanism.	• Develop and test detailed measures of exposure dose and specific aspects of setup/selection, delivery, and postprocessing; these should use clear operational definitions for replicability. • Studies of exposure in group format should incorporate multilevel modeling and other novel approaches (e.g., social relation, one-with-many modeling) to disentangle group-level dynamics. • Trials should incorporate detailed measurement of homework exposure and deploy novel methods (e.g., complier average causal effect) to account for individual differences in homework completion.	• Report detailed descriptions of methods in articles or extended supplementary material. • Create and publicly share detailed treatment manuals. • Deploy detailed measures of exposure dose and delivery in all clinical trials; report related data (including psychometric data) in articles and/or supplementary material. • Use clear operational definitions when describing delivery techniques; avoid using labels that infer mechanism or fail to convey what the technique "looks like."

(continued)

Table 2. *(continued)*

Priority area	Current challenges	Research agenda *Areas of focus for the field*	Action steps *Concrete actions for individual researchers*
Rigor and reproducibility	• Variation in exposure delivery, quality monitoring, measurement procedures, and/or allegiance to a particular mechanism likely contribute to discrepant mechanistic findings across studies. • Subjective measures of mechanism (e.g., self-report) are commonly used and least burdensome but increase risk for bias; this is exacerbated without masking.	• Develop and validate low-burden objective measures (e.g., using wearables or brief computerized tasks). • Optimize the utility of publicly available platforms for quickly locating and sharing "tools."	• Develop and publicly share detailed "tools" that support measurement and treatment procedures, including clinician training/monitoring (e.g., rating manuals, standard operating procedures). • Preregister hypotheses when possible. • Increase sample size in clinical trials to the fullest extent possible (e.g., by participating in multisite research consortiums) to offset decreased power resulting from measurement error. • Design clinical trials to be masked and use independently rated primary outcomes, particularly if using subjective measures of mechanism.
Outcome measurement/ analysis in clinical trials	• In traditional randomized controlled trials, the mediator is not randomized, leading to overestimation of a causal effect on outcome. • Relationships between mediators and outcomes can be driven by a third variable (i.e., other plausible mediators).	• Deploy enhanced cross-lagged panel models (e.g., random-intercept, latent change score) to most accurately assess the causal effect of a mediator.	• Assess candidate mediators and outcomes frequently. • Power trials to investigate all possible mechanisms and their interactions rather than focusing only on outcomes. • Assess delivery and/or contextual factors that may affect outcomes (e.g., significant life events, see also Measuring and Accounting for Exposure-Delivery Variation section). • Collaborate with researchers bringing relevant statistical expertise during both design and analysis.

(continued)

Table 2. *(continued)*

Priority area	Current challenges	Research agenda *Areas of focus for the field*	Action steps *Concrete actions for individual researchers*
Machine-learning applications	• Assumptions about mechanistic processes are typically made on the within-subjects level, but most analyses focus on the between-subjects level. • Mediation is often tested in linear models; nonlinear relationships may be present, but traditional modeling risks model misspecification. • Exposure use/techniques are not always randomized (e.g., homework exposures). • Machine-learning models are more flexible, make fewer assumptions, and are better suited for handling multicollinearity versus enhanced cross-lag panel models (see Outcome Measurement/Analysis in Clinical Trials section).	• Deploy supervised machine learning to simultaneously integrate many predictors and nonlinear variable relationships. • Deploy unsupervised machine learning to identify patient "clusters" (e.g., those that respond best to specific exposure conditions).	• Consider machine-learning-augmented inverse propensity score weighting with doubly robust estimation to accurately estimate causal effects. • Consider machine-learning methods based on mixture modeling, which are designed *a priori* to address measurement error. • Collaborate with researchers bringing relevant statistical expertise during both design and analysis.
Individual and contextual differences			
Magnitude/type of emotion elicited during exposure	• Exposure often targets emotions other than or in addition to fear (e.g., disgust, incompleteness); pattern of change and clinical outcomes can differ for these. • Emotional distress can be amplified or qualitatively altered by individual factors (e.g., anxiety sensitivity).	• Studies should measure multiple affective states to enable comparison of change patterns and outcomes by type of emotion. • Develop targeted augmentation and/or exposure-delivery strategies for different emotions.	• Design studies to measure a range of negative emotions across time; also consider measuring positive emotions. • Include measures of individual factors relevant for elicited emotions (e.g., anxiety sensitivity). • Increase sample sizes to the fullest extent possible (e.g., by participating in multisite research consortiums) to power subgroup analyses related to emotion.

(continued)

Table 2. *(continued)*

Priority area	Current challenges	Research agenda *Areas of focus for the field*	Action steps *Concrete actions for individual researchers*
Developmental considerations	• Exposure is efficacious for adolescents, but response may be less robust and enduring versus that for younger children and adults; adolescents experience normative deficits in extinction learning. • It is not known whether the higher response rate in younger children relates to true developmental differences or compensatory strategies (e.g., greater involvement of caregivers). • Some treatment packages for youths include less exposure in favor of other cognitive-behavioral therapy ingredients (e.g., relaxation); higher proportion of exposure strongly predicts better outcomes. • Younger patients have more difficulty articulating fear cognitions and/or understanding differences between mechanistic constructs (e.g., fear vs. expectancy).	• Longitudinal studies are needed to understand changes in extinction learning across development. • Compare patterns of mechanistic change and clinical outcomes across age cohorts in clinical trials, particularly when holding treatment components constant. • Conduct psychometric examination of mechanistic measures across age cohorts. • Establish "best practices" for streamlined use of self-report measures by age. • For adolescents, test intervention strategies that circumvent reliance on the prefrontal cortex (e.g., targeted exposure-augmentation strategies; emphasize rewarding stimuli).	• When possible, include objective measures (e.g., psychophysiology, observer ratings) as part of multimodal mechanism measurement for youths. • Include detailed measures of treatment delivery, including developmental adaptations and caregiver/peer involvement. • Increase sample size in clinical trials to the fullest extent possible (e.g., by participating in multisite research consortiums) to power tests across age cohorts.
Considerations for culturally diverse populations	• Exposure appears to be effective for diverse populations, but studies have been limited by small sample sizes and inability to test for subgroup effects (e.g., based on race, ethnicity, intersectionality). • Very few studies have examined mechanistic differences in marginalized individuals. • Factors such as discrimination and race can influence key mechanistic constructs (e.g., experiential avoidance, response to threat). • Patient reports of symptoms and mechanistic changes differ across cultures (e.g., endorsement of somatic vs. cognitive distress).	• Design lab-based and clinical studies to explicitly test mechanisms among marginalized individuals. • Conduct psychometric examination of mechanistic and symptom measures across cultural groups and languages. • Identify optimal strategies for providing culturally responsive exposure therapy (i.e., tailored for cultural context). • Examine elements of culturally responsive care in relation to exposure mechanisms.	• All studies should include measures that characterize historically marginalized groups (e.g., based on race, gender, sexual orientation, disability, economic status) *and* their experiences (e.g., discrimination, negative life events). • Increase sample size in clinical trials to the fullest extent possible (e.g., by participating in multisite research consortiums) to detect subgroup differences for minoritized individuals.

and testing different theory-based mechanisms as "competitors" risks obscuring the considerable overlap across theories, relationships among different mechanisms, and links with earlier translational work—and does not align with the idea that exposure most likely works via parallel and successive mechanisms (Himle, 2015). First, variable use of terminology may foster artificial boundaries between different approaches and hinder translational progress. For example, "habituation" is often discussed as a mechanism, referring to the reduction of fear responses within or between exposure exercises. However, in basic research, habituation is strictly defined as a short-term, implicit adaptation in the form of diminished responding to an inconsequential stimulus not due to fatigue (a nonassociative learning process; Rankin et al., 2009). This definition does not align with the idea of associative learning presumed to underlie exposure according to multiple theories, which is more accurately described as "extinction" (i.e., decrement of fear behaviors when associative relationships underlying original fear learning have changed; Lovibond, 2004; Table 1). The field will benefit from establishing consistent, precise terminology and concrete operational definitions for both mechanistic constructs and related exposure-optimization strategies (see Psychometric Properties of Measures and Rigor and Reproducibility in Clinical Trials sections).

Second, much of the research on mechanisms of change has been narrow in that studies have typically focused on testing one phase of treatment (e.g., before, during, or after in-session exposure; between-session exposure homework). This misses an important opportunity to understand relations among optimization strategies used across different phases of mechanistic learning, which include initial learning (during exposure; Benito et al., 2018), consolidation (in the hours after exposure; Kredlow et al., 2018), retrieval (during repeated exposures; Craske et al., 2014), and generalization (when encountering new stimuli and/or contexts; Richter et al., 2021). When considering all phases, there is an opportunity to fully optimize outcomes by deploying the most promising strategies targeting each.

Third, many constructs typically associated with different theories (e.g., EPT, IL, ACT) share a high degree of conceptual overlap and may not be distinct (i.e., represent measured variables of a shared latent construct). A case in point is extinction and expectancy violation. The reduction of fear is incongruent with the idea that fear will be of infinite intensity or duration (a common prediction in anxiety). How does this differ from expectancy violation or change? It seems plausible that changes in fear and expectancy represent emotional and cognitive aspects of the same underlying mechanism of change (e.g., Hofmann, 2008). Similar concerns have been raised regarding overlap between self-efficacy and outcome expectations (Breuninger et al., 2019) and between distress tolerance, experiential avoidance, and emotional engagement (Schloss & Haaga, 2011). In addition, many of these models are consistent with neurocognitive, computational models of reinforcement learning contending that prediction errors (online predictions followed by feedback) are the basis of most learning, including fear extinction (Koban et al., 2017; Song et al., 2022). In some cases, constructs with different labels share nearly identical definitions (e.g., change in maladaptive fear cognition, expectancy change, threat reappraisal) but have become associated with seemingly different clinical approaches (Himle, 2015). A more comprehensive, integrated theoretical conceptualization is needed to account for shared core features across existing theories—and will aid in accelerating optimization research with the greatest potential to produce clinical gains. Understanding shared and unique core features may also help to pinpoint whether these features qualify as common mechanism of change or whether some are unique to specific phenomena of anxiety disorders (e.g., see discussion on fear vs. anxiety; Öhman, 2008).

Mechanism or facilitator

Some candidate mechanisms appear to be important for therapeutic change but are not causes of symptom improvement. Consider, for example, emotional engagement/activation. Patients show emotional activation in everyday life without lasting change in anxiety symptoms. Brief emotional activation does not qualify as a mechanism of change because it does not directly cause anxiety reduction. However, it may be a "facilitator" (i.e., strategy that facilitates mechanism engagement) or a moderator in that it is necessary but not sufficient for mechanism engagement and/or strengthens the influence of another mechanism. A similar argument can be made about therapeutic alliance. Disentangling mechanisms from moderators and facilitators will expedite future research on proximal "levers" that cause mechanistic change—a critically important area of work that will aid in identifying the most promising strategies for treatment optimization. As part of this work, it will be necessary to consider that facilitators may trigger change in more than one mechanism, including those they are not theoretically paired with (e.g., exposures conducted with focus on fear extinction likely produce change in expectancy or vice versa). Thus, individual studies should include a comprehensive plan for measuring all plausible mechanisms—a necessary step for definitive mechanism testing (see Outcome

Measurement/Analysis in Clinical Trials section). This will also speed identification of the most potent optimization strategies for various mechanisms and may help identify alternate optimization strategies when tailoring is needed (e.g., for youths who have difficulty articulating fear cognitions; see Developmental Considerations section).

Mechanism or readout measure

Testing mechanisms is inherently dependent on imperfect "signals" or readout measures of those mechanisms, that is, observable responses used to measure the mechanism. For example, human fear-extinction research is at risk of failing to distinguish between readout measures and the underlying latent process of fear extinction (see Cronbach & Meehl, 1955). Typical readout measures in fear-extinction research are assessed on the subjective (e.g., threat expectancy), physiological (e.g., skin-conductance responses), and behavioral levels (e.g., safety behavior). Readout measures can thus be understood as the observable expression of a latent mechanism. Despite the assumption that they all measure the same underlying process, these readouts often do not converge (see Multimodal Measurement and Convergence Across Measurement Levels section), and subtle changes in measurement may differentially influence readouts (see Lonsdorf et al., 2017).

The distinction between mechanism and readout also applies to exposure theories. In the original description of EPT, within-session fear reduction is a readout that indicates successful emotional processing. Findings that within-session fear reduction is not consistently associated with treatment success (Craske et al., 2008) could thus be an issue with the readout measure, not the mechanism itself. Conversely, a readout measure might signal change in an underlying mechanism other than the one with which it is commonly associated. For example, within-session extinction could be consistent with IL if it serves as a signal of initial learning (followed later by consolidation and retrieval). The distinction between mechanism and readout has long been discussed in behaviorism, for example, by demonstrating that learning can occur without changes in observable behavior ("latent learning"; Tolman & Honzik, 1930). In addition to having implications for measurement error and power in clinical trials (see Measurement section), this issue is of notable clinical importance because therapists depend on observable readouts to guide exposure. Going forward, it will be important to distinguish between mechanisms and readout measures, which may be influenced by multiple factors apart from the mechanism itself. This calls for more basic research on the link between specific readout measures and a proposed underlying mechanism and consistent operational definitions and measurements for mechanisms of change.

Measurement

Mechanism measurement has largely followed traditional approaches used in experimental laboratory research. This has commonly included assessment of self-reported distress (e.g., Subjective Units of Distress Scale [SUDS]) or psychophysiological indices (e.g., skin conductance) during exposure. More recent studies have included a broader array of measurement approaches in response to advances in mechanistic theory, new efficacy data, and emphasis on the Research Domain Criteria for studies funded by the U.S. National Institute of Mental Health (Insel et al., 2010). Despite a long tradition of incorporating mechanistic measures in both analog and clinical studies of exposure, relatively little psychometric data exist to guide selection and/or administration timing. As a result, researchers have developed a variety of "homegrown" measures and supporting procedures (e.g., for rater training). This presents a challenge for replicability and likely contributes to inconsistent findings across trials. This problem may become even more apparent during forward translation, as resources decrease and exposure delivery shifts to fit various "real-world" contexts. It will be important to develop and follow a coordinated research agenda that emphasizes both psychometric rigor and pragmatism in the measurement and reporting of exposure mechanisms, an issue receiving ongoing attention in the National Institutes of Health Science of Behavior Change program, including recommendations for standards in the reporting of mechanistic behavior-change research (Birk et al., 2023).

Psychometric properties of measures

Future work will need to evaluate mechanistic measures in a rigorous way, with emphasis on establishing reliability, construct validity, and predictive validity (i.e., with treatment outcomes). This is particularly true for psychotherapy process-based constructs (i.e., events or interactions that occur during psychotherapy sessions) traditionally assessed at the exposure practice or session levels (e.g., expectancy, extinction, distress tolerance). It will be important to establish interrater reliability for measures rated by therapists or other observers and to detail the procedures needed for training them to a reliable criterion and maintaining it over time. These procedures might include developing a detailed manual to guide ratings, double-rated observations, and monitoring to prevent drift. Although

time-consuming, such procedures reduce measurement error, boost power, and support rigor and replicability across research groups, particularly when made publicly available. Innovations in advanced computing (e.g., artificial intelligence [AI] using natural language processing) might also be tested to replace human reporters, reducing or eliminating the need for ongoing monitoring (Ewbank et al., 2020). It will also be essential to establish construct validity through comparison with established measures of similar constructs (convergent validity) and those of different constructs (divergent validity). It is not yet clear whether measures commonly used to assess different theorized mechanisms will separate from one another or whether they represent overlapping aspects of the same underlying process (e.g., cognitive vs. emotional; see Conceptual Considerations section). This conceptual ambiguity has often resulted in studies using different construct labels despite identical measurement approaches (Cooper et al., 2017). Although a number of studies have measured more than one mechanistic construct, virtually none have deployed them in a way that facilitates examination of construct validity (i.e., with the same timing of administration during exposure). It will be important for the field to develop consistent definitions and measurement approaches for mechanism-relevant constructs that can be adopted across research groups and through forward translation. In contrast, numerous studies have examined predictive validity of mechanistic measures with clinical outcomes after treatment. This is particularly true for constructs based on SUDS (e.g., fear activation, extinction, variability), although studies have increasingly focused on others (e.g., expectancy, experiential avoidance, self-efficacy). However, few measures have consistently predicted outcomes across studies and research groups. At present, it is not clear whether inconsistent findings relate to differences in study procedures and/or samples, weak relations of these constructs with outcomes, or one or more common measurement challenges (as described below, e.g., limited reliability, different measurement approaches and timing).

Limited reliability and construct heterogeneity can drastically affect findings in mechanistic studies. Wilcox et al. (2013) presented a seemingly benign example in which 10% of analyzed data depart from the main sample-data pattern. This scenario could readily be observed in measurements of exposure and related symptom reduction. In this case, between-groups effect sizes are reduced by more than 70%, and statistical power is reduced by more than 60%, which increases the required sample size more than 10-fold (Faul et al., 2009). These problems are compounded in mechanistic research designs because there are multiple stages of analysis that sequentially worsen the problem. A key design in exposure research involves mediation with two parts, for example, one in which an intervention first predicts level of exposure engagement and then one in which exposure engagement predicts symptom response. In this case, commonly accepted levels for reliability (e.g., .80) would be expected to reduce standardized mediation effects by more than 50%. These examples fall under the umbrella of "data pollution" (De Nadai et al., 2022)—unintentional errors in data that have substantial impact on results. Data pollution can be addressed post hoc through analysis (e.g., robust estimation, latent variable estimation) and a priori through research design (e.g., selection of measures that exceed conventional norms for reliability). Once elements of data pollution are addressed, improved effect sizes and model precision substantially reduce the number of participants needed, which will contribute to accelerated scientific progress.

It will also be critical to establish "best practices" for handling common challenges that arise in process-based mechanism measurement. First, optimal assessment timing and variable calculation are unknown. Calculation of the same construct varies widely across studies, for example, using one time point (van Minnen & Hagenaars, 2002), average of time points (e.g., Jaycox et al., 1998), change scores that sum all time points (Benito et al., 2018), all time points in a multilevel model to compare effects across time (e.g., Sripada & Rauch, 2015), or calculation of reliable change indices (Cooper et al., 2017). Calculations based on more frequent assessment within and across sessions will facilitate better understanding of the time course of change and reduce measurement error (thus improving power; Faul et al., 2009). Some constructs are traditionally sampled more frequently than others (e.g., within-session variability vs. extinction); to ensure equally robust measurement/power across constructs, it may be important to use similar time points in the calculation of each. Finally, a majority of process-based measures focus on acute changes (i.e., within-session) or longer-term (i.e., between-session) changes during exposure tasks. From a learning perspective, these may be most relevant for understanding initial learning and later retrieval but miss the opportunity to understand processes related to consolidation (e.g., during postexposure processing). Understanding these processes provides an opportunity for memory editing (Phelps & Hofmann, 2019). Consolidation is thought to occur within a 6-hr period, leaving open the possibility that memory may be modified even well after the exposure session (Kredlow et al., 2018). Likewise, process-based measures have largely missed the opportunity to assess generalization of learning to new stimuli and/or contexts. Future

studies should enhance efforts to assess consolidation (e.g., by following up with participants during the 6 hr following exposure) and generalization (e.g., with regular assessment of potential experiences with new stimuli/contexts). For assessing these and other constructs outside of treatment sessions, studies might deploy low-burden ecological momentary assessment strategies to capture detailed temporal changes (e.g., Walz et al., 2014). Researchers might also deploy methods for controlled observation. For example, Richter et al. (2021) used a posttreatment behavioral-approach task to demonstrate modest generalization of exposure learning to new stimuli. Future studies should also work to identify the optimal window for measuring both learning generalization and change during the consolidation period, which may vary across development, individuals, and/or context.

A second challenge is that therapy content itself may create bias and measurement error. For example, psychoeducation could introduce expectancy or social-desirability bias in later self-report ratings of mechanistic constructs (e.g., patients may be more likely to report changes in expectancy vs. other constructs after psychoeducation focused on expectancy violation as a goal). Likewise, measures of mechanism depend on the quality of the exposure context, which is influenced by a wide variety of potential external factors (e.g., changes in stimulus, therapist actions) and patient behaviors (e.g., avoidance). Successful measures may need to account for the occurrence of these and their timing in relation to mechanism measurement (e.g., decrease in expectancy may not reflect mechanistic learning if preceded by avoidance). Selection of the exposure task may also influence mechanistic ratings, for example, if "easy" tasks produce a floor effect in ratings. A long-standing issue in the calculation of between-session changes relates to the introduction of increasingly difficult exposure tasks across treatment, making it difficult to determine whether mechanistic changes are occurring. This remains problematic even when variable task selection is an explicit goal (Craske et al., 2014) because many patients become more willing to try challenging exposures later in treatment. For additional discussion about delivery variation, see Measuring and Accounting for Exposure-Delivery Variation section. Measuring and accounting for these variations is likely to be complex, particularly as delivery becomes less controlled and patient presentation more variable through forward translation. Future studies will need to explore these possibilities in greater detail and consider innovative solutions for handling complexity in a practical manner (e.g., with AI approaches; Ewbank et al., 2020).

Multimodal measurement and convergence across measurement levels

One central measurement issue is whether multimodal assessment across measurement levels (e.g., self-report, behavioral, physiological, neural) and/or internal process (e.g., cognitive, emotional) is needed to ensure an accurate accounting of mechanistic change. Threat-conditioning paradigms provide a useful perspective on the challenges inherent in this issue. Although some studies have indicated correspondence between physiologic (e.g., Skin Conductance Level) and expectancy ratings (Fanselow & Pennington, 2018), many have not. Lubin and associates (2023) found that these measures shared less than 6% of variance, and Constantinou et al. (2021) reported shared variance ranging from a low of 0.5% during extinction-phase assessment to near 11% when early and late responding were parsed. These results caution against the assumption that these measures accurately assess the same construct, even in the context of tightly controlled laboratory studies. However, the distinction between mechanistic construct (e.g., extinction learning) and observable readout measures (e.g., Skin Conductance Response, expectancy ratings) also needs to be considered. Readout measures are influenced by a variety of factors and may thus indicate the same mechanism despite low overlap between measures (see Conceptual Considerations section).

Desynchrony among measures of fear is an old problem, with some indication that degree of desynchrony among measures—for example, reflecting self-report (verbal/cognitive), motor, and physiologic response systems (Lang, 1971)—may be moderated by degree of fear, external demands, type of intervention employed, time passed since improvement, and measure used (Hodgson & Rachman, 1974). Each spells trouble for the assessment of mechanism in clinical trials, raising questions about whether the study of mechanism should attend to what Hodges and Rachman (1974) proposed 5 decades ago, that (mechanistic) "effectiveness of a therapeutic technique should be assessed across all response systems" (p. 322). Indeed, examination of separate measures in the verbal/cognitive realm—self-report of fear (SUDS) versus expectancy—reveals that both have had predictive successes and failures (e.g., Pittig et al., 2016; Smits et al., 2013). This might be explained by individual differences in interoception accuracy (Khalsa et al., 2018) or reliance on expectancies or habit in information processing (Fradkin et al., 2020) that influence the meaning/accuracy of self-reported fears or expectancies, respectively. These factors may be even more relevant earlier in development

(see Developmental Considerations section), raising questions about whether/how multimodal assessment might be streamlined for youths while optimizing predictive value. Furthermore, the link between these measures and the core concept of fear continues to be debated in relation to whether verbal/cognitive, motor, and physiologic response systems arise from the same systems or represent separate pathways to fear and fear-related behaviors (LeDoux & Pine, 2016). Indeed, the distinction between higher- and lower-order processes can aid the understanding of discordance among measures and associated processes (e.g., forecasted action vs. conditioned action, self-report of fear or expectancies vs. physiologic reactivity; Taschereau-Dumouchel et al., 2022) and encourages investigators to pause and conceptually map linkages that might underlie a given measure of mechanism.

Some have argued that the essence of the emotional experience of fear and anxiety is the subjective experience and that the associated behaviors and physiological responses are, at best, indirect indicators of those inner experiences (LeDoux & Hofmann, 2018; Taschereau-Dumouchel et al., 2022). The subjective experience can be assessed only through verbal report, which constitutes a methodological barrier to studying conscious feelings in animals and humans who are nonverbal or earlier in development (Hofmann, 2008). Researchers often ignore desynchrony between these measures or arbitrarily choose one measure over others (often favoring objective over subjective measures). However, different outcome measures do not necessarily measure the same construct or may measure the construct at a different level of cognitive awareness (e.g., verbal report at the conscious level and physiological measures at the automatic and subconscious level). Likewise, some measures might provide a meaningful comparison between animals and humans (e.g., physiologic), but others may be more valuable for understanding the human experience (e.g., self-report). Thus, each measure may contribute unique information that can be seen as informative rather than a problem that needs rectifying (e.g., through error correction; Guolo, 2008). Moving forward, extinction research should focus greater attention on ontological issues, including explicit examination of relationships among readout measures and development of broader frameworks that incorporate measurement of extinction learning relative to other indices of declarative and associative learning and encompass related concepts (e.g., perseveration and behavioral persistence; Eisenberg et al., 2019).

Psychometric properties of laboratory models with humans

A fair amount of mechanistic research is based on laboratory models used as proxies for exposure therapy. Fear and avoidance conditioning and extinction are among the most prominent (Krypotos et al., 2018; Pittig et al., 2020). Briefly, fear conditioning involves repeated presentations of a neutral stimulus (designated "conditioned-threat stimulus" [CS+]) with an aversive unconditioned stimulus (US), resulting in the CS+ triggering anticipatory fear responses (e.g., elevated threat expectancy or skin conductance; Lonsdorf et al., 2017). Fear extinction, as a proxy for exposure tasks, subsequently involves repeated unreinforced presentations of the CS+, leading to a reduction in fear responses to that CS+. In avoidance conditioning, the aversive US associated with the CS+ can be prevented by a predefined avoidance response to the CS+. During avoidance extinction, the CS+ is no longer followed by a US, and changes in avoidance responses are examined (Pittig et al., 2020).

Fear and avoidance extinction has been studied across species, enabling insights into theoretical (Bouton, 2002) and neurobiological (Milad & Quirk, 2012) underpinnings (for additional detail about animal models, see Advances From Models With Nonhuman Animals section). Despite being highly influential, only a handful of studies have assessed the test-retest reliability of fear-response measures during fear extinction, such as self-report, psychophysiological, or neural indices (e.g., Klingelhöfer-Jens et al., 2022). Results indicate limited reliability across measures, although floor effects may contribute given that fear responses decreased during extinction and reliability was better during fear acquisition. Even less is known about psychometric properties of avoidance conditioning and other laboratory models (e.g., emotion-regulation tasks to model experiential avoidance, laboratory stress challenges to model distress tolerance, false feedback to manipulate self-efficacy beliefs). The psychometric properties of responses within these models have, to the best of our knowledge, not been investigated.

Furthermore, less is known about the external validity of laboratory models of exposure therapy. To what extent do methods and findings of laboratory models match clinical phenomena and translate to practice? Conceptually, validity criteria for laboratory models include face validity, construct validity, content validity, and predictive validity (Haynes et al., 1995; Krypotos et al., 2018; Scheveneels et al., 2016). However, few

studies have tested these in a robust manner (Pittig et al., 2020; Scheveneels et al., 2016). For example, face validity of stimuli used is relatively weak. More complex, multisensory, and life-like stimuli will better simulate the naturalistic situations confronted during exposure (Scheveneels et al., 2016). Moreover, recently acquired fear or avoidance responses or brief inductions of emotion regulation or self-efficacy may be substantially different from patient responses that have been elaborated, generalized, and socially reinforced over years. Recent studies have made efforts to improve relevance of laboratory models for patient populations and exposure therapy, with promising results (e.g., inclusion of overnight consolidation after fear acquisition, use of explicit instructions to ensure fear responses are reliably acquired; Hollandt et al., 2020). With respect to predictive validity, only a handful of studies have investigated whether/how individual differences in laboratory responses predict clinical response to exposure therapy (Scheveneels et al., 2021). This is critical because pretreatment laboratory assessments could have prognostic value or guide treatment tailoring. In summary, despite the importance of laboratory research on mechanisms of exposure, the psychometric properties and validity of models are still largely unknown.

Advances from models with nonhuman animals

Early translational studies using nonhuman animals have long been important for understanding exposure mechanisms and will continue to be a key avenue for future research. Animal literature has informed a variety of paradigms/behaviors relevant for laboratory models with humans and exposure therapy. These include fear potentiated startle, in which fear is initially conditioned by pairing a CS with mild shock (the US) and then a sudden loud noise is presented during the CS. Conditioned fear results in an exaggerated startle response (Davis, 2006). Passive versus active avoidance can also be assessed using tasks in which an individual is trained to inhibit the natural inclination to avoid negative consequences. A simple distinction between passive and active avoidance is that in the former, individuals choose to not engage in a response that will increase discomfort, and in the latter, they explicitly choose to use behaviors that decrease discomfort (Krypotos et al., 2018). Conditioned suppression of reward seeking can also be assessed using paradigms in which a rat is first trained to carry out a task for reward (e.g., press a bar) and then fear conditioned (perhaps also with extinction). Willingness to engage in bar pressing is assessed during CS presentation. Conditioned responding provides insight about willingness to engage in pleasurable activities following an anxiety-inducing experience. For example, a rat may show a significant reduction in fear response (freezing) following extinction but fail to resume reward seeking (Shumake & Monfils, 2015). Likewise, a patient may report distress reduction after treatment but refrain from activities they used to enjoy—a distinction that likely has important ramifications for long-term functional outcomes.

In fear-extinction paradigms (a proxy for exposure therapy, see Psychometric Properties of Laboratory Models With Humans), individuals are susceptible to the return of fear. This is evident through the phenomena of spontaneous recovery, reinstatement, and renewal, in which spontaneous recovery occurs after the passage of time, reinstatement is brought on by a stressful event, and renewal occurs in new contexts. Animal research has provided important insights into these phenomena and potential mechanisms through which novel treatment approaches could improve them. Nader proposed that once retrieved, memories become susceptible to disruption, and blocking the molecular cascade engaged at retrieval can attenuate fear expression and prevent return of fear—known as reconsolidation blockade. In an initial proof-of-concept study (Nader et al., 2000) and many that followed, reconsolidation blockade was achieved with protein synthesis inhibitors that cannot be used in humans. Since then, studies have explored pharmacological approaches to target fear memories with varying degrees of success. In a first study with humans, administration of propranolol after retrieval of a traumatic memory reduced physiological responding compared with placebo (Brunet et al., 2008). Additional studies of propranolol after memory reactivation demonstrated some success (e.g., Elsey & Kindt, 2017; Kindt et al., 2009). Since then, additional drugs have been tested for reconsolidation blockade (e.g., Sirolimus, Surís et al., 2013; mifepristone, Pitman et al., 2011). Despite promising results, this approach is not universally effective (e.g., Wood et al., 2015). Future studies should continue to examine whether targeting mechanisms that underlie reconsolidation blockade—first identified in animals—will effectively translate into improved outcomes in humans.

Focus on reconsolidation offered an opportunistic window for targeting fear memories. Although this area has received substantial focus to date, our discussion of it here serves to emphasize a recognizable example of cross-translation between basic neuroscience and the clinic. Monfils et al. (2009) proposed retrieval-extinction, in which extinction training is used during reconsolidation—a finite period when memories are susceptible to disruption. In this paradigm, an isolated retrieval trial is presented; then, after sufficient time has

elapsed for the memory to destabilize, extinction training is introduced. This has been successful in some studies with rodents and humans (e.g., Agren et al., 2012; Rao-Ruiz et al., 2011) and translated with some success in clinical and analog samples (e.g., Lancaster et al., 2020). One advantage of reconsolidation-based approaches is that they are theoretically less susceptible to return of fear (Monfils & Holmes, 2018)—and unlike extinction, need not engage the prefrontal cortex (Agren et al., 2012). They could thus prove effective in treating populations in which engaging the prefrontal cortex proves challenging (e.g., individuals with traumatic brain injury or youths). The above approaches are in their infancy, and future research should continue to test whether vulnerability opened by memory retrieval will prove an effective therapeutic window on a larger scale in humans.

Ultimately, animal-to-human translation should be viewed as an iterative process or "feedback loop" that also serves to improve basic models and enable continued treatment refinement. This will facilitate ongoing consideration of species-specific methods and measures, feeding back practical information about what can reasonably be translated across domains and what future adjustments should be made (e.g., Haaker et al., 2019). Rodent study successes over the years suggest that most inroads have evolved through feedback from behavioral developments and approaches. Human behavioral studies can serve to inspire rodent researchers to further develop novel assays and improve understanding of the broad repertoire of rodent fear expression and processing. Although more homogeneous than humans, rodents still display vast individual differences in fear responding. Rodent research is uniquely suited for explaining individual variance given the ability to isolate individual factors of interest under tightly controlled conditions. Ultimately, close collaboration between basic and clinical researchers will facilitate development of methods with a balance of internal and external validity and will speed treatment and study development.

Study Design and Analysis

Experiment and clinical-trial design

In a seminal review of mechanisms of change in psychotherapy research, Kazdin (2007) highlighted the advantages of using an experimental approach for testing therapeutic change mechanisms. This emphasis is complemented by calls to redesign clinical trials with specific tests of how treatment elements are linked to mechanism(s) as a prelude to evaluating clinical significance (Nielsen et al., 2018). Thus, progress in understanding mechanisms relies on well-designed experiments that specify putative mechanisms and clinical trials that can show the effects of targeting specific mechanisms. Across both preclinical experiments and clinical trials, there is a need for traditional and innovative randomized group-based designs and single-subject designs (e.g., multiple baseline, alternating treatment, knock-out). Small-*N* designs capture within-subjects change that is often missed in group-based designs and are ideally suited for isolating the contribution of specific exposure strategies. They are also valuable as part of a study series that follows the experimental-therapeutics framework by testing initial mechanism engagement before moving on to a larger pilot trial. For example, one recent study used a multiple-baseline design to test a therapist-training tool for optimizing exposure delivery in community treatment settings (Benito, Herren, et al., 2021). Results demonstrated that therapists met delivery benchmarks and patients showed highest levels of within-session extinction only after the training tool was introduced—demonstrating that successful manipulation of exposure delivery via the training tool likely had a causal effect on a theorized mechanism.

Experimental designs. Experimental designs in laboratory settings fill an important translational gap between basic laboratory models (e.g., for fear and avoidance conditioning/extinction) and full clinical trials and will continue to be part of a meaningful research agenda. These studies aim to manipulate putative mechanisms using a brief exposure protocol and examine how changes in these relate to subsequent reduction in fear and/or anxiety (or other dimensional symptom outcomes). Samples are often composed of healthy participants and/or individuals with elevated symptoms (i.e., a clinical diagnosis is not required), which facilitates speedy recruitment. The ability to isolate mechanisms of interest and rapidly determine which "levers" engage those mechanisms best is a key advantage of these designs. Because they require fewer resources than full treatment trials, these studies are also highly efficient. For example, in one study focused on testing threat reappraisal as a core mechanism of change, Kamphuis and Telch (2000) examined the independent and joint effects of distraction and threat focus with disconfirmation on fear reduction during exposure. Participants with severe claustrophobia symptoms were randomly assigned to exposure (a) with guided threat reappraisal, (b) with a cognitive-load distractor task, (c) with both, or (d) without either. The cognitive-load task had a marked detrimental effect on threat reappraisal and fear reduction, whereas guided threat reappraisal (regardless of cognitive load) had a facilitative effect. The greatest level of

threat reappraisal and fear reduction and the lowest level of return of fear were observed in the guided threat reappraisal without cognitive-load condition—providing a promising avenue for future clinical trials.

Clinical-trial designs. Full clinical-treatment trials will continue to be important for testing the efficacy of exposure optimization strategies (and corresponding mechanisms) before making recommendations for clinical practice and are essential for ensuring that promising strategies hold up under increasingly generalizable conditions (e.g., with treatment-seeking clinical populations, using a multisession course of treatment). For example, optimization strategies could function differently when moving from experimental protocols with a limited number of sessions into a full course of exposure therapy (e.g., with repeated consolidation and retrieval of learning over a period of months). In addition, individuals with anxiety disorders show impaired extinction learning; this is even more pronounced among individuals with more than one anxiety diagnosis (Marin et al., 2017). Optimal strategies for engaging exposure mechanisms may differ when moving from a population with subclinical/mild symptoms into a treatment-seeking clinical population. For example, it is possible that individuals with more severe/impairing symptoms need additional compensatory strategies, are less able to experience long-term benefit from exposure in the presence of factors that interfere with learning (e.g., safety behaviors), and/or need a higher "dose" of various strategies.

As possible, secondary studies that examine exposure delivery in past clinical trials will be important for advancing knowledge about which mechanism-related strategies were actually used (intentionally or unintentionally). One common assumption is that older trials were conducted using a "habituation" model guided by EPT, yet few treatment manuals explicitly called for this, and the limited available evidence suggests considerable variation in delivery, with exposures more commonly including features consistent with other models (e.g., IL; Benito et al., 2018). Determining whether clinical outcomes from prior trials can be improved with mechanism-informed exposure techniques is perhaps the highest priority goal for future research. To that end, information about exposure delivery in older trials will be needed to design realistic "exposure as usual" conditions in future trials, thus enhancing significance. In addition, given the conceptual overlap of mechanistic theories (see Conceptual Considerations section), it will be challenging—if not impossible—to design trials that directly compare various models (e.g., extinction vs. inhibitory learning vs. cognitive) without inadvertently creating artificial "straw man" comparison conditions. Instead, it may be more valuable to follow an integrated theoretical framework with focus on testing the most promising mechanisms and corresponding strategies (e.g., for before, during, and after exposure) identified in earlier translational work.

Innovative clinical trial designs, such as sequential multiple assignment randomized trial (SMART) designs, are likely to be helpful for testing multiple exposure-optimization strategies and/or related mechanisms. SMARTs test adaptive interventions via multiple randomization points based on time and/or tailoring factors (Almirall et al., 2014). Information about mechanisms might be used as a tailoring factor that triggers a subsequent treatment decision (i.e., rerandomization; Sauer-Zavala et al., 2024; Southward & Sauer-Zavala, 2020). To the degree that change in mechanism precedes clinical improvement, mechanisms can serve as a proximal indicator of treatment response and guide relevant decision-making earlier in treatment. For example, patients could be initially randomly assigned to a condition (e.g., different sets of exposure-optimization strategies vs. exposure as usual), followed by a second random assignment after a specified number of weeks. At that time, individuals who show evidence of mechanism engagement might continue within their assigned treatment arm, while individuals that do not are randomly assigned again (e.g., to continue vs. switch arms). This example design would accommodate mechanistic heterogeneity across individuals (e.g., with respect to relevant mechanisms and/or optimization strategies) and explicitly test alternate treatment options for individuals who do not show initial change in a given mechanism. Alternatively, individuals that do show evidence of mechanism engagement early in treatment could be randomly assigned to continue or terminate treatment (Sauer-Zavala et al., 2024). This type of design would test whether individuals showing evidence of mechanism engagement could end treatment early, reducing excess treatment and allowing providers to treat additional patients. These are just a few examples of the ways in which SMART or other adaptive designs might use information about mechanisms to make complex treatment and/or dosing decisions and optimize outcomes.

Measuring and accounting for exposure-delivery variation

Exposure delivery can vary widely even in tightly controlled clinical trials with highly adherent therapists (Benito et al., 2018; Benito, Machan, et al., 2021)—and this can have profound implications for conclusions that might be drawn about mechanism. Most trials use some form of adherence or fidelity monitoring, yet these are often trial-specific, lack psychometric support,

and are designed to capture only basic information about "ingredients" used in each session (e.g., whether exposure was used, whether proscribed elements from another treatment arm were used). Moving forward, studies should routinely and rigorously assess nuanced aspects of delivery with direct relevance to theorized mechanisms (e.g., exposure dose, targeted core fear, use of compound exposures, process of task selection, key elements of postexposure processing, and many more). Researchers should also seek to develop psychometric support for measures of exposure delivery, including interrater reliability and validity. As part of this process, it will be essential to develop clear operational definitions (i.e., concrete descriptions of what a delivery strategy "looks like") to facilitate replication across studies and translation into service settings (e.g., through clinician training). Much like problems with labeling and defining mechanisms (i.e., different mechanistic labels often used for the same measurement approach and vice versa; see Conceptual Considerations and Psychometric Properties of Measures sections), specific delivery techniques are likely labeled and defined differently by different research groups. Some clinical trials have failed to find group differences when comparing different exposure approaches (e.g., Cooper et al., 2017), leading some to question whether the robust effect of exposure simply overshadows relatively smaller effects associated with variation in delivery. However, other well-powered studies have found modest to meaningful differences (e.g., Cooper et al., 2017; Gloster et al., 2011; Pittig et al., 2021) and/or association between delivery variation and clinical outcomes (e.g., Benito, Herren, et al., 2021). Incorporating detailed measurement of a full range of exposure-delivery variables will be important for identifying/ruling out reasons that some trials find differences while others do not (e.g., whether other, uncontrolled delivery factors may have interfered).

Group-based exposure. Group-based exposure has many desirable elements that are unavailable in individual therapy, including peer modeling/encouragement and efficiency (i.e., treating multiple patients at once). However, it is difficult to separate specific causes of exposure effects because of added variables introduced by group dynamics (e.g., personality characteristics of group members at the subject level, cohesion/morale at the group level). In addition to traditional multilevel-modeling approaches, social-relation modeling (Kenny et al., 2006) can explicitly separate effects between individual participants and the broader group, and one-with-many modeling (Brinberg et al., 2022) can address the therapist role (the therapist as the "one" and individual group members as the "many"). The one-with-many design is easier to implement in practice, in which patients simply rate their therapist. Social-relations modeling requires group members to rate all other group members, which presents additional logistical challenges. The one-with-many design has been used in a number of clinical settings, including evaluation of clinician effects in primary care and adolescent substance-abuse treatment (Anagnostopoulos et al., 2012; Marcus et al., 2011). These group-based designs allow researchers to separate specific effects of exposure and provide an opportunity to identify specific elements that enhance exposure effects at both individual and group levels. For example, it is possible to quantify how effects of a particularly resilient group member shape more positive outcomes for the entire group or how a suboptimal experience of one group member can affect outcomes of other individual group members and overall group results.

Variation in exposure dose. There is strong theoretical rationale for the importance of a high dose of in-session and out-of-session exposure, and a higher number of in-session exposures robustly predicts improved treatment outcomes (Peris et al., 2017). Nevertheless, treatment protocols vary in the amount of exposure called for (e.g., number of sessions with exposure), and even highly adherent therapists deliver different amounts (e.g., number/duration of exposures) while remaining within protocol parameters (Benito et al., 2018; Benito, Herren, et al., 2021). It will be important for treatment studies to measure and report these exposure-dose variables, which may explain variation in outcomes across patients and facilitate comparison of findings across trials. Given that homework completion can be highly variable (Kazantzis et al., 2017), a major problem is that patients can never be fully randomly assigned regarding out-of-session exposure dose. As a result, traditional intent-to-treat approaches lose the ability to draw causal inferences and can underestimate the magnitude of intervention effects (Connell, 2009). A creative analytic approach designed to address this problem is complier average causal effects (CACE) modeling (Yau & Little, 2001). The CACE approach relies on latent variable modeling and has been applied in multiple domains, including substance abuse (e.g., Cordovilla-Guardia et al., 2017). Data generated during exposure-therapy research fits well into CACE modeling strategies but requires careful measurement of exposure dose and outside exposure practice. The CACE modeling approach can also be used to address participants who prematurely discontinue treatment, which is common (Swift et al., 2017).

Rigor and reproducibility in clinical trials

Although a full discussion is beyond the scope of this article, several issues that relate specifically to exposure-therapy mechanisms are worth highlighting here. As

described in the Psychometric Properties of Measures and Measuring and Accounting for Exposure Delivery Variation sections, researchers can work to enhance both rigor and reproducibility by using well-validated measures to assess mechanisms and clearly articulating detailed measurement methods for mechanisms and aspects of exposure delivery (e.g., through rating manuals, standard operating procedures for training/monitoring). Ideally, these could be publicly shared—for example, as part of a larger exposure research "toolbox" available for researchers. Transparency and consistency of mechanism measurement and treatment delivery will accelerate work in the field by reducing discrepant findings across studies (and contributing to knowledge gained when findings do differ). Another important area for future work relates to the need for objective mechanistic measures (e.g., psychophysiological, observed behavior) and clinical outcomes (e.g., rated by masked independent evaluator), which will greatly enhance rigor. Relying primarily on subjective measures of mechanism (e.g., self-report) introduces bias; in trials that also lack double-masking, the risk for bias related to subjective measures increases further (e.g., to 22%, as reported in Savović et al., 2012). As described above (Psychometric Properties of Measures section), this degree of error has a profound impact on power and dramatically alters findings (Wilcox et al., 2013). This issue is likely compounded in trials that use subjective ratings of both mechanism and clinical outcome, particularly if both rely on the same rater, in which shared forms of bias can inflate the risk of Type I error. At present, objective measures of mechanism are infeasible in many trials; it will be particularly important that these trials use masking and objective clinical outcomes. Future studies should work to improve feasibility of objective measures (e.g., psychophysiologic or behavioral indices assessed through wearables) for use in clinical trials; for example, functional clinical outcomes might be assessed based on geographic range of location in daily life during and after exposure for agoraphobia. Finally, many exposure trials to date have suffered from modest sample sizes that, coupled with measurement error, lead to underpowered studies. Leveraging multisite research consortia can dramatically increase total sample size, thereby producing well-powered studies of putative exposure mechanisms that are more likely to be replicated (Smits et al., in press).

Outcome measurement/analysis in clinical trials

Perhaps guided by seminal articles on research on the mechanisms of change of psychosocial treatments (e.g., Kazdin, 2007), many studies that have aimed to test mechanisms have done so with secondary analyses of clinical-trial data. In these studies, repeated measures of symptom severity and candidate mechanisms have been modeled to test mediation (degree to which change in mechanistic variables accounts for clinical change). However, the mediator is not randomized, which introduces confounding that limits causal interpretations of associations between mediator and outcome. Hence, mediation analyses that are correlational (e.g., parallel-process models that relate the slope of the mediator to the slope of the outcome or use of time-varying mediators to relate to time-varying outcomes [even lagged outcomes] without controlling for previous outcomes) will often overestimate the causal effect of the mediator on outcome. Researchers have developed a number of new approaches using variations of cross-lag panel models that more accurately assess the causal effect of a mediator on outcome, such as random-intercept cross-lag panel models, (Hamaker et al., 2015), latent change score models (McArdle & Hamagami, 2001), integrated cross-lag and fixed-effects models (Falkenström et al., 2022), and difference in differences. To enhance accuracy of causal estimates in these cross-lag models, randomized controlled trials (RCTs) can include the following methods.

First, assessment of mediators and outcomes should be frequent. In any cross-lag model, the lag must accurately reflect the time that it takes for an effect to be observed in the dependent variable. Many studies assess mediators and outcomes infrequently (e.g., every 4 weeks). However, if the effect of the mediator is relatively rapid, any effect on the outcome may be washed out by the time the lagged outcome is measured. Likewise, if the effect is longer than the time between assessments, then no effect will be found. It is best to assess measures too often versus not often enough because time-series analysis can initially determine the appropriate lag time for cross-lag models. Furthermore, these methods assume that data are missing at random (MAR). Frequent assessments also enhance the likelihood that data are MAR because if dropout is related to scores on the outcome, recent assessments of the outcome should be related to the missingness, thereby meeting MAR.

Second, clinical trials should measure many possible mediators and include them simultaneously in the model; the vast majority of trials measure just one or two potential mediators. Any relation between these mediators and outcome may not reflect true causal relations but, rather, a correlation of measured mediators with true mediators (i.e., third-variable confounds). This problem is present even with advanced cross-lag panel analyses that examine within-persons relations between mediators and outcomes. Thus, if there are multiple

plausible mechanisms through which the intervention affects the outcome, whether those are specific (e.g., expectancy violation) or nonspecific (e.g., therapeutic alliance), all should be measured and included in analysis to ensure that relations between mediators and outcomes are not overestimated because of omitted third variables. Sensitivity analyses can be conducted for models with and without covariates to examine model robustness.

Machine-learning applications

An agenda that proposes increasingly complex theories about exposure mechanisms has a better chance of succeeding if it embraces methodological advances in statistical approaches designed to rigorously test such complex theories. Although the enhanced cross-lag panel models described above can be generalized to multiple predictors, these models are very complex and often require multiple constraints to converge to a solution even when considering only one predictor of outcome. Thus, they may be impractical for teasing apart the exact mechanisms by which exposure affects outcome. Machine-learning (ML) methods provide a tool kit of flexible algorithms that may overcome some of the aforementioned limitations. Although ML has become increasingly mainstream in psychological science, the focus of application has been on prediction and classification (as opposed to causal inference). Yet many of the tools of ML can be applied to test novel mechanistic hypotheses in a robust manner designed to protect against pitfalls that often affect replicability and generalizability. In this section, we provide an overview of how ML can be applied to examine both mediation and moderation.

ML in mediation analysis. Mediation is often tested in (linear) structural equation models, which can be adapted to incorporate multiple mediators, different types of confounding, and repeated measures (Moerkerke et al., 2015). Increasing the complexity of mediation models may facilitate the identification of the causal effects of interest; however, this also increases the potential for statistical misspecification (e.g., assuming linear associations when they are actually nonlinear; Valente et al., 2017). To address this, many tools of ML that have been honed in the predictive-modeling framework are increasingly applied in the causal-inference literature (Blakely et al., 2021). ML methods (vs. traditional regression) are more flexible and make fewer assumptions. This means that if the true relations between treatments, mediators, and outcomes are nonlinear, estimates from a data-adaptive ML algorithm are more likely to reflect reality, especially as sample sizes increase.

Modern approaches to specifying causal models and evaluating them statistically are heavily influenced by the counterfactual framework (Pearl, 2019b). A key feature of this framework is that cause-and-effect questions are mathematically translated to methods for the estimation of potential outcomes at the individual level (Broadbent, 2015). Models are developed that use relevant pretreatment variables (e.g., participant and contextual characteristics) to estimate outcomes after experimental treatment and after control treatment. These outcome models can be applied to each individual to get an estimate of outcome after the individual's actual treatment and after the individual's counterfactual treatment. This is different from standard analysis, in which treatment effect is estimated through a comparison of group averages. Standard regression approaches can be used to develop outcome models but make restrictive parametric assumptions that may be unrealistic. As an alternative, outcome models can be developed using data-adaptive ML algorithms that model complex nonparametric associations, including high-dimensional interactions (van der Laan & Rose, 2011).

A variety of ML methods have been developed that can incorporate estimates from outcome models along with estimates from treatment-propensity models to enhance estimation of the treatment effect of interest (e.g., targeted maximum likelihood estimation and augmented inverse probability weighting; van der Laan & Rose, 2011). Although adoption of these methods has been slow in the behavioral sciences, the combination of open-source ML software, increased access to high-performance computing, and larger sample sizes will facilitate increased application of modern causal-inference approaches in future research on mechanisms of exposure therapy. Researchers interested in applying ML methods to test causal hypotheses could use a superlearner (stacked ensembling) approach that includes a diverse set of ML algorithms that allows researchers to make fewer assumptions about which algorithm may be best suited for their data by incorporating estimates from multiple, diverse algorithms. The potential for overfitting is reduced by applying cross-estimation (or cross-fitting) procedures (Zivich & Breskin, 2021) that provide outcome estimates for each individual that are based on models that were developed in data sets that excluded that individual (analogous to cross-validation in the predictive-modeling framework).

ML in moderation analysis. Exposure outcomes vary by patient, yet the field lacks clear information about predictive patient profiles and/or mechanisms through which variation occurs. Work in this area has highlighted

relevant factors such as motivation and insight (Abramowitz et al., 2019). To complement this variable-focused work (which has been based on traditional statistics or supervised ML), unsupervised ML takes a person-centered analytic approach. It can provide a more comprehensive data-driven profile of patients, especially regarding how exposure mechanisms differentially affect individuals. Unsupervised ML encompasses multiple approaches that find homogeneous participant-subgroup clusters within a set of heterogeneous variable responses. In the case of exposure, unsupervised ML can identify groups of patients who are more or less likely to respond to specific strategies and the specific conditions in which a number of candidate-exposure mechanisms have effects. For example, De Nadai et al. (2023) investigated how functional MRI (fMRI)-based neural mechanisms differentiated response to exposure-based treatment for obsessive compulsive disorder (OCD). By applying unsupervised ML to results across many OCD-relevant brain areas, they found three groups of patients. One group showed a neuroimaging pattern consistent with healthy control subjects and exhibited greater reduction in OCD symptoms with exposure therapy versus general stress-management training.

There are many options for implementing unsupervised ML, and some are better suited for exposure research. Compared with areas in which ML methods are often developed (e.g., computer-vision research), behavioral research has substantially more measurement error, and ML methods developed in other fields often do not accommodate the resulting bias. Methods based on mixture-modeling fit explicitly account for measurement error and are more likely to be robust in the face of such error. Examples include Gaussian mixture modeling (for continuous variable responses; also called "latent profile analysis") and latent class analysis (when applied to categorical responses). Other unsupervised ML approaches (e.g., *k*-means, hierarchical clustering) rely on post hoc approaches to model calibration (e.g., cross-validation), which can reduce robustness. Post hoc calibration is problematic for clinical research, which often cannot readily provide the steady stream of data needed to continually refine and test the model. Mixture modeling also provides a more precise framework for identifying causes and consequences of identified patient clusters through partial assignment of patients to each cluster in a way that acknowledges that imperfect cluster assignment and accounts for this in subsequent covariate analyses (Asparouhov & Muthén, 2014).

Promise and pitfalls of ML. The above approaches must be considered in the context of potential trade-offs. Many ML-based models are "black box" models in which the specific contributing role of each individual variable remains unknown and thus provide little insight into theoretical underpinnings related to who will benefit from what treatment. This property of ML-based models can make it difficult to correct models that behave incorrectly, especially related to bias against various demographic groups (Pearl, 2019a). Representative samples and humility regarding individual predictors are essential. Out-of-sample cross-validation, although rare in the behavioral sciences, can mitigate some of these problems. Data analysis also does not occur in isolation, and many innovative methods require careful pretrial design to function properly.

Application of these methods also requires appropriate lags and assessment of relevant confounds (i.e., other possible mediators) as recommended for RCT design above (Outcome Measurement/Analysis in Clinical Trials section). Furthermore, ML-enhanced propensity-score approaches have been developed for dichotomous predictors/mediators, although few are dichotomous, and transformation into such leads to loss of accuracy and power. Fortunately, researchers have recently proposed extensions of the propensity-score approach to include continuous predictors. These methods, once fully developed, will provide an alternative analytical framework for causal-mediation analysis.

Individual and Contextual Differences

The most relevant question for future research may not be "Which mechanism matters most?" but "Which mechanism matters most for whom, and under what circumstances?" It is highly likely that successful exposure depends on multiple interrelated mechanisms and that these mechanisms (and optimal methods/measures) will vary by individual, context, and/or timing (Rothman & Sheeran, 2021). As just one example, earlier translational work points to several contextual features that may underlie return of fear/avoidance following within-sessions extinction (Bouton, 2002). Contexts moderating extinction retention in animal models include internal affective state (e.g., degree of arousal), external contexts (e.g., smells, cage design), and passage of time. Human studies have replicated these effects with location (room), imagined location, and degree of arousal (e.g., Mystkowski et al., 2006). Accordingly, assessment of overall treatment outcome and session-by-session retention of extinction effects may need to consider context as a moderating variable in exposure efficacy. Likewise, failure to find meaningful differences in trials that compare exposure-delivery approaches could be explained by moderation of treatment effects by individual or contextual differences. For example, trials may enroll individuals with varying deficits in fear-extinction learning (e.g., some with deficits in

initial learning while others experience problems with consolidation); the delivery variations being tested may not target the relevant deficit for all patients. Little work has been done to elucidate important contextual, individual, or timing variations in mechanisms over a full course of clinical treatment; below, we explore potentially important factors to consider for future studies in this area.

Magnitude and type of emotion elicited during exposure

Factors that influence the magnitude and/or quality of distress during exposure may have implications for understanding mechanisms. For example, anxiety sensitivity (AS) has been conceptualized as an important amplifier of emotional/somatic distress (Otto et al., 2016), and this amplification may have important effects on safety learning during exposure. Patients high in AS may simultaneously face two types of fear cues, for example, external cues arranged by the therapist and internal cues amplified by AS. Threat expectancies could be tied to both, and although one type may be disconfirmed (e.g., actual harm in the presence of the external cue), another is inadvertently confirmed (e.g., feeling overwhelmed by "dangerous" levels of anxiety). When this occurs, there is potential for the exposure to appear successful relative to external dangers but constitute a sensitization experience relative to internal dangers.

Furthermore, exposure classically targets emotions of fear and anxiety, but other types of distress can be prominent in individual presentations and/or symptom clusters and are often targeted with exposure (e.g., disgust, incompleteness, shame, guilt). Disgust has been associated with contamination-related OCD, blood-injection-injury phobia, spider phobia (Cisler et al., 2009), and some forms of posttraumatic stress disorder (PTSD; Badour & Feldner, 2018). Different emotional targets may show different response patterns or have unique implications for treatment strategies or outcomes. Disgust has distinct neural and physiological substrates and is less closely tied to cognitions about negative outcomes, instead characterized by visceral discomfort (Cisler et al., 2009). In fact, disgust has been linked to evaluative conditioning, in which a CS acquires the valence (e.g., dislike) of a paired US. Disgust also has shown greater resistance to extinction than fear in some studies (e.g., Olatunji et al., 2007).

Incompleteness is a sensory-affective experience that things are "not just right," involved in some presentations of OCD and OC-spectrum disorders (Summerfeldt, 2004). Like disgust, incompleteness is less tied to specific negative cognitions, and relevant exposures are self-confirming: Uncomfortable feelings occur, and individuals learn to tolerate them. Also like disgust, incompleteness may show distinct response patterns, such as a shallower decline or even a flat trajectory during exposure (Milgram et al., 2022). In one study of pediatric OCD, greater baseline disgust and incompleteness, but not fear, predicted poorer response to treatment that included exposure (Cervin & Perrin, 2021).

Shame and guilt have been associated with PTSD. Shame is a negative affective-cognitive state that entails judgment of the self, whereas guilt entails judgment of one's actions (Saraiya & Lopez-Castro, 2016). In contrast to the present- or future-oriented focus of fear and anxiety, shame and guilt are past-focused emotions (Pugh et al., 2015). There is a small, mixed literature on the role of shame and guilt in the treatment of PTSD. Some studies suggest that traditional or prolonged exposure (PE) is less effective at targeting shame and guilt. These findings have facilitated development of tailored treatments, such as PE with imagery rescripting or cognitive-processing therapy, to more directly target related cognitions (e.g., Saraiya & Lopez-Castro, 2016). In sum, extant studies suggest that several unique processes are involved in "nonfear" exposure targets, with potential implications for exposure strategies and outcomes. These emotions may decline more slowly during exposure or benefit from augmentation strategies. Future laboratory and clinical studies will benefit from concurrent measurement of multiple affective states, which will enable direct comparisons.

Developmental considerations

Although exposure-based cognitive-behavioral therapy (CBT) is a first-line treatment with strong efficacy for youths with anxiety disorders, OCD, or PTSD (Freeman et al., 2018; Higa-McMillan et al., 2016), not all youths experience symptom remission, and many fail to maintain gains over time (Kodal et al., 2018). Adolescents in particular appear to experience less benefit from exposure than both younger children and adults (Turner et al., 2018) and may also be less likely to maintain symptom remission through long-term follow-up—particularly when receiving exposure alone (i.e., without medication; 20%–50% remission rate at 6 years; Ginsburg et al., 2018). These clinical outcomes are paralleled in the laboratory by marked differences in extinction learning. Adolescent rodents and humans show developmentally normative deficits in the retention of extinction learning, as indicated by higher threat responding at later extinction recall and in some cases, poorer initial extinction learning (e.g., Pattwell et al., 2012). One benefit of animal models is the ease of testing species-homologous modifications in exposure therapy

(via fear-extinction protocols) and other interventions while taking into account discrete age boundaries. Such evidence can systematically track the influence of stages of brain development on responses to treatment. In contrast, despite apparent mechanistic and clinical differences across development, the vast majority of clinical trials include youths of all ages and have not been powered to examine differences based on age (Seligman & Ollendick, 2011). It will be critically important for future clinical studies to be sufficiently powered to test the most relevant mechanisms and examine treatment-outcome differences across different phases of development.

Continued maturation of the prefrontal cortex and associated regulatory behaviors across adolescence has emerged as a potential explanation for why treatments that rely on prefrontal functioning may be less effective in pediatric populations. Extinction is largely dependent on the prefrontal cortex, development of which continues into the early 20s (Giedd et al., 1999). In contrast, the amygdala matures on a much faster timescale, reaching structural maturity during adolescence (Hu et al., 2013) while simultaneously exhibiting greater functional activation relative to adults in response to fear-associated stimuli (Lau et al., 2011). As structural and functional connectivity between prefrontal control regions and the amygdala strengthens into early adulthood, approaching a greater degree of top-down inhibition of the amygdala, the capacity for affective control increases. Clinical populations of youths (i.e., youths with disorders treated by exposure) appear to experience even greater deficits during this developmental process (e.g., Liu et al., 2016; Zugman et al., 2021), leading researchers to suggest that some youths may benefit from targeted exposure-augmentation strategies to improve top-down control (Fitzgerald et al., 2021). Available evidence also suggests the potential benefit of intervention strategies that emphasize rewarding stimuli and circumvent the prefrontal cortex (e.g., Meyer et al., 2023). Although continued maturation of prefrontal circuitry among youths may contribute to deficits in exposure learning, it also presents a unique opportunity for intervention during a developmental window in which neural mechanisms for cognitive control are most amenable to modulation (Fitzgerald & Taylor, 2015). Novel augmentation strategies that bolster functioning in key prefrontal circuitry during this time (e.g., noninvasive neuromodulation, Garnaat et al., 2019; behavioral strategies, Fitzgerald et al., 2021) have potential to produce potent and enduring effects into adulthood.

In addition to neurobiological changes, the socioemotional landscape shifts dramatically across development, particularly during adolescence, when school and social activities reorient toward same-age peers and away from caregivers and immediate family. Another potential explanation for attenuated exposure outcomes during adolescence is that caregivers are less involved in treatment (vs. for younger children), but adolescents have not yet independently mastered skills deployed consistently by most adults. Additional research is needed to determine whether age-related effects relate to true developmental differences, use of compensatory treatment strategies (e.g., caregiver involvement), or both. Longitudinal studies will also be essential for future work in this area given that work to date has largely been cross-sectional. This will facilitate clearer understanding of the developmental trajectory underpinning mechanistic changes for youths and may aid in identifying critical or sensitive windows for intervention.

Another potential challenge for understanding mechanisms in exposure-based treatments for youths relates to the common inclusion of various developmental adaptations in treatment packages. In general, packages for youths (vs. adults) are more likely to include tailored delivery strategies (e.g., psychoeducational content, community/school-based delivery) and greater emphasis on other components in addition to exposure (e.g., relaxation, cognitive restructuring, caregiver training; Chorpita & Daleiden, 2009). Although there is clear empirical support specifically for the inclusion of caregiver/family-based treatment components (e.g., Freeman et al., 2018), note that a higher proportion of in-session exposure robustly predicts clinical improvement among youths (Peris et al., 2017)—suggesting that exposure should continue to be emphasized as a primary and necessary treatment component in CBT packages for youths. Note that findings in the pediatric OCD literature (in which protocols typically include a higher proportion of sessions with exposure compared with those for anxiety disorders; Franklin et al., 2019; Freeman & Garcia, 2008) seem to show better long-term maintenance of gains (e.g., 73% remission at 3 years, Melin et al., 2020; 61% at 7–9 years, Fatori et al., 2020). Although these trials cannot be directly compared for a variety of reasons, future studies may wish to explore whether a higher dose of exposure can mitigate against deficits in extinction learning experienced by youths.

Finally, additional consideration should be given to measurement differences that arise across development. Younger children are not yet able to articulate cognitions regarding threat and distress in the same way as adults or even adolescents (Freeman et al., 2011). Youths also exhibit less insight, manifesting as difficulty with cognitive processing and effectively articulating symptoms (Geller et al., 2001). This likely makes it more difficult to leverage cognitive constructs (e.g., expectancy violation) as a clinical tool. It also presents challenges for measuring multiple constructs via self-report

because it may be difficult for youths to reliably report on differential changes across those constructs (e.g., distress vs. expectancy vs. stimulus valence). Although enhanced training (e.g., through psychoeducation) could be considered for bolstering this skill in youths, it will likely be difficult to do so without inadvertently contributing to response bias (see Psychometric Properties of Measures section). Burden associated with multiple measures is also likely to be higher among youths than adults (Norris et al., 2023), particularly through forward translation into clinical settings. Going forward, it will be important to empirically identify optimal streamlined approaches to self-report measurement of mechanism in youths of various ages. Future studies may also wish to place greater emphasis on objective measures (e.g., psychophysiology, observer ratings) as part of multimodal mechanism measurement for youths.

Considerations for culturally diverse populations

Thus far, findings from RCTs examining exposure-based interventions have found mostly comparable outcomes for individuals from diverse backgrounds on measures of clinical response, remission, symptom severity, and overall functioning (e.g., Kline et al., 2020; McLean et al., 2022). However, many of these have been limited by small sample sizes, inadequate analyses examining race and ethnicity, and lack of consideration of intersectionality or sociocultural context (beyond a checkbox identity). Few studies have examined whether hypothesized mechanisms are differentially important for individuals from different historically marginalized communities (e.g., based on race, ethnicity, sexual/gender identity, income). It is likely that sociocultural processes experienced by these groups have an impact on exposure mechanisms. As an example, discrimination is an adverse experience associated with physiological, cognitive, and behavioral changes that increase vulnerability for experiencing anxious symptoms (Lara-Cinisomo et al., 2016; Pascoe & Smart Richman, 2009). According to models of stigmatized identity, experiences of discrimination are associated with increased vigilance to negative cues in the environment (Helzer et al., 2009; Pachankis, 2007), a process associated with heightened levels of anxiety and increased levels of avoidance (Cisler & Koster, 2010). Such cognitive vigilance and physiological reactivity may negatively affect mechanisms relevant to exposure therapy, such as experiential avoidance and threat expectancies. Findings from a primary-care sample of Latine adults suggest that discrimination has an indirect effect on depression, social-anxiety, and anxious-arousal symptoms through experiential avoidance (Zvolensky et al., 2022). Indeed, overreliance on such strategies may contribute to overall psychological distress and affect willingness to experience distressing emotions during exposure. Neurobiological findings also suggest that exposure to negative life experiences (inclusive of violence exposure, low family income, and neighborhood disadvantage) affect mental health and treatment through influence on brain regions that support threat-related emotional functioning. In a prospective study of adolescents who identified as Black Americans and White Americans, which included self-report questionnaires and a Pavlovian fear-conditioning task during fMRI, racial differences in neural (fMRI activity), behavioral (threat expectancy), and psychophysiological responses to threat (Harnett et al., 2019) were found. Racial differences in brain activity to threat were smaller after accounting for negative life experiences. These findings highlight the impact that such experiences may have on symptoms and exposure mechanisms. Sociocultural background may also influence the reliability and validity of self-report measures used to assess exposure mechanisms. For example, individuals from a wide variety of cultures worldwide may be most likely to endorse anxiety using somatic or culture-specific terms (Lewis-Fernández et al., 2010), and men may be more likely than women to describe feelings of anger or irritability when experiencing anxiety and/or sadness (Genovese et al., 2017). Reliance on assessment of limited cognitive or emotional experiences during exposure may miss important information about the experience of mechanistic change for these groups. Significant work remains for the field to understand exposure mechanisms, best methods for measuring them, and treatment response across different sociocultural groups, particularly for those that have been historically marginalized. This information will be critically important for future efforts to optimize culturally responsive assessment and care in the context of exposure therapy.

Discussion

Coordinated action is needed to accelerate research on exposure-therapy mechanisms, particularly related to key issues of conceptualization, measurement, design, analysis, and individual differences. In this article, we outline an initial agenda for addressing specific questions in each of these priority areas. In addition to specific recommendations for future work in each area (see Table 2), several overarching themes emerge. First, considerable work is needed to conceptually integrate overlapping mechanistic constructs, disentangle these from facilitators or readout measures, and establish consistent mechanistic terminology with clear operational definitions. As has been initiated for other psychological

constructs with clear ties to important clinical outcomes (e.g., Eisenberg et al., 2019), extinction research can benefit from focused ontological research to provide greater specification of the link between the construct of extinction and its observable measurement. This will improve replicability and aid in identifying the most potent mechanisms and is likely to require ongoing collaborative work across research groups. Likewise, there is a pressing need for open-source, empirically supported tools that might be deployed across trials, settings, and/or research groups (e.g., reliable and valid measures of mechanism, detailed fidelity measures, clinician/rater training and monitoring protocols, treatment manuals, analytic code, variable calculation procedures, other standard operating procedures). Such a toolbox could improve rigor and replicability across nearly all priority areas outlined in this article and—as empirically supported tools become available—might be hosted by the ETC (see Smits et al., in press) or other groups with a similar mission. A final overarching theme relates to the need for larger sample sizes in mechanistic trials, which will facilitate examination of individual differences, bolster power in the presence of measurement error, and enable use of advanced analytic techniques. A central goal of the ETC is to facilitate research collaborations across sites, leading to larger sample sizes through enhanced recruitment and/or data sharing (Smits et al., in press). As research in these areas progresses and new priorities emerge, we anticipate that the present agenda will continue to evolve. We anticipate that this will eventually lead to new or improved theoretical models of fear and anxiety, which in turn will produce new research questions for the future.

Transparency

Action Editor: DeMond M. Grant
Editor: Jennifer L. Tackett
Author Contributions

Kristen Benito: Conceptualization; Writing – original draft; Writing – review & editing.
Andre Pittig: Conceptualization; Writing – original draft; Writing – review & editing.
Jonathan Abramowitz: Conceptualization; Writing – original draft.
Joanna J. Arch: Conceptualization; Writing – original draft.
Denise Chavira: Conceptualization; Writing – original draft.
Rianne de Kleine: Conceptualization; Writing – original draft.
Alessandro S. De Nadai: Conceptualization; Writing – original draft.
Dirk Hermans: Conceptualization; Writing – original draft.
Stefan G. Hofmann: Conceptualization; Writing – review & editing.
Jürgen Hoyer: Conceptualization; Writing – original draft.
Jonathan D. Huppert: Conceptualization; Writing – original draft.
Katharina Kircanski: Conceptualization; Writing – original draft.
Peter M. McEvoy: Conceptualization; Writing – original draft.
Heidi Meyer: Conceptualization; Writing – original draft.
Marie-H. Monfils: Conceptualization; Writing – original draft.
Santiago Papini: Conceptualization; Writing – original draft.
Winfried Rief: Conceptualization; Writing – original draft.
David Rosenfield: Conceptualization; Writing – original draft.
Eric A. Storch: Conceptualization; Writing – original draft.
Michael J. Telch: Conceptualization; Writing – original draft.
Michael W. Otto: Conceptualization; Writing – original draft; Writing – review & editing.
Jasper A. J. Smits: Conceptualization; Writing – original draft; Writing – review & editing.
Exposure Therapy Consortium: Conceptualization; Writing – original draft.

Declaration of Conflicting Interests

Dr. Otto received compensation for his advisory role with Big Health and received grant support from the National Institute on Drug Abuse (NIDA) and Big Health for effort outside the submitted work. Dr. Smits received funding from NIDA (R21DA049539, R01DA047933) and the National Cancer Institute (NCI; R01CA273221) and compensation from Big Health, Elsevier, American Psychological Association, and Oxford University Press outside the submitted work. Dr. Storch received research funding to his institution from the Ream Foundation, International OCD Foundation, and NIH. He was a consultant for Brainsway and Biohaven Pharmaceuticals. He owned stock less than $5000 in NView (for distribution of the Y-BOCS and CY-BOCS). He received book royalties from Elsevier, Wiley, Oxford, American Psychological Association, Guildford, Springer, Routledge, and Jessica Kingsley. Dr. Rief received research grants from the German Research Foundation DFG (e.g., CRC 289; RTG 2271) and the Hessian Ministry of Science and Arts, and book royalties from Hogrefe publisher. He also received speaker's honorarium from Boehringer Ingelheim. Dr. Hoyer received a research grant from Mindable Health GmbH for effort outside the submitted work.

Funding

Author effort was supported as follows: K. Benito (National Institute of Mental Health [NIMH] R61MH133666, R01MH128595, R01MH112516; National Institute of General Medical Sciences P20GM139743, P20GM130452; Patient Centered Outcomes Research Institute IHS-2017C1-6400), A. Pittig (German Research Foundation PI1269/2-2 389569971), D. Hermans (U Leuven Research Grant C16/19/002), J. D.

Huppert (Israel Science Foundation Grant 1905/20 and Sam Helen Beber Chair of Clinical Psychology), K. Kircanski (NIMH Intramural Research Program ZIAMH002781), H. Meyer (NIMH R00MH119320), M.-H. Monfils (NIMH R01MH125951), S. Papini (National Institute on Drug Abuse R36DA049122), M. W. Otto (NIMH R01MH125949, National Institutes of Health Columbia University Science of Behavior Change Resource and Coordinating Center U24AG052175), J. A. J. Smits (NIMH R01MH125951).

ORCID iDs

Kristen Benito https://orcid.org/0000-0002-1908-9282
Andre Pittig https://orcid.org/0000-0003-3787-9576
Jonathan D. Huppert https://orcid.org/0000-0002-0537-4701
Marie-H. Monfils https://orcid.org/0000-0001-8971-6651
Michael J. Telch https://orcid.org/0000-0003-3444-8500
Jasper A. J. Smits https://orcid.org/0000-0003-1633-9693

Acknowledgments

The listed funding sources had no role in the content or the decision to publish this article.

References

Abramowitz, J. S., Deacon, B. J., & Whiteside, S. P. H. (2019). *Exposure therapy for anxiety: Principles and practice* (2nd ed.). The Guilford Press.

Agren, T., Engman, J., Frick, A., Björkstrand, J., Larsson, E. M., Furmark, T., & Fredrikson, M. (2012). Disruption of reconsolidation erases a fear memory trace in the human amygdala. *Science*, *337*(6101), 1550–1552. https://doi.org/10.1126/science.1223006

Almirall, D., Nahum-Shani, I., Sherwood, N. E., & Murphy, S. A. (2014). Introduction to SMART designs for the development of adaptive interventions: With application to weight loss research. *Translational Behavioral Medicine*, *4*(3), 260–274. https://doi.org/10.1007/s13142-014-0265-0

Anagnostopoulos, F., Liolios, E., Persefonis, G., Slater, J., Kafetsios, K., & Niakas, D. (2012). Physician burnout and patient satisfaction with consultation in primary health care settings: Evidence of relationships from a one-with-many design. *Journal of Clinical Psychology in Medical Settings*, *19*(4), 401–410. https://doi.org/10.1007/s10880-011-9278-8

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 329–341. https://doi.org/10.1080/10705511.2014.915181

Badour, C. L., & Feldner, M. T. (2018). The role of disgust in posttraumatic stress: A critical review of the empirical literature. *Journal of Experimental Psychopathology*, *9*(3), Article pr.032813. https://doi.org/10.5127/pr.032813

Bandura, A. (1988). Self-regulation of motivation and action through goal systems. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 37–61). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-94-009-2792-6_2

Benito, K. G., Herren, J., Freeman, J. B., Garcia, A. M., Block, P., Cantor, E., Chorpita, B. F., Wellen, B., Stewart, E., Georgiadis, C., Frank, H., & Machan, J. (2021). Improving delivery behaviors during exposure for pediatric OCD: A multiple baseline training trial with community therapists. *Behavior Therapy*, *52*(4), 806–820. https://doi.org/10.1016/j.beth.2020.10.003

Benito, K. G., Machan, J., Freeman, J. B., Garcia, A. M., Walther, M., Frank, H., Wellen, B., Stewart, E., Edmunds, J., Kemp, J., Sapyta, J., & Franklin, M. (2018). Measuring fear change within exposures: Functionally-defined habituation predicts outcome in three randomized controlled trials for pediatric OCD. *Journal of Consulting and Clinical Psychology*, *86*(7), 615–630. https://doi.org/10.1037/ccp0000315

Benito, K. G., Machan, J., Freeman, J. B., Garcia, A. M., Walther, M., Frank, H., Wellen, B., Stewart, E., Edmunds, J., Sapyta, J., & Franklin, M. E. (2021). Therapist behavior during exposure tasks predicts habituation and clinical outcome in three randomized controlled trials for pediatric OCD. *Behavior Therapy*, *52*(3), 523–538. https://doi.org/10.1016/j.beth.2020.07.004

Birk, J. L., Otto, M. W., Cornelius, T., Poldrack, R. A., & Edmondson, D. (2023). Improving the rigor of mechanistic behavioral science: The introduction of the checklist for investigating mechanisms in behavior-change research (CLIMBR). *Behavior Therapy*, *54*(4), 708–713. https://doi.org/10.1016/j.beth.2022.12.008

Blakely, T., Lynch, J., Simons, K., Bentley, R., & Rose, S. (2021). Reflection on modern methods: When worlds collide—Prediction, machine learning and causal inference. *International Journal of Epidemiology*, *49*(6), 2058–2064. https://doi.org/10.1093/ije/dyz132

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*(10), 976–986. https://doi.org/10.1016/s0006-3223(02)01546-9

Breuninger, C., Tuschen-Caffier, B., & Svaldi, J. (2019). Dysfunctional cognition and self-efficacy as mediators of symptom change in exposure therapy for agoraphobia—Systematic review and meta-analysis. *Behaviour Research and Therapy*, *120*, Article 103443. https://doi.org/10.1016/j.brat.2019.103443

Brinberg, M., Ram, N., Conroy, D. E., Pincus, A. L., & Gerstorf, D. (2022). Dyadic analysis and the reciprocal one-with-many model: Extending the study of interpersonal processes with intensive longitudinal data. *Psychological Methods*, *27*(1), 65–81. https://doi.org/10.1037/met0000380

Broadbent, A. (2015). Causation and prediction in epidemiology: A guide to the 'Methodological Revolution.*" Studies in History and Philosophy of Science Part C: Biological and Biomedical Sciences*, *54*, 72–80. https://doi.org/10.1016/j.shpsc.2015.06.004

Brunet, A., Orr, S. P., Tremblay, J., Robertson, K., Nader, K., & Pitman, R. K. (2008). Effect of post-retrieval propranolol on psychophysiologic responding during subsequent

script-driven traumatic imagery in post-traumatic stress disorder. *Journal of Psychiatric Research*, *42*(6), 503–506. https://doi.org/10.1016/j.jpsychires.2007.05.006

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., & 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. https://doi.org/10.1002/da.22728

Cervin, M., & Perrin, S. (2021). Incompleteness and disgust predict treatment outcome in pediatric obsessive-compulsive disorder. *Behavior Therapy*, *52*(1), 53–63. https://doi.org/10.1016/j.beth.2020.01.007

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*(3), 566–579. https://doi.org/10.1037/a0014565

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review*, *30*(2), 203–216. https://doi.org/10.1016/j.cpr.2009.11.003

Cisler, J. M., Olatunji, B. O., & Lohr, J. M. (2009). Disgust, fear, and the anxiety disorders: A critical review. *Clinical Psychology Review*, *29*(1), 34–46. https://doi.org/10.1016/j.cpr.2008.09.007

Connell, A. M. (2009). Employing complier average causal effect analytic methods to examine effects of randomized encouragement trials. *The American Journal of Drug and Alcohol Abuse*, *35*(4), 253–259. https://doi.org/10.1080/00952990903005882

Constantinou, E., Purves, K. L., McGregor, T., Lester, K. J., Barry, T. J., Treanor, M., Craske, M. G., & Eley, T. C. (2021). Measuring fear: Association among different measures of fear learning. *Journal of Behavior Therapy and Experimental Psychiatry*, *70*, Article 101618. https://doi.org/10.1016/j.jbtep.2020.101618

Cooper, A. A., Clifton, E. G., & Feeny, N. C. (2017). An empirical review of potential mediators and mechanisms of prolonged exposure therapy. *Clinical Psychology Review*, *56*, 106–121. https://doi.org/10.1016/j.cpr.2017.07.003

Cordovilla-Guardia, S., Fernández-Mondéjar, E., Vilar-López, R., Navas, J. F., Portillo-Santamaría, M., Rico-Martín, S., & Lardelli-Claret, P. (2017). Effect of a brief intervention for alcohol and illicit drug use on trauma recidivism in a cohort of trauma patients. *PLOS ONE*, *12*(8), Article e0182441. https://doi.org/10.1371/journal.pone.0182441

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. https://doi.org/10.1016/j.brat.2007.10.003

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. https://doi.org/10.1016/j.brat.2014.04.006

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Davis, M. (2006). Neural systems involved in fear and anxiety measured with fear-potentiated startle. *The American Psychologist*, *61*(8), 741–756. https://doi.org/10.1037/0003-066X.61.8.741

De Nadai, A. S., Fitzgerald, K. D., Norman, L. J., Russman Block, S. R., Mannella, K. A., Himle, J. A., & Taylor, S. F. (2023). Defining brain-based OCD patient profiles using task-based fMRI and unsupervised machine learning. *Neuropsychopharmacology*, *48*(2), 402–409. https://doi.org/10.1038/s41386-022-01353-x

De Nadai, A. S., Hu, Y., & Thompson, W. K. (2022). Data pollution in neuropsychiatry—An under-recognized but critical barrier to research progress. *JAMA Psychiatry*, *79*(2), 97–98. https://doi.org/10.1001/jamapsychiatry.2021.2812

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), Article 2319. https://doi.org/10.1038/s41467-019-10301-1

Elsey, J. W. B., & Kindt, M. (2017). Tackling maladaptive memories through reconsolidation: From neural to clinical science. *Neurobiology of Learning and Memory*, *142*(Pt. A), 108–117. https://doi.org/10.1016/j.nlm.2017.03.007

Ewbank, M. P., Cummins, R., Tablan, V., Bateup, S., Catarino, A., Martin, A. J., & Blackwell, A. D. (2020). Quantifying the association between psychotherapy content and clinical outcomes using deep learning. *JAMA Psychiatry*, *77*(1), 35–43. https://doi.org/10.1001/jamapsychiatry.2019.2664

Falkenström, F., Solomonov, N., & Rubel, J. A. (2022). How to model and interpret cross-lagged effects in psychotherapy mechanisms of change research: A comparison of multilevel and structural equation models. *Journal of Consulting and Clinical Psychology*, *90*(5), 446–458. https://doi.org/10.1037/ccp0000727

Fanselow, M. S., & Pennington, Z. T. (2018). A return to the psychiatric dark ages with a two-system framework for fear. *Behaviour Research and Therapy*, *100*, 24–29. https://doi.org/10.1016/j.brat.2017.10.012

Fatori, D., Polanczyk, G. V., de Morais, R. M. C. B., & Asbahr, F. R. (2020). Long-term outcome of children and adolescents with obsessive-compulsive disorder: A 7–9-year follow-up of a randomized clinical trial. *European Child & Adolescent Psychiatry*, *29*(11), 1613–1616. https://doi.org/10.1007/s00787-019-01457-8

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fitzgerald, K. D., Schroder, H. S., & Marsh, R. (2021). Cognitive control in pediatric obsessive-compulsive and anxiety disorders: Brain-behavioral targets for early intervention. *Biological Psychiatry*, *89*(7), 697–706. https://doi.org/10.1016/j.biopsych.2020.11.012

Fitzgerald, K. D., & Taylor, S. F. (2015). Error-processing abnormalities in pediatric anxiety and obsessive compul-

sive disorders. *CNS Spectrums*, *20*(4), 346–354. https://doi.org/10.1017/S1092852915000036

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, *99*(1), 20–35. https://doi.org/10.1037/0033-2909.99.1.20

Fradkin, I., Adams, R. A., Parr, T., Roiser, J. P., & Huppert, J. D. (2020). Searching for an anchor in an unpredictable world: A computational model of obsessive-compulsive disorder. *Psychological Review*, *127*(5), 672–699. https://doi.org/10.1037/rev0000188

Franklin, M. E., Freeman, J. B., & March, J. S. (2019). *Treating OCD in children and adolescents: A cognitive-behavioral approach*. The Guilford Press.

Freeman, J., Benito, K., Herren, J., Kemp, J., Sung, J., Georgiadis, C., Arora, A., Walther, M., & Garcia, A. (2018). Evidence base update of psychosocial treatments for pediatric obsessive-compulsive disorder: Evaluating, improving, and transporting what works. *Journal of Clinical Child and Adolescent Psychology*, *47*(5), 669–698. https://doi.org/10.1080/15374416.2018.1496443

Freeman, J., Flessner, C. A., & Garcia, A. M. (2011). The Children's Yale-Brown Obsessive-Compulsive Scale: Reliability and validity for use among 5 to 8 year olds with obsessive-compulsive disorder. *Journal of Abnormal Child Psychology*, *39*, 877–883. https://doi.org/10.1007/s10802-011-9494-6

Freeman, J. B., & Garcia, A. M. (2008). *Family based treatment for young children with OCD: Therapist guide*. Oxford University Press.

Garnaat, S. L., Conelea, C. A., McLaughlin, N. C. R., & Benito, K. (2019). Pediatric OCD in the era of RDoC. *Journal of Obsessive-Compulsive and Related Disorders*, *23*, Article 100385. https://doi.org/10.1016/j.jocrd.2018.03.002

Geller, D. A., Biederman, J., Faraone, S., Agranat, A., Cradock, K., Hagermoser, L., Kim, G., Frazier, J., & Coffey, B. J. (2001). Developmental aspects of obsessive-compulsive disorder: Findings in children, adolescents, and adults. *The Journal of Nervous and Mental Disease*, *189*(7), 471–477. https://doi.org/10.1097/00005053-200107000-00009

Genovese, T., Dalrymple, K., Chelminski, I., & Zimmerman, M. (2017). Subjective anger and overt aggression in psychiatric outpatients. *Comprehensive Psychiatry*, *73*, 23–30. https://doi.org/10.1016/j.comppsych.2016.10.008

Giedd, J., Blumenthal, J., Jeffries Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, *2*, 861–863. https://doi.org/10.1038/13158

Ginsburg, G. S., Becker-Haimes, E. M., Keeton, C., Kendall, P. C., Iyengar, S., Sakolsky, D., Albano, A. M., Peris, T., Compton, S. N., & Piacentini, J. (2018). Results from the child/adolescent anxiety multimodal extended long-term study (CAMELS): Primary anxiety outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *57*(7), 471–480. https://doi.org/10.1016/j.jaac.2018.03.017

Gloster, A. T., Wittchen, H. U., Einsle, F., Lang, T., Helbig-Lang, S., Fydrich, T., Fehm, L., Hamm, A. O., Richter, J., Alpers, G. W., Gerlach, A. L., Ströhle, A., Kircher, T., Deckert, J., Zwanzger, P., Höfler, M., & Arolt, V. (2011). Psychological treatment for panic disorder with agoraphobia: A randomized controlled trial to examine the role of therapist-guided exposure in situ in CBT. *Journal of Consulting and Clinical Psychology*, *79*(3), 406–420. https://doi.org/10.1037/a0023584

Guolo, A. (2008). Robust techniques for measurement error correction: A review. *Statistical Methods in Medical Research*, *17*(6), 555–580. https://doi.org/10.1177/0962280207081318

Haaker, J., Maren, S., Andreatta, M., Merz, C. J., Richter, J., Richter, S. H., Meir Drexler, S., Lange, M. D., Jüngling, K., Nees, F., Seidenbecher, T., Fullana, M. A., Wotjak, C. T., & Lonsdorf, T. B. (2019). Making translation work: Harmonizing cross-species methodology in the behavioural neuroscience of Pavlovian fear conditioning. *Neuroscience & Biobehavioral Reviews*, *107*, 329–345.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hamm, A. O. (2020). Fear, anxiety, and their disorders from the perspective of psychophysiology. *Psychophysiology*, *57*(2), Article e13474. https://doi.org/10.1111/psyp.13474

Harnett, N. G., Wheelock, M. D., Wood, K. H., Goodman, A. M., Mrug, S., Elliott, M. N., Schuster, M. A., Tortolero, S., & Knight, D. C. (2019). Negative life experiences contribute to racial differences in the neural response to threat. *NeuroImage*, *202*, Article 116086. https://doi.org/10.1016/j.neuroimage.2019.116086

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247. https://doi.org/10.1037/1040-3590.7.3.238

Helzer, E. G., Connor-Smith, J. K., & Reed, M. A. (2009). Traits, states, and attentional gates: Temperament and threat relevance as predictors of attentional bias to social threat. *Anxiety, Stress, and Coping*, *22*(1), 57–76. https://doi.org/10.1080/10615800802272244

Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child and Adolescent Psychology*, *45*(2), 91–113. https://doi.org/10.1080/15374416.2015.1046177

Himle, M. B. (2015). Let truth be thy aim, not victory: Comment on theory-based exposure process. *Journal of Obsessive-Compulsive and Related Disorders*, *6*, 183–190. https://doi.org/10.1016/j.jocrd.2015.03.001

Hodgson, R., & Rachman, S. (1974). II. Desynchrony in measures of fear. *Behaviour Research and Therapy*, *12*(4), 319–326. https://doi.org/10.1016/0005-7967(74)90006-0

Hofmann, S. G. (2008). Cognitive processes during fear acquisition and extinction in animals and humans: Implications for exposure therapy of anxiety disorders. *Clinical Psychology Review*, *28*, 199–210. https://doi.org/10.1016/j.cpr.2007.04.009

Hollandt, M., Wroblewski, A., Yang, Y., Ridderbusch, I. C., Kircher, T., Hamm, A. O., Straube, B., & Richter, J. (2020). Facilitating translational science in anxiety disorders by adjusting extinction training in the laboratory

to exposure-based therapy procedures. *Translational Psychiatry*, *10*(1), Article 110. https://doi.org/10.1038/s41398-020-0786-x

Hu, S., Pruessner, J. C., Coupé, P., & Collins, D. L. (2013). Volumetric analysis of medial temporal lobe structures in brain development from childhood to adolescence. *NeuroImage*, *74*, 276–287. https://doi.org/10.1016/j.neuroimage.2013.02.032

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Jaycox, L. H., Foa, E. B., & Morral, A. R. (1998). Influence of emotional engagement and habituation on exposure therapy for PTSD. *Journal of Consulting and Clinical Psychology*, *66*(1), 185–192. https://doi.org/10.1037//0022-006x.66.1.185

Kamphuis, J. H., & Telch, M. J. (2000). Effects of distraction and guided threat reappraisal on fear reduction during exposure-based treatments for specific fears. *Behaviour Research and Therapy*, *38*(12), 1163–1181. https://doi.org/10.1016/s0005-7967(99)00147-3

Kazantzis, N., Brownfield, N. R., Mosely, L., Usatoff, A. S., & Flighty, A. J. (2017). Homework in cognitive behavioral therapy: A systematic review of adherence assessment in anxiety and depression (2011-2016). *The Psychiatric Clinics of North America*, *40*(4), 625–639. https://doi.org/10.1016/j.psc.2017.08.001

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1–27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Kenny, D. A., West, T. V., Malloy, T. E., & Albright, L. (2006). Componential analysis of interpersonal perception data. *Personality and Social Psychology Review*, *10*(4), 282–294. https://doi.org/10.1207/s15327957pspr1004_1

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., . . . Interoception Summit 2016 Participants. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *3*(6), 501–513. https://doi.org/10.1016/j.bpsc.2017.12.004

Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: Erasing human fear responses and preventing the return of fear. *Nature Neuroscience*, *12*(3), 256–258. https://doi.org/10.1038/nn.2271

Kline, A. C., Feeny, N. C., & Zoellner, L. A. (2020). Race and cultural factors in an RCT of prolonged exposure and sertraline for PTSD. *Behaviour Research and Therapy*, *132*, Article 103690. https://doi.org/10.1016/j.brat.2020.103690

Klingelhöfer-Jens, M., Ehlers, M. R., Kuhn, M., Keyaniyan, V., & Lonsdorf, T. B. (2022). Robust group—But limited individual-level (longitudinal) reliability and insights into cross-phases response prediction of conditioned fear. *eLife*, *11*, Article e78717. https://doi.org/10.7554/eLife.78717

Knowles, K. A., & Tolin, D. F. (2022). Mechanisms of action in exposure therapy. *Current Psychiatry Reports*, *24*(12), 861–869. https://doi.org/10.1007/s11920-022-01391-8

Koban, L., Schneider, R., Ashar, Y. K., Andrews-Hanna, J. R., Landy, L., Moscovitch, D. A., Wager, T. D., & Arch, J. J. (2017). Social anxiety is characterized by biased learning about performance and the self. *Emotion*, *17*(8), 1144–1155. https://doi.org/10.1037/emo0000296

Kodal, A., Fjermestad, K., Bjelland, I., Gjestad, R., Öst, L. G., Bjaastad, J. F., Haugland, B. S. M., Havik, O. E., Heiervang, E., & Wergeland, G. J. (2018). Long-term effectiveness of cognitive behavioral therapy for youth with anxiety disorders. *Journal of Anxiety Disorders*, *53*, 58–67. https://doi.org/10.1016/j.janxdis.2017.11.003

Kredlow, M. A., Eichenbaum, H., & Otto, M. W. (2018). Memory creation and modification: Enhancing the treatment of psychological disorders. *The American Psychologist*, *73*(3), 269–285. https://doi.org/10.1037/amp0000185

Krypotos, A. M., Vervliet, B., & Engelhard, I. M. (2018). The validity of human avoidance paradigms. *Behaviour Research and Therapy*, *111*, 99–105. https://doi.org/10.1016/j.brat.2018.10.011

Lancaster, C. L., Monfils, M. H., & Telch, M. J. (2020). Augmenting exposure therapy with pre-extinction fear memory reactivation and deepened extinction: A randomized controlled trial. *Behaviour Research and Therapy*, *135*, Article 103730. https://doi.org/10.1016/j.brat.2020.103730

Lang, P. J. (1971). The application of psychophysiological methods to the study of psychotherapy and behavior modification. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 75–125). John Wiley & Sons.

Lara-Cinisomo, S., Girdler, S. S., Grewen, K., & Meltzer-Brody, S. (2016). A biopsychosocial conceptual framework of postpartum depression risk in immigrant and U.S.-born Latina mothers in the United States. *Women's Health Issues*, *26*(3), 336–343. https://doi.org/10.1016/j.whi.2016.02.006

Lau, J. Y., Britton, J. C., Nelson, E. E., Angold, A., Ernst, M., Goldwin, M., Grillon, C., Leibenluft, E., Lissek, S., Norcross, M., Shiffrin, N., & Pine, D. S. (2011). Distinct neural signatures of threat learning in adolescents and adults. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4500–4505. https://doi.org/10.1073/pnas.1005494108

LeDoux, J. E., & Hofmann, S. G. (2018). The subjective experience of emotion: A fearful view. *Current Opinion in Behavioral Sciences*, *19*, 67–72. https://doi.org/10.1016/j.cobeha.2017.09.011

LeDoux, J. E., & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: A two-system framework. *The American Journal of Psychiatry*, *173*(11), 1083–1093. https://doi.org/10.1176/appi.ajp.2016.16030353

Lewis-Fernández, R., Hinton, D. E., Laria, A. J., Patterson, E. H., Hofmann, S. G., Craske, M. G., Stein, D. J., Asnaani, A., & Liao, B. (2010). Culture and the anxiety disorders: Recommendations for DSM-V. *Depression and Anxiety*, *27*(2), 212–229. https://doi.org/10.1002/da.20647

Liu, Y., Bilek, E. L., & Fitzgerald, K. D. (2016). Developmental neuroimaging in pediatric obsessive-compulsive disorder. *Current Behavioral Neuroscience Reports*, *3*(3), 193–203. https://doi.org/10.1007/s40473-016-0086-1

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., . . . 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 and Biobehavioral Reviews*, *77*, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026

Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning & Memory*, *11*(5), 495–500. https://doi.org/10.1101/lm.79604

Lubin, R. E., Fitzgerald, H. E., Rosenfield, D., Carpenter, J. K., Papini, S., Dutcher, C. D., Dowd, S. M., Hofmann, S. G., Pollack, M. H., Smits, J. A. J., & Otto, M. W. (2023). Using pre-treatment de novo threat conditioning outcomes to predict treatment response to DCS augmentation of exposure-based CBT. *Journal of Psychiatric Research*, *164*, 357–363. https://doi.org/10.1016/j.jpsychires.2023.06.008

Marcus, D. K., Kashy, D. A., Wintersteen, M. B., & Diamond, G. S. (2011). The therapeutic alliance in adolescent substance abuse treatment: A one-with-many analysis. *Journal of Counseling Psychology*, *58*(3), 449–455. https://doi.org/10.1037/a0023196

Marin, M. F., Zsido, R. G., Song, H., Lasko, N. B., Killgore, W. D. S., Rauch, S. L., Simon, N. M., & 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. https://doi.org/10.1001/jamapsychiatry.2017.0329

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). American Psychological Association. https://doi.org/10.1037/10409-005

McLean, C. P., Levy, H. C., Miller, M. L., & Tolin, D. F. (2022). Exposure therapy for PTSD: A meta-analysis. *Clinical Psychology Review*, *91*, Article 102115. https://doi.org/10.1016/j.cpr.2021.102115

McNally, R. J. (2007). Mechanisms of exposure therapy: How neuroscience can improve psychological treatments for anxiety disorders. *Clinical Psychology Review*, *27*(6), 750–759. https://doi.org/10.1016/j.cpr.2007.01.003

Melin, K., Skarphedinsson, G., Thomsen, P. H., Weidle, B., Torp, N. C., Valderhaug, R., Højgaard, D. R. M. A., Hybel, K. A., Nissen, J. B., Jensen, S., Dahl, K., Skärsäter, I., Haugland, B. S., & Ivarsson, T. (2020). Treatment gains are sustainable in pediatric obsessive-compulsive disorder: Three-year follow-up from the NordLOTS. *Journal of the American Academy of Child and Adolescent Psychiatry*, *59*(2), 244–253. https://doi.org/10.1016/j.jaac.2019.01.010

Meyer, H. C., Fields, A., Vannucci, A., Gerhard, D. M., Bloom, P. A., Heleniak, C., Opendak, M., Sullivan, R., Tottenham, N., Callaghan, B. L., & Lee, F. S. (2023). The added value of crosstalk between developmental circuit neuroscience and clinical practice to inform the treatment of adolescent anxiety. *Biological Psychiatry Global Open Science*, *3*(2), 169–178. https://doi.org/10.1016/j.bpsgos.2022.02.002

Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience: Ten years of progress. *Annual Review of Psychology*, *63*, 129–151. https://doi.org/10.1146/annurev.psych.121208.131631

Milgram, L., Sheehan, K., Cain, G., Carper, M. M., O'Connor, E. E., Freeman, J. B., Garcia, A., Case, B., & Benito, K. B. (2022). Comparison of patient-reported distress during harm avoidance and incompleteness exposure tasks for youth with OCD. *Journal of Obsessive-Compulsive and Related Disorders*, *35*, Article 100760. https://doi.org/10.1016/j.jocrd.2022.100760

Moerkerke, B., Loeys, T., & Vansteelandt, S. (2015). Structural equation modeling versus marginal structural modeling for assessing mediation in the presence of posttreatment confounding. *Psychological Methods*, *20*(2), 204–220. https://doi.org/10.1037/a0036368

Monfils, M. H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: Key to persistent attenuation of fear memories. *Science*, *324*(5929), 951–955. https://doi.org/10.1126/science.1167975

Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: Opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. The Lancet. *Psychiatry*, *5*(12), 1032–1042. https://doi.org/10.1016/S2215-0366(18)30270-0

Mystkowski, J. L., Craske, M. G., Echiverri, A. M., & Labus, J. S. (2006). Mental reinstatement of context and return of fear in spider-fearful participants. *Behavior Therapy*, *37*(1), 49–60. https://doi.org/10.1016/j.beth.2005.04.001

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726. https://doi.org/10.1038/35021052

Nielsen, L., Riddle, M., & King, J. W., NIH Science of Behavior Change Implementation Team, Aklin, W. M., Chen, W., Clark, D., Collier, E., Czajkowski, S., Esposito, L., Ferrer, R., Green, P., Hunter, C., Kehl, K., King, R., Onken, L., Simmons, J. M., Stoeckel, L., Stoney, C., . . . Weber, W. (2018). The NIH science of behavior change program: Transforming the science through a focus on mechanisms of change. *Behaviour Research and Therapy*, *101*, 3–11. https://doi.org/10.1016/j.brat.2017.07.002

Norris, L. A., Rabner, J. C., Crane, M. E., Cervin, M., Ney, J. S., Benito, K. G., Kendall, P. C., & Frank, H. E. (2023). What caregivers like the most (and least) about cognitive behavioral therapy for youth anxiety: A mixed methods approach. *Journal of Anxiety Disorders*, *98*, Article 102742. https://doi.org/10.1016/j.janxdis.2023.102742

Öhman, A. (2008). Fear and anxiety. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 709–729). The Guilford Press.

Olatunji, B. O., Forsyth, J. P., & Cherian, A. (2007). Evaluative differential conditioning of disgust: A sticky form of relational learning that is resistant to extinction. *Journal of Anxiety Disorders*, *21*(6), 820–834. https://doi.org/10.1016/j.janxdis.2006.11.004

Otto, M. W., Eastman, A., Lo, S., Hearon, B. A., Bickel, W. K., Zvolensky, M., Smits, J. A., & Doan, S. N. (2016). Anxiety sensitivity and working memory capacity: Risk factors and targets for health behavior promotion. *Clinical Psychology Review*, *49*, 67–78. https://doi.org/10.1016/j.cpr.2016.07.003

Pachankis, J. E. (2007). The psychological implications of concealing a stigma: A cognitive-affective-behavioral model. *Psychological Bulletin*, *133*(2), 328–345. https://doi.org/10.1037/0033-2909.133.2.328

Pascoe, E. A., & Smart Richman, L. (2009). Perceived discrimination and health: A meta-analytic review. *Psychological Bulletin*, *135*(4), 531–554. https://doi.org/10.1037/a0016059

Pattwell, S. S., Duhoux, S., Hartley, C. A., Johnson, D. C., Jing, D., Elliott, M. D., Ruberry, E. J., Powers, A., Mehta, N., Yang, R. R., Soliman, F., Glatt, C. E., Casey, B. J., Momam, I., & Lee, F. S. (2012). Altered fear learning across development in both mouse and human. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16318–16323. https://doi.org/10.1073/pnas.1206834109

Pearl, J. (2019a). The limitations of opaque learning machines. In J. Brockman (Ed.), *Possible minds: 25 ways of looking at AI* (pp. 13–19). Penguin Press.

Pearl, J. (2019b). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, *62*(3), 54–60. https://doi.org/10.1145/3241036

Peris, T. S., Caporino, N. E., O'Rourke, S., Kendall, P. C., Walkup, J. T., Albano, A. M., Bergman, R. L., McCracken, J. T., Birmaher, B., Ginsburg, G. S., Sakolsky, D., Piacentini, J., & Compton, S. N. (2017). Therapist-reported features of exposure tasks that predict differential treatment outcomes for youth with anxiety. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(12), 1043–1052. https://doi.org/10.1016/j.jaac.2017.10.001

Phelps, E. A., & Hofmann, S. G. (2019). Memory editing: From science fiction to clinical practice. *Nature*, *572*, 43–50. https://doi.org/10.1038/s41586-019-1433-7

Pitman, R. K., Milad, M. R., Igoe, S. A., Vangel, M. G., Orr, S. P., Tsareva, A., Gamache, K., & Nader, K. (2011). Systemic mifepristone blocks reconsolidation of cue-conditioned fear; Propranolol prevents this effect. *Behavioral Neuroscience*, *125*(4), 632–638. https://doi.org/10.1037/a0024364

Pittig, A., Heinig, I., Goerigk, S., Thiel, F., Hummel, K., Scholl, L., Deckert, J., Pauli, P., Domschke, K., Lueken, U., Fydrich, T., Fehm, L., Plag, J., Ströhle, A., Kircher, T., Straube, B., Rief, W., Koelkebeck, K., Arolt, V., . . . Wittchen, H. U. (2021). Efficacy of temporally intensified exposure for anxiety disorders: A multicenter randomized clinical trial. *Depression and Anxiety*, *38*(11), 1169–1181. https://doi.org/10.1002/da.23204

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. https://doi.org/10.1097/YCO.0000000000000220

Pittig, A., Wong, A. H. K., Glück, V. M., & Boschet, J. M. (2020). Avoidance and its bi-directional relationship with conditioned fear: Mechanisms, moderators, and clinical implications. *Behaviour Research and Therapy*, *126*, Article 103550. https://doi.org/10.1016/j.brat.2020.103550

Pugh, L. R., Taylor, P. J., & Berry, K. (2015). The role of guilt in the development of post-traumatic stress disorder: A systematic review. *Journal of Affective Disorders*, *182*, 138–150. https://doi.org/10.1016/j.jad.2015.04.026

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., Coppola, G., Geyer, M. A., Glanzman, D. L., Marsland, S., McSweeney, F. K., Wilson, D. A., Wu, C. F., & Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, *92*(2), 135–138. https://doi.org/10.1016/j.nlm.2008.09.012

Rao-Ruiz, P., Rotaru, D. C., van der Loo, R. J., Mansvelder, H. D., Stiedl, O., Smit, A. B., & Spijker, S. (2011). Retrieval-specific endocytosis of GluA2-AMPARs underlies adaptive reconsolidation of contextual fear. *Nature Neuroscience*, *14*(10), 1302–1308. https://doi.org/10.1038/nn.2907

Richter, J., Pané-Farré, C. A., Gerlach, A. L., Gloster, A. T., Wittchen, H. U., Lang, T., Alpers, G. W., Helbig-Lang, S., Deckert, J., Fydrich, T., Fehm, L., Ströhle, A., Kircher, T., Arolt, V., & Hamm, A. O. (2021). Transfer of exposure therapy effects to a threat context not considered during treatment in patients with panic disorder and agoraphobia: Implications for potential mechanisms of change. *Behaviour Research and Therapy*, *142*, Article 103886. https://doi.org/10.1016/j.brat.2021.103886

Rosenfield, D., Smits, J. A. J., Hofmann, S. G., Mataix-Cols, D., de la Cruz, L. F., Andersson, E., Rück, C., Monzani, B., Pérez-Vigil, A., Frumento, P., Davis, M., de Kleine, R. A., Difede, J., Dunlop, B. W., Farrell, L. J., Geller, D., Gerardi, M., Guastella, A. J., Hendriks, G. J., . . . Otto, M. W. (2019). Changes in dosing and dose timing of D-Cycloserine explain its apparent declining efficacy for augmenting exposure therapy for anxiety-related disorders: An individual participant-data meta-analysis. *Journal of Anxiety Disorders*, *68*, Article 102149. https://doi.org/10.1016/j.janxdis.2019.102149

Rothman, A. J., & Sheeran, P. (2021). The operating conditions framework: Integrating mechanisms and moderators in health behavior interventions. *Health Psychology*, *40*(12), 845–857. https://doi.org/10.1037/hea0001026

Saraiya, T., & Lopez-Castro, T. (2016). Ashamed and afraid: A scoping review of the role of shame in post-traumatic stress disorder (PTSD). *Journal of Clinical Medicine*, *5*(11), Article 94. https://doi.org/10.3390/jcm5110094

Sauer-Zavala, S., Southward, M. W., Terrill, D. R., Semcho, S. A., & Stumpp, N. E. (2024). Mechanism engagement as a potential evidence-based approach to personalized treatment termination. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *34*(1), 124–136. https://doi.org/10.1080/10503307.2023.2168574

Savović, J., Jones, H. E., Altman, D. G., Harris, R. J., Jüni, P., Pildal, J., Als-Nielsen, B., Balk, E. M., Gluud, C., Gluud, L. L., Ioannidis, J. P., Schulz, K. F., Beynon, R., Welton, N. J., Wood, L., Moher, D., Deeks, J. J., & Sterne, J. A. (2012). Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Annals of Internal Medicine*, *157*(6), 429–438. https://doi.org/10.7326/0003-4819-157-6-201209180-00537

Scheveneels, S., Boddez, Y., & Hermans, D. (2021). Predicting clinical outcomes via human fear conditioning: A narrative review. *Behaviour Research and Therapy*, *142*, Article 103870. https://doi.org/10.1016/j.brat.2021.103870

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. https://doi.org/10.1016/j.brat.2016.08.015

Schloss, H. M., & Haaga, D. A. (2011). Interrelating behavioral measures of distress tolerance with self-reported experiential avoidance. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, *29*(1), 53–63. https://doi.org/10.1007/s10942-011-0127-3

Seligman, L. D., & Ollendick, T. H. (2011). Cognitive-behavioral therapy for anxiety disorders in youth. *Child and Adolescent Psychiatric Clinics of North America*, *20*(2), 217–238. https://doi.org/10.1016/j.chc.2011.01.003

Shumake, J., & Monfils, M. H. (2015). Assessing fear following retrieval + extinction through suppression of baseline reward seeking vs. freezing. *Frontiers in Behavioral Neuroscience*, *9*, Article 355. https://doi.org/10.3389/fnbeh.2015.00355

Smits, J. A. J., Abramowitz, J. S., Arch, J. J., Papini, S., Anderson, R., Dixon, L., Graham, B. M., Hofmann, S. G., Hoyer, J., Huppert, J. D., Jacquart, J., Johnson, D., McEvoy, P. M., McKay, D., Newby, J., Otto, M. W., Pittig, A., Rief, W., Rosenfield, D., . . . Exposure Therapy Consortium. (in press). Improving exposure therapy: Rationale for and design of a global consortium. *Psychiatric Clinics of North America*.

Smits, J. A. J., Rosenfield, D., Otto, M. W., Marques, L., Davis, M. L., Meuret, A. E., Simon, N. M., Pollack, M. H., & Hofmann, S. G. (2013). D-cycloserine enhancement of exposure therapy for social anxiety disorder depends on the success of exposure sessions. *Journal of Psychiatric Research*, *47*(10), 1455–1461. https://doi.org/10.1016/j.jpsychires.2013.06.020

Song, M., Jones, C. E., Monfils, M.-H., & Niv, Y. (2022). *Explaining the effectiveness of fear extinction through latent-cause inference*. arXiv. https://doi.org/10.48550/arXiv.2205.04670

Southward, M. W., & Sauer-Zavala, S. (2020). Experimental manipulations to test theory-driven mechanisms of cognitive behavior therapy. *Frontiers in Psychiatry*, *11*, Article 603009. https://doi.org/10.3389/fpsyt.2020.603009

Sripada, R. K., & Rauch, S. A. (2015). Between-session and within-session habituation in prolonged exposure therapy for posttraumatic stress disorder: A hierarchical linear modeling approach. *Journal of Anxiety Disorders*, *30*, 81–87. https://doi.org/10.1016/j.janxdis.2015.01.002

Summerfeldt, L. J. (2004). Understanding and treating incompleteness in obsessive-compulsive disorder. *Journal of Clinical Psychology*, *60*(11), 1155–1168. https://doi.org/10.1002/jclp.20080

Surís, A., Smith, J., Powell, C., & North, C. S. (2013). Interfering with the reconsolidation of traumatic memory: Sirolimus as a novel agent for treating veterans with posttraumatic stress disorder. *Annals of Clinical Psychiatry*, *25*(1), 33–40.

Swift, J. K., Greenberg, R. P., Tompkins, K. A., & Parkin, S. R. (2017). Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: A meta-analysis of head-to-head comparisons. *Psychotherapy*, *54*(1), 47–57. https://doi.org/10.1037/pst0000104

Taschereau-Dumouchel, V., Michel, M., Lau, H., Hofmann, S. G., & LeDoux, J. E. (2022). Putting the "mental" back in "mental disorders": A perspective from research on fear and anxiety. *Molecular Psychiatry*, *27*, 1322–1330. https://doi.org/10.1038/s41380-021-01395-5

Tolman, E. C., & Honzik, C. H. (1930). *Introduction and removal of reward, and maze performance in rats*. University of California Publications in Psychology.

Tuerk, P. W. (2014). Starting from something: Augmenting exposure therapy and methods of inquiry. *The American Journal of Psychiatry*, *171*(10), 1034–1037. https://doi.org/10.1176/appi.ajp.2014.14070880

Turner, C., O'Gorman, B., Nair, A., & O'Kearney, R. (2018). Moderators and predictors of response to cognitive behaviour therapy for pediatric obsessive-compulsive disorder: A systematic review. *Psychiatry Research*, *261*, 50–60. https://doi.org/10.1016/j.psychres.2017.12.034

Twohig, M. P., Abramowitz, J. S., Bluett, E. J., Fabricant, L. E., Jacoby, R. J., Morrison, K. L., Reuman, L., & Smith, B. M. (2015). Exposure therapy for OCD from an acceptance and commitment therapy (ACT) framework. *Journal of Obsessive-Compulsive and Related Disorders*, *6*, 167–173. https://doi.org/10.1016/j.jocrd.2014.12.007

U.S. Preventive Services Task Force. (2020). *Screening for anxiety in adults*. https://uspreventiveservicestaskforce.org/uspstf/document/final-research-plan/anxiety-adults-screening

Valente, M. J., Pelham, W. E., Smyth, H., & MacKinnon, D. P. (2017). Confounding in statistical mediation analysis: What it is and how to address it. *Journal of Counseling Psychology*, *64*(6), 659–671. https://doi.org/10.1037/cou0000242

van der Laan, M. J., & Rose, S. (2011). *Targeted learning*. Springer.

van Minnen, A., & Hagenaars, M. (2002). Fear activation and habituation patterns as early process predictors of response to prolonged exposure treatment in PTSD. *Journal of Traumatic Stress*, *15*(5), 359–367. https://doi.org/10.1023/A:1020177023209

Walz, L. C., Nauta, M. H., & aan het Rot, M. (2014). Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: A systematic review. *Journal of Anxiety Disorders*, *28*(8), 925–937.

Wilcox, R., Carlson, M., Azen, S., & Clark, F. (2013). Avoid lost discoveries, because of violations of standard assumptions, by using modern robust statistical methods. *Journal of Clinical Epidemiology*, *66*(3), 319–329. https://doi.org/10.1016/j.jclinepi.2012.09.003

Wood, N. E., Rosasco, M. L., Suris, A. M., Spring, J. D., Marin, M. F., Lasko, N. B., Goetz, J. M., Fischer, A. M., Orr, S. P., & Pitman, R. K. (2015). Pharmacological blockade of memory reconsolidation in posttraumatic stress disorder: Three negative psychophysiological studies. *Psychiatry Research*, *225*(1–2), 31–39. https://doi.org/10.1016/j.psychres.2014.09.005

Yau, L. H. Y., & Little, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, *96*(456), 1232–1244. http://www.jstor.org/stable/3085886

Zivich, P. N., & Breskin, A. (2021). Machine learning for causal inference: On the use of cross-fit estimators. *Epidemiology*, *32*(3), 393–401. https://doi.org/10.1097/EDE.0000000000001332

Zugman, A., Winkler, A. M., & Pine, D. S. (2021). Recent advances in understanding neural correlates of anxiety disorders in children and adolescents. *Current Opinion in Psychiatry*, *34*(6), 617–623. https://doi.org/10.1097/YCO.0000000000000743

Zvolensky, M. J., Rogers, A. H., Mayorga, N. A., Shepherd, J. M., Bakhshaie, J., Garza, M., Viana, A. G., Ochoa-Perez, M., Lemaire, C., Ruiz, A., & Peraza, N. (2022). Perceived discrimination, experiential avoidance, and mental health among Hispanic adults in primary care. *Transcultural Psychiatry*, *59*(3), 337–348. https://doi.org/ 10.1177/13634615211038159