Retrieved-context theory of memory in emotional disorders

Rivka T. Cohen and Michael J. Kahana

University of Pennsylvania

Author Note

Rivka T. Cohen, Department of Psychology, University of Pennsylvania

Michael J. Kahana, Department of Psychology, University of Pennsylvania

This work was funded by the National Institute of Mental Health at the National Institutes of Health, grant number MH55687 to MJK. We thank Ayelet Ruscio, Sudeep Bhatia, Nora Herweg, Dan Schonhaut, Christoph Weidemann, Ethan Solomon, Paul Wanda, Nick Diamond, Noa Hertz, Alice Healy, Katherine Hurley, Jessie Pazdera, Effie Li, Nicole Miller, Anne Park, Dilara Berkay, and Michael Barnett for helpful conversations and comments on this draft. Portions of this work were presented at the Context and Episodic Memory Symposium in 2018 and the Society for Mathematical Psychology in 2019.

Abstract

Learning and memory play a central role in psychopathology, particularly in depression and posttraumatic stress disorder. We present a new, transdiagnostic theory of how memory and mood interact in emotional disorders. Drawing upon retrieved-context models of episodic memory, we propose that memories form associations with the contexts in which they are encoded, including emotional valence and arousal. Later, encountering contextual cues retrieves their associated memories, which in turn reactivate the context that was present during encoding. We first show how our retrieved-context model accounts for findings regarding the organization of emotional memories in list-learning experiments. We then show how this model predicts clinical phenomena, including persistent negative mood after chronic stressors, intrusive memories of painful events, and the efficacy of cognitive-behavioral therapies.

Keywords: memory, emotional disorders, depression, posttraumatic stress disorder, PTSD

Retrieved-context theory of memory in emotional disorders

Over 70% of adults will experience traumatic stress at some point in their lifetime, including threatened or actual physical assault, sexual violence, motor vehicle accidents, and natural disasters (Benjet et al., 2016). Following trauma exposure, 20-32% of adults suffer distressing, intrusive memories of the event, hyperarousal, avoidance of event reminders, and persistent negative mood (Brewin, Andrews, Rose, & Kirk, 1999; Koren, Arnon, & Klein, 1999). Trauma-exposed adults also commonly experience negative mood, decreased motivation, and disrupted sleep or appetite (Goenjian et al., 2001). Historically, the Diagnostic and Statistical Manual of Mental Disorders (DSM) has divided these symptoms into diagnostic categories of posttraumatic stress disorder (PTSD) or major depressive disorder (MDD), according to the prominence of memory versus mood disturbances, respectively (American Psychiatric Association, 2013). However, although memory intrusions are not among the diagnostic criteria for MDD, patients with depression also report high rates of intrusive memories, replete with sensory and reliving experiences (Reynolds & Brewin, 1999). Conversely, although PTSD is not considered to be a mood disorder, affective symptoms are part of the disorder and as many as 50-75% of patients have co-occurring MDD (Brady, Killeen, Brewerton, & Lucerini, 2000). The high comorbidity and symptom overlap between these disorders has led to a search for transdiagnostic processes (Insel et al., 2010; Sanislow et al., 2010) that hold promise for the development of more efficient and efficacious treatments. Our goal in this paper is to leverage computational modeling to develop a new, transdiagnostic theory of how distressing events interact with the memory system to produce the negative mood and intrusive memories that cut across depression and PTSD.

Learning and memory have long been implicated as a transdiagnostic process underlying psychopathology (Beck, Rush, Shaw, & Emery, 1979; Bower, 1981; Brewin, 2006). For example, patients with anxiety and depression exhibit more frequent recall of negative events,

suggesting processes of *mood-congruent recall* (Matt, Vázquez, & Campbell, 1992; Watkins, Mathews, Williamson, & Fuller, 1992). In clinical settings, therapeutic learning is enhanced when it occurs in an emotional context that will match the later retrieval context, suggesting processes of *emotion-state dependent recall* (Craske et al., 2008). Patients with PTSD experience intrusive memories of traumatic events (Ehlers, Hackmann, & Michael, 2004), and patients with other emotional disorders, including depression, social anxiety, agoraphobia, and other conditions, also report high rates of intrusive memories of painful events, which evoke distress and often involve visual/sensory and reliving experiences of the recalled event (Berntsen, 2010; Day, Holmes, & Hackmann, 2004; Hackmann, Clark, & McManus, 2000; Muse, Mcmanus, Hackmann, Williams, & Williams, 2010; Osman, Cooper, Hackmann, & Veale, 2004; Price, Veale, & Brewin, 2012; Reynolds & Brewin, 1999; Speckens, Hackmann, Ehlers, & Cuthbert, 2007).

Psychopathology research is increasingly drawing upon findings and methods from cognitive neuroscience to identify transdiagnostic processes that can explain these effects (Montague, Dolan, Friston, & Dayan, 2012; Sanislow et al., 2010). This has led to ongoing debate over whether intrusive memories and other memory disturbances in emotional disorders reflect the involvement of basic episodic memory processes (Rubin, Berntsen, & Bohni, 2008; Rubin, Dennis, & Beckham, 2011), a specialized memory system that handles memory for high-arousal negative events (Brewin, 2014), or other systems outside of episodic memory, such as fear-conditioning (Foa & Kozak, 1986) or semantic memory (Bower, 1981). We propose that one of the reasons this debate remains unresolved is that clinical theories have not yet integrated with recent theoretical developments in episodic memory, which concerns memory for personally experienced events that are associated with a specific context in time and place (Moscovitch, Cabeza, Winocur, & Nadel, 2016). A further reason for the persisting debate may be the difficulty of comparing narrative theories whose predictions under different conditions are not fully or formally specified. In memory research, computational models have been a highly

useful tool for clarifying whether a theory that can account for one phenomenon can simultaneously count for other phenomena of interest. In this paper, we draw upon insights from prior clinical theories of memory in emotional disorders (Brewin, 2014; Ehlers & Clark, 2000; Foa & Kozak, 1986; Rauch & Foa, 2006) and join them with computational modeling techniques of emotion in memory (Bower, 1981; Talmi et al., 2019) to develop a new theory of memory, mood, and their mutual influence in emotional disorders. We then test whether this new model, using our updated theory of emotion in episodic memory, can account for the persistent negative mood and intrusive memories observed in clinical settings.

Retrieved-context theory. In this paper, we propose a new theory of memory in psychopathology that builds upon a rich body of research into models of human episodic memory. In retrieved-context models, the episodic context that is associated with a memory during its encoding guides the later activation of that memory and determines whether it will be consciously retrieved (Healey & Kahana, 2014; Howard & Kahana, 2002a; Lohnas, Polyn, & Kahana, 2015; Polyn, Norman, & Kahana, 2009; Talmi et al., 2019). Each memory's perceptual features, such as sights and sounds, are encoded in a network of associations with the memory's context, such as time of day, physical location, and emotions felt during the event. Later, encountering a similar context cues retrieval of the associated memories in the stored network, activating the memories whose encoded contexts have the highest overlap with the context that is present at the time of retrieval. Once a memory is retrieved, it is reencoded in association with the new context into which it was re-introduced.

The Emotional Context Maintenance and Retrieval (eCMR) model. In a recent update to retrieved-context models, Talmi et al. (2019) broke new ground by conceptualizing emotion as a component of memories and their contexts, and by proposing that high arousal during encoding strengthens the associations between items and their contexts. The resulting model, eCMR, captures key features of emotional memory in laboratory settings (Talmi et al., 2019). However, eCMR has limited ability to describe memory in emotional disorders. First, eCMR treats emotion

in memory as a binary feature: that is, emotion is either present or absent, without having negative or positive valence. In addition, eCMR resets prior learning at the end of each encoding list. Thus, this model cannot distinguish between recalls from a target vs. non-target prior context, a necessary ability for modeling memory intrusions (Lohnas et al., 2015), which are of interest to accounting for intrusive trauma-memories.

Overview. This paper introduces a new retrieved-context model of emotion and memory (CMR3), that goes beyond previous work in allowing us to simulate a lifetime of memories and to examine the role of intrusions in the dynamics and persistence of affective states. First, we use a comparative modeling approach to determine the representation of emotional valence in the episodic memory system (Experiment 1, Simulation 1). Then, we demonstrate CMR3's ability to account for mood-congruent recall (Simulation 2) and emotion-state dependent recall (Simulation 3). We then test CMR3's ability to account for the effects of environmental stressors on negative mood (Simulation 4) and clarify the model's predictions that repeated stressors will have a greater likelihood of becoming activated as intrusive memories due to being associated with a wider variety of cueing contexts (Simulation 5). Then, we show the model's ability to capture the efficacy of behavioral activation therapy for depression (Simulation 6). We demonstrate CMR3's ability to predict that high arousal during encoding will lead to the development of intrusive memories (Simulation 7), and then show CMR3's ability to capture the efficacy of prolonged exposure therapy for addressing intrusive memories of high-arousal negative events (Simulation 8). We conclude with a discussion of how retrieved-context theory relates to current theories of memory in psychopathology and generates novel predictions.

A retrieved-context model of memory and mood

According to retrieved-context theory, when people form memories of new experiences, they encode the memory in association with its contextual information, such as the time and place of the experience, the thoughts and emotions present during the experience, and other

internal states. Later, these contexts – such as revisiting a location where an event took place, or having an emotion that was present during the event – can cue recall of associated memories. Once context cues a memory's retrieval, the memory reactivates its associated contexts, thus reinstating thoughts or emotions that were present during the original experience. The network of associations formed between items and past contexts can also encode semantic relationships among items, as items that share meaning tend to occur in similar temporal and semantic contexts (Howard & Kahana, 2002b; Polyn et al., 2009). Together, these *episodic* (contextual) and *semantic* (conceptual) associations guide memory retrieval.

Our model introduces two major advances beyond prior, related, models. First, our model enables memories and their contexts to have negative, positive, and neutral emotional properties (emotional valence). Second, our model enables emotional learning to accrue over a lifetime of experiences. This also allows the model to distinguish between recall of a memory from an intended (target) context versus an unintended (non-target) context, thus allowing the model to distinguish between voluntary and intrusive memories. The ability to make this distinction is critical in characterizing the role of memory in the development and treatment of PTSD.

Model Description

Here we provide a formal description of our model (Figure 1), cast in terms of equations that define the representation of items and the mechanisms that result in memory storage and retrieval. Following earlier formulations of retrieved-context theory, we assume a multidimensional feature representation of items, each denoted by f_t and a multidimensional representation of context that evolves as a result of each new experience and the memories it evokes (we denote the context at time t as c_t). Our approach inherits the same equations for updating context, associating context and items, and determining recall dynamics as Polyn et al (2009). We also inherit the additional mechanisms added by Lohnas et al (2015) to simulate the

continuity of memory across lists. We follow Talmi et al (2019) in modeling emotion as a component of each item vector which then integrates into context. Then, we advance the model by allowing emotion to be not only present or absent as in eCMR, but further, to have positive or negative valence. In addition, we allow learning to accumulate over the course of multiple lists. This has two desirable properties: first, it allows memory to accrue over the course of long periods of time as in CMR2 (Lohnas et al., 2015), rather than a single experimental list. Second, it provides a method by which we can operationalize intrusive memories as instances of recalling an item from a non-target list, a phenomenon that is of great interest in psychopathology but which cannot be captured by a model that resets the memory system at the end of each individual list. We call this updated model CMR3 and present a mathematical description below.

Item Representation. In CMR3, each element of $\boldsymbol{f}_t$ represents whether that item is present or absent in the current memory episode, and each element of $\boldsymbol{c}_t$ represents the extent to which prior memories or cognitive states are still active during encoding. In addition to the representations of item features in $\boldsymbol{f}_t$ and temporal context in $\boldsymbol{c}_t$, Polyn et al. (2009) introduce an additional subregion in each vector to contain source-memory attributes, such as the encoding task conducted during learning (Polyn et al., 2009) or the presence of emotion (Talmi et al., 2019). In eCMR, emotion is represented as simply present or absent, taking a binary value of 0 or 1 in the source-memory subregion. To model how memory differentially contributes to negative or positive mood, we updated the emotional subregion of both feature and context vectors to contain two cells. In $\boldsymbol{f}_t$, one cell indicates whether the item is negative, the second whether it is positive, and neutral items have no content in either cell. In $\boldsymbol{c}_t$, one cell holds the accumulation of negative context and the other holds the accumulation of positive context. We chose two separate cells due to findings that negative and positive emotion may operate via separate cognitive systems (Cacioppo & Berntson, 1994; Lang, 1995).

List Representation. As in CMR2 (Lohnas et al., 2015), we concatenated all presented lists in both the feature and context vectors, to allow the model to carry information forward that was learned in prior lists. Each list is separated in the feature layer by *disruptor items*. These disruptor items induce a shift in temporal context between lists, but they do not form associations with other items in the Hebbian matrices, as described below for true items. We begin each simulation by presenting an initial disruptor item so that the initial context vector is not empty (Lohnas et al., 2015). In addition, we present CMR3 with a disruptor item after each encoding list to model the distractor task used in Experiment 1.

Context-Updating. During encoding, as each item is presented, it enters the current context, causing other contextual elements to decay. Thus, context evolves according to the equation:

$$(1) \qquad \boldsymbol{c}_{t+1} = \rho \boldsymbol{c}_t + \beta \boldsymbol{c}^{IN}$$

As such, context becomes a recency-weighted sum of past context states. The features from $\boldsymbol{f}_t$ that will enter into the new context, represented by the term $\boldsymbol{c}^{IN}$, are determined by multiplying the weight matrix of item-to-context associations, which we call M^{FC}, by $\boldsymbol{f}_t$, the vector of current features, and then norming the resulting vector, such that $\boldsymbol{c}^{IN} = \frac{M^{FC}\boldsymbol{f}_t}{\|M^{FC}\boldsymbol{f}_t\|}$. In equation 1, ρ scales the magnitude of prior context so that $\boldsymbol{c}$ does not grow without bounds (see Howard & Kahana, 2002a). The β parameter determines the rate at which new temporal context enters the updated context vector. This rate differs depending on whether the subject is encoding a new event (β_{enc}), retrieving a new memory (β_{rec}), registering a shift between list contexts (β_{post}), or experiencing a context shift due to a distractor task ($\beta_{distract}$). Emotional context updates according to the same equation, at the rate of β_{emot}.

Encoding. As each newly presented item evolves and updates cognitive context, its features form associations with the elements of context present during encoding. The memory and context representations, $\boldsymbol{f}_t$ and $\boldsymbol{c}_t$, interact via two Hebbian associative (outer-product)

weight matrices, which model the strength of associations from the studied items to their encoding context, M^{FC}, and from context to associated items, M^{CF}. Because $\boldsymbol{f}_t$ and $\boldsymbol{c}_t$ each have two subregions – the first devoted to individual features of the memory, or *item features,* and its temporal context, and the second consisting of two cells devoted to emotional valence (see item representation), M^{FC} and M^{CF} have corresponding subregions. In the upper-left quadrant, each matrix contains associations between item features, $\boldsymbol{f}_{items}$, and temporal (or non-emotional) context elements, $\boldsymbol{c}_{temp}$. In the lower-left quadrant of M^{FC} and the upper-right quadrant of M^{CF}, each matrix contains the associations between item features and emotional context elements, $\boldsymbol{c}_{emot}$, as below:

$$(2) \qquad M^{CF} = \begin{bmatrix} \boldsymbol{f}_{items}\,\boldsymbol{c}_{temp}^{\top} & \boldsymbol{f}_{items}\boldsymbol{c}_{emot}^{\top} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

and

$$(3) \qquad M^{FC} = \begin{bmatrix} \boldsymbol{c}_{temp}\boldsymbol{f}_{items}^{\top} & \mathbf{0} \\ \boldsymbol{c}_{emot}\boldsymbol{f}_{items}^{\top} & \mathbf{0} \end{bmatrix}$$

Prior to the experiment, we initialize each weight matrix as an identity matrix of rank $i + 2$, where i is the total number of items presented in the experiment. Two is the number of elements contained in the emotional subregion of the feature and context vectors. In the process, the subregions containing associations between items' temporal features and emotional context are initialized to zero. The pre-experimental Hebbian matrices, called M_{Pre}^{CF} and M_{Pre}^{FC}, are then scaled by (1 – γ_{CF}) and M^{CF} by (1 – γ_{FC}), respectively. The parameters γ_{CF} and γ_{FC} are the learning rates for context-to-item and item-to-context associations, respectively (see *Learning Rates,* below).

Semantic associations. To represent semantic associations between pairs of items, we add a matrix of inter-item semantic associations, called M^{S}, to the quadrant in M_{Pre}^{CF} that contains associations between items' temporal features and temporal context (upper-left quadrant). Each entry in this matrix equals the dot product similarity between vector

representations of each item's features (Polyn et al., 2009). We define these similarities using Google's Word2Vec algorithm (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), which uses the co-occurrence of different words across multiple texts to determine vector representations of each word. The similarity between a pair of items can then be calculated using the dot product between each pair of item vectors. When adding M^S to the initialized M^{CF} , we scale M^S by the parameter *s*. Thus, *s* determines the degree to which semantic associations guide the retrieval process.

Learning Rates. New episodic associations accumulate as different memory features and contextual elements are active at the same time. We preserved the learning rule from CMR (Polyn et al., 2009), in which memories form associations with the context at the current time step, rather than at the preceding timestep. Thus, as each new memory is encoded, the change in associations stored in each weight matrix equals:

$$(4) \qquad \Delta M^{CF} = \boldsymbol{f}_t \boldsymbol{c}_t^\top$$

and

$$(5) \qquad \Delta M^{FC} = \boldsymbol{c}_t \boldsymbol{f}_t^\top$$

The rate at which these associations integrate into the M^{FC} and M^{CF}weight matrices is determined by three parameters: γ_{FC}, γ_{CF}, and γ_{emot}. The parameters γ_{CF} and γ_{FC} are the learning rates of new context-to-item and item-to-context associations, respectively. When updating both matrices, the γ_{emot} is the learning rate of new associations between item features and emotional context in M^{CF}, which allows memories to form new associations with emotional vs. non-emotional context at different rates. The learning rate in the other quadrants is set to 0.0, since these associations do not contribute to the encoding and retrieval processes in the

current model equations (Polyn et al., 2009; Talmi et al., 2019). Thus, before ΔM^{FC} and ΔM^{CF} integrate into M^{FC} and M^{CF}, we scale each ΔM elementwise by its respective matrix of learning rates, L^{FC} and L^{CF}:

$$(6) \qquad L^{CF} = \begin{bmatrix} \gamma_{CF} & \gamma_{emot} \\ 0 & 0 \end{bmatrix}$$

$$(7) \qquad L^{FC} = \begin{bmatrix} \gamma_{FC} & 0 \\ \gamma_{FC} & 0 \end{bmatrix}$$

In addition, to model increased attention to the items at the beginning of lists, each value in equation (4) is scaled by $\phi_t = \phi_S e^{-\phi_D(t-1)} + 1$, where ϕ_S scales the overall level of the primacy effect and ϕ_D determines the rate of decay in this primacy effect, as the ith item is presented. Thus, at a given point in the experiment, the strength of the associations between items and contexts stored in each weight matrix are given by the equations:

$$(8) \qquad M^{CF} = M^{CF}_{Pre} + L^{CF} \sum_t \phi_t \boldsymbol{f}_t \boldsymbol{c}_t^\top$$

$$(9) \qquad M^{FC} = M^{FC}_{Pre} + L^{FC} \sum_t \boldsymbol{c}_t \boldsymbol{f}_t^\top$$

Drawing upon work by Talmi et al. (2019), we additionally implement a parameter, ϕ_{emot}, which modulates the strength of context-to-item associations when an emotional item evokes arousal during encoding. For items that evoke no arousal, ϕ_{emot} takes a value of 1.0, and for items that evoke emotional arousal, ϕ_{emot} takes a value greater than 1.0 (Talmi et al., 2019).

Recall. Following the accumulation of feature-to-context associations over the course of multiple items and lists, the model proposes that the context present at retrieval serves to activate its associated memory elements. This takes place by multiplying M^{CF}, by the current context vector at retrieval, $\boldsymbol{c}_R$. In the resulting vector of item activations, each item is thus activated to the extent that its context during encoding is similar to the context that is present at the start of the retrieval process, such that the vector of item activations is given by:

$$(10) \qquad \boldsymbol{a} = M^{CF} \boldsymbol{c}_R.$$

Thus, each memory's features are activated according to the strength of the dot-product similarity between the context vector present during its encoding and the context that is present during retrieval, as well as the strength of pre-existing semantic associations. The activated elements then enter an evidence accumulation process, in which the evidence for the retrieval of any given item is driven by its activation value in $\boldsymbol{a}$. On each step of this leaky accumulator process (Usher & Mcclelland, 2001), the vector of evidence, $\boldsymbol{x}$, evolves according to the following equation:

$$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} - \tau\kappa\boldsymbol{x}_{n-1} - \lambda\tau N\boldsymbol{x}_{n-1} + \tau\boldsymbol{a} + \boldsymbol{\epsilon}. \tag{11}$$

That is, the level of evidence on the current step equals the level of evidence on the last step ($\boldsymbol{x}_{n-1}$), minus the decay in that evidence over time, minus lateral inhibition from other activated items (N is a matrix with 1's along the diagonal and -1's at all other entries, such that the summed activation of other items is subtracted from the given item's activation, and then scaled by λ and the time-decay scalar, τ), plus incoming item activations, and plus noise, $\boldsymbol{\epsilon}$. When the evidence in $\boldsymbol{x}_n$ for an item i passes a certain threshold, θ_i, that item emerges as a candidate for recall. For computational efficiency, we limit this evidence accumulation race to the 42 (list length x 2) items with the highest activations, since the other items are very unlikely to emerge as successful candidates for recall (Lohnas et al., 2015).

This item then updates the $\boldsymbol{f}_t$ vector with its features, and the item's associated context is then retrieved by multiplying M^{FC} by the newly updated $\boldsymbol{f}_t$ vector. If the recalled item's evoked context, c^{IN}, is sufficiently similar to the current context ($\boldsymbol{c}^{IN} \cdot \boldsymbol{c}_i > \boldsymbol{c}_{thresh}$, with $\boldsymbol{c}_{thresh}$ being a threshold scalar), then that item is reinstated into $\boldsymbol{f}_t$. Whether or not the retrieved item is reinstated, it cues its associated context, which evolves the context vector according to the same equations that update $\boldsymbol{c}_t$ during encoding. Once an item has been retrieved, the threshold of evidence needed to subsequently recall this just-retrieved item is equal to the threshold $\theta_i = 1 + \omega\alpha^j$, where j represents the number of times that a different item has been subsequently

retrieved. Here, $\alpha = [0.0, 1.0)$, such that ω raises the absolute level of the threshold, and α decays this newly raised threshold each time the given item is reactivated.

Experiment 1: The role of emotional valence in memory

To develop a model of how episodic memory processes can promote negative versus positive emotion in emotional disorders, it is crucial to understand how emotional valence is represented in human memory and, further, how it may guide and be evoked by the process of recalling a memory. In their retrieved context model of emotional memory, Talmi et al. (2019) modeled emotion as a binary source attribute whose value would indicate the presence or absence of emotional content in items and their contexts. In free recall tasks, however, subjects tend to recall negative, positive, and neutral items in clusters together with other items that have the same valence (Long et al., 2015). This *emotional clustering effect* suggests that not only the presence of emotion, but also the valence of that emotion, guides the organization and retrieval of items in memory. Here, we first sought to replicate the emotional clustering effect in an independent free recall dataset. Then, in Simulation 1, we fit the observed data using two variants of the retrieved context theory: one in which a single binary features indicates whether a memory possesses emotional information (Talmi et al., 2019) and another in which two binary attributes separately indicate whether a memory possesses positive or negative information (the current CMR3 model). We then assessed the two model variants by comparing their ability to fit emotional clustering effects, and other benchmark phenomena, in free recall.

Subjects and Stimuli. We fit the model to data from Experiment 4 of the Penn Electrophysiology of Encoding and Retrieval Study (Kahana, Aggarwal, & Phan, 2018), an ongoing project to collect a large set of free recall data within and across subjects. The full dataset consisted of 120 adults. Among these, 97 subjects completed the full 24 sessions of delayed free recall. The mean age among completers was 22.26 (SEM = 3.1), and reported genders were 51.5% female, 47.5% male, 1.0% other. In each session, subjects studied 24 lists

of 24 unique nouns (576 words total) drawn from a subset of 1638 nouns from the University of South Florida word pool (Nelson, McEvoy, & Schreiber, 2004). See Kahana et al. (2018) for full details of the experiment stimuli and methods. To assess the valence of each word, we used ratings from a prior norming study (Long et al., 2015). In this study, 120 subjects on Amazon's Mechanical Turk (MTurk; Mason & Suri, 2012) generated valence ratings for each item, rating them on a scale from 1-9, where 1 indicated negative valence and 9 indicated positive valence. Words were binned as negative if their valence rating fell below 4, neutral if between 4-6, and positive if above 6. The resulting valences of the 864 words in the experiment word pool were 25.6% positive, 8.9% negative, and 65.5% neutral.

Method. To test the effects of emotional valence on episodic memory, we presented subjects with a multi-trial delayed free recall task, as described above. In a free recall task, subjects study lists of items (here, words). At the end of each study list, subjects then recall as many words as they can, in any order. Because subjects must recall words from the most recent list, rather than all words they have ever learned in their lifetime, this task measures the ability to localize memory to a specific context (i.e., this specific list, in this specific session). Because subjects may recall the study words in any order (i.e., freely), the order in which these words are recalled suggests how the items were organized in memory. Finally, in a "delayed" free recall task, the experimenter places a distractor task (i.e., a delay) after each study list and just prior to the recall period. This reduces the recency effect, or better recall for the last 3-4 items on the list, thus preventing these items from being overrepresented in the subsequent analyses. Here, subjects completed a distractor task of 16 seconds of mental addition.

Behavioral Analyses. In this experiment, we sought to replicate the emotional clustering effect. This replication not only served to establish the reliability of the emotional clustering phenomena, but it also allowed us to obtain individual subject parameter sets for use in subsequent simulations of theoretically-generated data (see Simulations 2-8). We designed PEERS Experiment 4 to comprise sufficient data to allow for stable modeling at the individual

subject level (i.e., each subject contributed data from 576 lists obtained across 24 experimental sessions).

To evaluate the effect of emotional valence on recall organization we conducted a conditional probability analysis of transitions among items according to their positive, negative, or neutral valence (see Long et al., 2015 for full methods). Following Long et al. (2015), we adjusted probabilities to account for different rates of negative, positive, and neutral items in each study list. We then examined subjects' tendencies to transition between same-valent or differently-valent items, with higher probabilities for same-valent items indicating a higher tendency to cluster these items together during recall.

Next, we assessed the presence of classic, non-emotional patterns in free recall. These included the *serial position curve* (SPC), or how an item's probability of being recalled changes as a function of its presentation order during study (Murdock, 1962). The SPC allowed us to assess *primacy* and *recency* effects, or subjects' tendencies to recall the first and most recent items that they have seen, respectively. Second, we calculated the *probability of first recall* (PFR), or each item's probability of being first-recalled, plotted as a function of its serial position during study. This allowed us to assess what processes guide the start of subjects' recall outputs. Third, we used conditional probability analyses to calculate the tendency of items to cluster during recall to cluster according to their temporal proximity to one another during study (*temporal contiguity;* Kahana, 1996).

In addition, we used conditional probability analyses to measure subjects' tendency to cluster items during according to their semantic associations (*semantic clustering;* Howard & Kahana, 2002b). To represent semantic relationships among the words, we used Google's word vectorization method (Word2Vec; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Word2Vec identifies the semantic meaning of a word based on what other words accompany it across many different settings in which it is used. Using machine-learning, Word2Vec trains on a database of online texts and then generates vectors representing the dimensions of each

word based on the other words that typically accompany it. The inter-item similarity for each pair of words is then calculated as the dot-product between each word-vector in the pair. Using Word2Vec, we generated inter-item similarities for each pair of words in the 1638 noun word pool.

Finally, we calculated the average rate at which subjects mistakenly recalled items from lists preceding the target list (*prior-list intrusions,* or PLI's), and mistaken recalls of items that were never presented during the experiment (*extra-list intrusions,* or ELI's). For each behavioral analysis, we first averaged each effect across lists within session, and then across sessions within each subject.

Results. We assessed the patterns of recall that characterized the behavioral data across multiple sessions of delayed free recall (Figure 2). The lag-conditional response probability (Lag-CRP) analysis showed that subjects had a tendency to cluster their recalls according to how close in time those items were encoded relative to one another (temporal clustering). Subjects also showed a tendency to cluster their recalls according to how semantically similar those items were (semantic clustering). The serial position curve (SPC), or the probability of recall for each item based on where it was presented in the study list, and the probability of first recall (PFR), revealed both recency and primacy effects. That is, subjects were most likely to begin recall with the last item presented or the first item presented, and they were also more likely to recall these items in general. As expected based on previous research (Kahana, Dolan, Sauder, & Wingfield, 2005; Zaromb et al., 2006), our healthy young adult subjects committed low rates of PLI's and ELI's.

In addition, we replicated findings by Long et al. (2015), that subjects had a small, but significant tendency to recall items in clusters of the same emotional valence (Figure 3). That is, upon recalling a negative item, subjects were more likely to next recall another negative item than a positive item, and upon recalling a positive item, subjects were more likely to next recall another positive item than a negative item (Figure 3). To test the significance of this effect, we

conducted a 3 (Initial Item Valence: Negative, Positive, or Neutral) x 3 (Transition-Item Valence: Negative, Positive, or Neutral) linear mixed effects model predicting transition probabilities (lme4 R package; Bates, Mäechler, Bolker, & Walker, 2015). We set neutral valence as the base category, and we used a random intercept to account for the dependency of observations within each subject. The fixed-effects estimates and significance testing using the Satterthwaite's method for approximating degrees of freedom (lmerTest R package; Kuznetsova, Brockhoff, & Christensen, 2017) are presented in Table 1. The resulting interactions indicated that negative items were significantly more likely to be recalled together with other negative items than with non-negative items, $\beta = .057, t(768) = 35.1, p < 10^{-16}$, and positive items were significantly more likely to be recalled together with other positive items than with non-positive items, $\beta = .021, t(768) = 12.8, p < 10^{-16}$. Taken together, our results replicate findings that positive and negative items tend to cluster together with same-valent items during recall (Long et al., 2015). Next, we use a computational approach to identify which representation of emotion in memory is optimal to predict this effect.

Modeling emotional valence in memory

In retrieved-context models, emotional valence – the negative or positive emotional content of a memory – might serve one of several roles. As proposed in spreading activation models (Bower, 1981), emotional valence might serve as part of a memory's semantic content, contributing to how similar the features are of one memory to another (Bower, 1981). Alternatively, emotion might serve as a type of episodic context, yet have broad effects on memory, such that the valence of the emotion does not play a role (Talmi et al., 2019). Third, emotional valence might serve as a type of episodic context, such that negative or positive emotional states can become associated with an event and later serve as retrieval cues. To determine the optimal representation of emotion in memory, we fit CMR2, eCMR, and CMR3 to the behavioral effects observed in Experiment 1 and tested differences between each model's

ability to capture these patterns. This allowed us to compare three possible representations of emotion in memory encoding and retrieval processes. In CMR2 (Lohnas et al., 2015), we implemented emotion solely as part of items' semantic content: that is, as a component of inter-item semantic associations. In eCMR (Talmi et al., 2019), we additionally included an item's status as emotional versus neutral (i.e., the presence or absence of emotional content) as a part of item and context features. Because we were interested in capturing patterns of item intrusions across lists, and the original eCMR does not have multilist capabilities, we fitted eCMR with multilist capabilities as in CMR2 and CMR3. In CMR3 (see *Model Overview*), we additionally included emotional valence as part of item and context features.

Model Specification. To isolate the effects of emotional valence, we held the arousal parameter, ϕ_{emot} (Talmi et al., 2019), constant at 1.0 for each model. At this value, arousal neither enhances or weakens the context-to-item associations in M^{CF}. This allowed us to test each model's ability to capture emotional patterns in the behavioral data solely due to how it represents emotion's dimensions in the feature and context vectors. In addition, the words presented in the free recall tasks have emotional valence properties but likely induce minimal or absent levels of arousal during the free recall task (Long et al., 2015). As in the behavioral analyses in Experiment 1, we used Word2Vec (Mikolov et al., 2013) to generate inter-item similarities for each pair of items, which were added on to the pre-experimental M^{CF} matrix that represents associations between items and contexts.

Simulation 1: Method. We sought to test each model's ability to fit the variability in the dataset within each subject, as well as the aggregate behavioral patterns taken across the dataset. Therefore, we obtained parameter fits for each model to each individual subject's data. Obtaining parameter fits to individual subjects served two purposes. First, for each model, we used the resulting sets of parameters to generate three simulated datasets, comprised of the responses predicted by each model (CMR2, eCMR, or CMR3) for each subject. Then, on each simulated dataset of "virtual subjects," we repeated the full behavioral analyses that we ran on

the empirical data. This enabled us to test each model's ability to fit the aggregate behavioral responses observed in the original data. Second, obtaining the best-fitting parameters to each individual subject allowed us to examine the predictions of retrieved-context theory for how different people, whose minds operate according to different cognitive parameters, respond to life stressors with quick recovery, vs. negative mood or intrusive memories (Simulations 2-8).

To obtain the best-fitting parameters for CMR2, eCMR, and CMR3 for each subject, we used a particle swarm with 200 particles evolved over 30 iterations. In each iteration, the model generates the predicted outputs for subject when the model is outfitted with the given set of parameters ("particle"). Then the algorithm calculates a goodness-of-fit index for each particle that quantifies the error between the model predictions and the data, to identify which parameter set produced the best fit. For our fit index, we used the χ^2 error, defined as the sum of squared residuals, with each squared residual normed by the unbiased sample variance of the corresponding data point (Bevington & Robinson, 2003). After the χ^2 error for each particle is determined, the optimal parameter set is identified as the one having the lowest χ^2 error in this iteration, and other parameters move their values closer to the current optimal value. Over the course of iterations, the particles converge on an optimal set of parameters, thereby minimizing the χ^2 error. The 56 points contributing to the χ^2 error for each model fit were the first 15 and last 3 positions of the SPC; the first 15 positions of the PFR; the inner 6 points of the lag-CRP curve (lags of -3, -2, -1, 1, 2, and 3); the 9 conditional probabilities of transitioning among negative, positive, and neutral item types (negative-to-negative, negative-to-positive, negative-to-neutral, and so forth); all 6 points on the semantic CRP curve; the frequency of PLI's per list (1 data point); and the frequency of ELI's per list (1 data point).

Results and Discussion. In order to determine how best to represent emotional valence in CMR, we fit three models to the data from each individual subject. The first model, CMR2 (Lohnas et al., 2015), represented emotional valence solely as a contributor to items' semantic

associations with one another. In the second model, eCMR, an item's emotional content is represented as a binary value, being either present or absent in item features and episodic source context. Finally, in the new model we present in this paper, CMR3, we enabled items to have negative or positive valence. To obtain the best-fitting parameters for each model, fit to each subject, we minimized the χ^2 error between each model's predictions and the actual data for that subject. At the end of the particle swarm process, the mean χ^2 error across model fits to individual subjects was 36.60 (SEM = 4.95) for CMR2, 30.45 (SEM = 2.05) for eCMR, and 29.74 (SEM = 1.64) for CMR3, where lower χ^2 error values indicate better model fits. Degrees of freedom for the χ^2 error values were determined by *n – m,* where *n* equals the number of data points fit (56 points) and *m* equals the number of free parameters in each model (15 for CMR2 and 17 for eCMR and CMR3). Thus, for 97 subjects, CMR2 provided 86 model fits with non-significant $\chi^2(41)$ error; eCMR provided 92 model fits with non-significant $\chi^2(39)$ error; and CMR3 provided 90 model fits with non-significant $\chi^2(39)$ error. The means and standard errors of the sets of parameters obtained for CMR2, eCMR, and CMR3, taken across individual-subject fits, are presented in Table 2.

Next, we used these sets of parameters to simulate three full datasets, with each dataset indicating what the data would look like if subjects' cognitive operations followed the rules of the given model. Having generated the dataset predicted by each model, we assessed each model's ability to predict the aggregate behavioral effects that we observed in the actual data (Figures 2 and 3). Using each simulated dataset, we repeated the behavioral analyses conducted in Experiment 1, and compared the results to those obtained from the empirical data. The graphs for each behavioral analysis are presented in Figures 2 and 3. To assess each model's ability to fit the aggregate data, we calculated three measures of fit: (1) the same χ^2 goodness of fit index that was minimized while fitting the model; (2) the Bayesian Information Criterion (BIC) to account for the different number of parameters across models (Kahana, Zhou,

Geller, & Sekuler, 2007; Polyn et al., 2009; Schwarz, 1978); and (3) the RMSE, to identify which specific behavioral analyses determined each model's overall ability to fit the data. We calculated each goodness-of-fit test with respect to all data points in all analyses, *n* = 75, to obtain a total measure of each model's ability to fit the aggregate behavioral effects.

The resulting χ^2 error values were $\chi^2(60) = 74.1,\ p = .10$ for CMR2, $\chi^2(58) = 55.2,$ $p = .58$ for eCMR, and $\chi^2(58) = 56.7,\ p = .52$ for CMR3, indicating that all three model fits had non-significant error. For direct model comparisons, it is not valid to directly compare the size of χ^2 error values because eCMR and CMR3 have two more parameters than CMR2, which gives them an advantage over CMR2 in their ability to fit the data. Therefore, we calculated the Bayesian Information Criterion (BIC) values (Kahana, Zhou, Geller, & Sekuler, 2007; Polyn et al., 2009; Schwarz, 1978) for each model's fits to the aggregate data. The BIC accounts for each model's ability to fit the data while penalizing models that have a greater number of parameters, thus placing the three models on equal footing for comparison. Under the assumption of normally distributed residuals, the BIC formula simplifies to:

$$BIC = k\ ln(n) + n \ln\left(\frac{RSS}{n}\right)$$

Here, *k* represents the number of free parameters in the model, *n* represents the number of data points, and RSS is the residual sum of squares. To ensure that all points contribute equally to model fits, we multiplied the emotional clustering effect by a factor of 10 to place it on the same scale as the other data points from the set of behavioral analyses (Polyn et al., 2009). This was not necessary for the χ^2 error values because norming the squared residuals by the unbiased sample variance in the data (Bevington & Robinson, 2003) already sets the contributing residuals to comparable scales. The resulting BIC's were -347.06 for CMR2, -345.65 for eCMR, and -353.82 for CMR3, where lower (i.e., more-negative) values represent improved model fits. The results indicate that CMR3 provided the best balance of model parsimony and error minimization, followed by CMR2 and then eCMR.

To identify which behavioral effects distinguished each model's ability to fit the aggregate data, we calculated RMSE values for each behavioral analysis (Table 3). For the total RMSE, taken across all points in all graphs (N = 75), CMR3 provided the smallest RMSE, followed by CMR2, and then eCMR, where smaller values indicate better model fit. Comparing eCMR and CMR3, eCMR provided lower RMSE's for the positive lags of the Lag-CRP and the frequency of extra-list intrusions. Conversely, CMR3 provided the lowest RMSE for the emotional clustering effect, followed by CMR2 and then eCMR. CMR2 provided worse fits to the semantic clustering effect and the probability of first recall, suggesting that the model may have had to sacrifice fits to these data in its attempts to capture emotional clustering patterns.

Visual inspection of Figures 2 and 3 confirms that all three models provided comparable fits to non-emotional memory effects but differed in their ability to capture the emotional clustering effect, the key effect of interest in this dataset for developing a model of emotional memory. To clarify why CMR3 provided optimal fits to the emotional clustering effect, we assessed the tendency of each model to predict the two core aspects of this effect: (1) the tendency of same-valent items to cluster together during recall, and (2) the tendency of oppositely-valent items to not cluster together during recall.

Figure 3, Panel B, displays the tendencies of each model to over-predict or under-predict item clustering according to the emotional valences of just-recalled and next-recalled items. By not including emotion as a type of episodic features and contexts, CMR2 under-predicted the tendency of same-valent items to cluster together. However, by not differentiating between items of different valences, eCMR over-predicted the tendency of opposite-valent items to cluster together. In CMR3, allowing the emotional content of items and contexts to have both negative and positive emotional valence best captured the tendency of same-valent items to cluster together and the tendency of oppositely-valent items to not cluster together.

In sum, prior theorists have proposed that emotion serves as a feature of items' representations and their contexts (Talmi et al., 2019), or alternatively, that it serves as a type of

semantic information about an item (Bower, 1981). However, neither theory has yet been tested by quantitative model fits to empirical data. In this paper, we conducted explicit model comparisons between CMR2 (emotion as semantic-information only), eCMR (representing emotion as present or absent in item features and their contexts, but not having multiple valences), and CMR3 (representing emotion in item features and their contexts, and allowing it to have positive or negative valence). Our results demonstrated that emotion serves as an element of items' episodic features and contexts, as proposed by Talmi et al. (2019) but not tested via quantitative fits in that paper.

Moreover, we show that to capture the effects of emotion on free recall, it is important to represent emotion as having positive and negative valence. First, including emotional valence improved the empirical fits presented above. Whereas CMR3 and eCMR yielded equivalent fits for the χ^2 fit index, the RMSE indices indicated that this was due to eCMR having better performance on non-emotional patterns in the data and CMR3 having better performance on emotional patterns in the data. Overall, CMR3 yielded an improvement in fit according to its BIC and total RMSE value. Second, including emotional valence in the model is necessary to allow retrieved-context theory to predict the activation of negative vs. positive emotional context when retrieving a given memory. In the subsequent sections, we show how the new retrieved-context theory, developed above and formalized in CMR3, can capture patterns of negative vs. positive mood and memory in emotional disorders.

Mood-congruent recall

In the prior section of this paper, we used comparative model fits to test how best to represent emotion in episodic memory. Next, we present a series of simulations testing the model's ability to predict patterns of memory and emotion that are relevant to psychopathology. We begin by evaluating the model's ability to account for mood-congruent and emotion-state dependent recall. Then, we present a new theory of how these principles contribute to negative

mood and intrusive memories as transdiagnostic processes, with a focus on patterns in major depression and PTSD.

Motivation. A patient in a depressive episode may readily recall memories of loss or failure, while memories of happy events elude him (Teasdale, 1983). Yet, once recovered, this patient may tend to recall positive memories and have difficulty remembering the events that occurred while they were depressed. This phenomenon, *mood-congruent recall* (MCR), refers to higher rates of recall for memories whose emotional attributes match those of the person's current emotional state. MCR has been observed widely across both clinical and healthy populations: e.g., for sadness and loss-related stimuli in patients with depression (Bradley, Mogg, & Williams, 1995; Teasdale & Fogarty, 1979; P. Watkins et al., 1992), for threat-related stimuli in patients with anxiety (Mathews, Mogg, May, & Eysenck, 1989; but see Bradley et al., 1995) and patients with PTSD (Paunovic, Lundh, & Öst, 2002), and for positive stimuli in healthy subjects (P. Watkins et al., 1992). Here, we sought to test whether retrieved-context theory can account for the process of mood-congruent recall.

Simulation 2. Method. We tested CMR3's ability to predict mood-congruent recall (MCR). Specifically, we simulated the results from a series of free recall experiments in which MCR would be expected to occur. First, we simulated 97 virtual subjects by specifying the model with each set of parameters generated while fitting CMR3 to the behavioral data in Experiment 1. Each set of parameters defines the virtual subject's memory system, determining the responses that the model (i.e., the virtual subject) will give while completing the simulated task. Then, we presented each virtual subject with the following task. The simulated task consisted of 60 encoding lists, with each containing a randomly permuted mixture of positive (30%), negative (30%), and neutral (40%) items.

To ensure that negative, positive, and neutral items would all have equivalent semantic structures, we drew the inter-item similarity for each pair of items at random and without replacement from a normal distribution ($\mu = 0.2, \sigma = 0.5$). A value of -1.0 represents maximum

dissimilarity, and 1.0 represents maximum similarity. To simulate a set of unique items, no item pairs were assigned an inter-item similarity of 1.0. Finally, we set the *s* parameter equal to 0.15, which scaled down the inter-item semantic similarities by that value. This allowed the model outputs to more strongly reflect the episodic dynamics of interest.

In this task, each virtual subject began studying each list in a neutral state. That is, we set the virtual subject's context vector to contain no emotional content. After each encoding list, and prior to the free recall section, we simulated a mood induction. That is, after each list, we set the subject's context vector to contain negative (20 lists), neutral (20 lists), or positive (20 lists) emotional content. The model then predicted the subject's responses during a free recall session, in which emotional context was allowed to evolve naturally in response to the emotional content of the retrieved items. We then ran this simulation for each of the 97 sets of parameters obtained in the fits to behavioral data in Simulation 1. Across simulations, we analyzed the probability of recall for positive, negative, and neutral stimuli, conditional upon the type of mood induction prior to recall.

Results and Discussion. Using the sets of parameters obtained in Simulation 1, we simulated a set of 97 virtual subjects by specifying CMR3 with each set of parameters and obtaining the model-predicted output for a virtual recall task. Specifically, we assessed the model-predicted effects of mood state during retrieval on the recall of emotionally valent and neutral items. Across parameter sets, model predicted that negative or positive mood states cue more-frequent recall of items that match the current emotional state (Figure 4, A). Further, the model predicted that, during a neutral mood, emotional items were more likely to be recalled overall, with probabilities of 0.66 (SEM = .021) for negative items, 0.65 (SEM = .023) for positive items, and 0.61 (SEM = .024) for neutral items. This matches findings that in laboratory tasks lacking mood induction manipulations, and in which most subjects likely have neutral mood during the task, emotional items have a recall advantage over neutral items (Long, Danoff, & Kahana, 2015; Siddiqui & Unsworth, 2011; Talmi et al., 2019). In CMR3, this occurs because

studying or recalling an item evokes emotional context consistent with its emotional valence, and these emotional features serve as the input to the context updating equation. As a result, items that induce a particular emotional context during their encoding have a heightened match when the emotional context evoked by the pre-recall mood induction has matching emotional elements. Next, we examine the model's ability to predict a related effect, emotion-state dependent recall.

Emotion-state dependent recall

Motivation. Whereas patients learn new skills and insights in a calm, therapeutic setting, they must often recall these skills for use in settings fraught with intense emotion. *Emotion-state dependent recall* refers to better recall of learned items or associations when they are retrieved in a context similar to that in which they were encoded. Emotion-state-dependent recall is a variant of the more general *state-dependent recall* effect, in which information learned in one environment or state is better recalled in a matching state (Godden & Baddeley, 1975; Smith & Vela, 2001). Whereas laboratory studies of emotion-state dependent recall have yielded mixed findings (Bower & Mayer, 1985; Eich, 1995), the dependency of new learning on emotional context is routinely observed in clinical settings (Craske et al., 2008). In a prominent example, patients with anxiety who form new associations between feared stimuli and neutral outcomes often find that when encountering these stimuli in new contexts, the associated anxiety reemerges. Here, we tested whether CMR3 can account for emotion-state dependent recall.

Simulation 3. Method. We simulated a delayed free recall task consisting of 60 lists of 10 neutral items. We simulated mood inductions (1) at the beginning of each encoding list and (2) immediately prior to recall. The order of mood inductions in each session was randomized, such that 30 lists had matching encoding and retrieval contexts (15 negative-negative, 15 positive-positive) and 30 lists had mis-matched encoding and retrieval contexts (15 negative-positive, 15 positive-negative). We then calculated the model-predicted recall probabilities and response

times conditional upon matching negative, matching positive, mismatched negative-positive, and mismatched positive-negative, encoding and retrieval contexts. Finally, we ran this simulation with each set of parameters obtained by fitting CMR3 to individual subjects' behavioral data in Simulation 1.

Results and Discussion. In this simulation, we sought to determine whether CMR3 can predict *emotion-state dependent recall,* or improved recall of new learning that occurred in an emotional state that matches that present during encoding. Specifically, we tested whether virtual subjects recalled neutral items better when their emotional state during retrieval matched the emotional state that was present during encoding. The model-predicted recall probabilities for neutral items encoded and retrieved in matching emotional states were .67 (SEM = .027) for negative states and .69 (SEM = .025) for positive states. The recall probability for neutral items encoded in a negative state and retrieved in a positive state was 0.41 (SEM = .033), and the recall probability for neutral items encoded in a positive state and retrieved in a negative state was 0.38 (SEM = .035). This decrease in recall for neutral items in non-matching (vs. matching) emotional states during encoding and retrieval reflects the model's core prediction: an item's activation for retrieval depends on the similarity between the current context and the context associated with that item during encoding. In CMR3, emotional content is part of those contexts. Thus, the model can account for findings that new learning in clinical settings is enhanced when patients learn new material or associations in an emotional state that matches the emotional state in which they will later need to activate the new learning (Craske et al., 2008).

Effects of frequent, unique stressors on mood

A core feature of major depressive disorder (MDD) is negative mood that persists for two weeks or longer (American Psychiatric Association, 2013). Stressful life events are a major precipitant for MDD, with the likelihood of a depressive episode increasing according to the frequency of recent stressors (Kendler, Karkowski, & Prescott, 1998). Even types of depression

considered to be more biologically based are often precipitated by a severe event or major difficulty, particularly for the first major depressive episode (Brown, Harris, & Hepworth, 1994; Johnson & Roberts, 1995). Early developmental stressors, such as childhood abuse and neglect, also contribute to later risk of depression in adulthood (Jaffee et al., 2002; Kendler et al., 1998, 1995), perhaps in part by influencing how individuals recall their experiences and respond to new events. Here, we test the ability of retrieved-context theory to characterize how the frequency of prior and recent stressors can lead to the development and persistence of negative mood.

Simulation 4. To model the effects of stressors on persistent mood, we again modeled a group of virtual subjects, for whom we then simulated a series of lifetime events. As in Simulations 2 and 3, we constructed each virtual subject by using the CMR3 model, specified by the best-fitting parameters obtained when fitting CMR3 to that subject's individual recall dynamics from the empirical data (see Simulation 1). The model equations thus defined how each virtual subject would respond to new events and recall old ones. As a brief review, in CMR3, each new event introduces its emotional features into the current context and then becomes associated with this temporal, emotional, and spatial context during encoding. Later, encountering a retrieval context that shares high perceptual or semantic similarity to prior contexts reactivates associated memories. When the activated memories are retrieved, they reinstate the context that was present during their encoding. The memory is then re-encoded together with its old context and the new context into which it was retrieved. This updated context serves as the retrieval cue for the next memory.

Using these operations, we modeled how each of the 97 virtual subjects would respond to three different conditions of life events. Unlike in the prior simulations, here an event does not represent a word or other item studied in a list during an experimental recall task. Rather, an event in this simulation models an event that might happen to a subject in the real world. We assigned different properties to each event, including the type of emotional valence and the

event's semantic relatedness to other events. For simplicity, we allowed each event to have only one valence: either negative, positive, or neutral. To simulate event properties and semantic associations, we drew values for inter-event similarities at random and without replacement from a normal distribution ($\mu = 0.2, \sigma = 0.5$), with positive values indicating similarity and negative values indicating dissimilarity. Although all events in this simulation are abstract, to provide a concrete example, a birthday party and an engagement party might have a positive similarity of 0.8. In comparison, going swimming and filling out a tax form might be more dissimilar, with a score of -0.2. No event pairs were allowed to have a perfect similarity of 1.0, to ensure that all simulated events would be unique. In addition, we set the parameter that controls how much semantic associations guide the retrieval process, *s,* to a value of 0.15. Drawing the inter-event similarity values for all events from the same distribution, regardless of their emotional content, and holding *s* constant across all virtual subjects allowed us to control whether the results of our simulations could be accounted for just due to the semantic associations among events.

For each virtual subject, we simulated a life-course of events, which began with a "developmental period" simulating the initial events experienced by each virtual subject. Each developmental period consisted of 50 sets, or "lists," of 10 encoding events. Then, we varied the contents of each virtual subject's developmental period across three conditions. Each condition varied the frequency of stressors, and in the process, the ratio of negative to positive events. In the Low-Frequency condition, lists contained 40% neutral, 30% negative, and 30% positive events, resulting in a 1:1 ratio of negative-to-positive events. In the Moderate-Frequency condition, we increased the frequency of negative encoding events, such that lists contained 20% neutral, 60% negative, and 20% positive events, resulting in a 3:1 ratio of negative to positive events. In the High-Frequency condition, we further increased the frequency of negative events, such that lists consisted of 20% neutral, 70% negative, and 10% positive events, resulting in a 7:1 ratio of negative to positive events. Each "list" (i.e., sequence) of encoding

events was followed by a free recall period, to model instances in which virtual subjects might spontaneously recollect these prior events. After the developmental period, we presented a *neutral period.* This neutral period consisted of 30 lists of 10 new, unique encoding events, 100% of which had neutral valence. This allowed us to isolate the effects of prior learning during the developmental period on later mood and memory retrieval, without needing to filter out the effects of new emotional events.

Method. Because each virtual subject is defined as a set of distinct parameters, this allowed us to run each condition for each virtual subject. As we presented the simulated encoding events to each virtual subject, we tracked the model-predicted levels of negative and positive emotion evoked by each event. In addition, during the session of free recall after each list, we calculated the proportion of correctly recalled events (i.e., events that had occurred in the prior, target encoding list), the number of response repetitions (i.e., events that were correctly recalled more than once within each simulation period), and the rate of memory intrusions (i.e., events that were recalled but that had occurred in a non-target encoding list), and the number of event repetitions that occurred within each simulation period. All recall metrics were averaged across lists within each period in the simulation.

Results and Discussion. In this simulation, we sought to test CMR3's predictions of how developmental learning histories affect subsequent mood and memory retrieval. In Figure 5, we first illustrate the dynamics of emotion in each "experimental" condition for one virtual subject, with the subject's cognitive operations defined by the parameters estimated by fitting CMR3 to the recall dynamics observed in a free recall dataset (see Simulation 1). The three sample developmental paths shown in Figure 5 illustrate how the same set of parameters responds to three different developmental learning histories to determine the subject's mood state in response to encoding and retrieval events. Each panel depicts the model-predicted mood for the same parameter set over the course of encoding and retrieval events for each of three conditions: low, medium, and high frequencies of stressors during the developmental period.

Within each panel, the graph shows model-predicted mood over the course of four simulated periods of time. The first two periods, the developmental and neutral periods, show the output of the simulations conducted in this section (*Effects of unique, frequent stressors on mood*). In the developmental period (Figure 5, *developmental period*) we show the model-predicted mood while encountering low, medium, or high frequencies of stressors during early learning. In the neutral period (Figure 5, *neutral period*) we show the model-predicted mood once these stressors have abated and the model is encountering only new, neutral events. The latter two periods of each panel, the treatment and post-treatment periods, show the virtual subject's response to a simulated course of behavioral activation therapy, which we discuss in *Behavioral Activation Therapy* (Simulation 6).

In each panel, height on the graph represents the level of model-predicted mood, calculated as the difference between the amount of positive emotion and negative emotion contained in the emotional valence subregion of the context vector at that point in the simulation. The color of each point on the graph represents the type of event, whether an encoding event or a retrieval event, that evoked the current mood state. Blue represents a negative event, pink a positive event, and gray a neutral event. Each simulation run has the same number of encoding events; however, because CMR3 models recall as a noisy process, the number of recalled events differs in each simulation, and as a result, the length of each simulation differs across panels.

Mood. In Figure 6, we present the mean level of mood, averaged across virtual subjects, within each period of the low-frequency stressors condition, the medium-frequency stressors condition, and the high-frequency stressors condition. As the frequency of developmental stressors increased, the model predicted an increase in mean levels of negative emotion and a decrease in mean levels of positive emotion during the developmental period (Figure 6, Panel A). Further, the emotional context that accrued during the developmental period persisted into the neutral period, even though the virtual subjects rarely recalled events from the

developmental period (Figure 6, Panel B). Instead of mood persisting due to recall of prior negative experiences, recalling the recent neutral events evoked the emotional context with which they had been encoded, which reflected the levels of negative or positive emotion evoked and accrued during the developmental period. That is, the negative emotion from recent stressors generalized to neutral events, causing them to reactivate negative mood even once all stressors, and active memories thereof, had abated.

Recall. In Figure 6, we present the mean levels of types of recall outputs within each period (Panels B-D), averaged across virtual subjects. During the neutral period, in which the virtual subjects experienced persistent low mood due to the prior stressors, the model predicted decreased probability of correct recall relative to the prior developmental period (Figure 6, Panel B). However, when mood elevated during the treatment period of the simulation introduced in a subsequent section (Figure 6, Panel A; see methods in *Behavioral Activation Therapy*, Simulation 6), the probability of correct recall also increased (Figure 6, Panel B).

One theory of reduced recall of recent neutral or positive events during a depressive episode is impairment of the neural structures that support the memory system (Austin, Mitchell, & Goodwin, 2001; Bremner et al., 2004). However, our simulations demonstrate reduced correct recall rates of recent events during a state of persistent negative emotion (Figure 6, Panel B) while using the same set of parameters for each virtual subject, throughout all periods of each simulation. Our results propose an alternative, or perhaps complementary, explanation for reduced memory during a depressive episode. Specifically, mood-congruent recall processes and interference from prior negative memories, which are activated by the patient's mood state, should decrease recall for any non-mood congruent memories, including neutral or positive stimuli. Considering that the stimuli in cognitive tests are often designed to be neutral, our simulations suggest the importance of future research to test whether findings of reduced memory in depressed patients may reflect the mismatch between items' emotional properties

and the subject's emotional state, rather than (or perhaps as a compounding factor to) damage to the memory system.

Repeated stressors

In human memory, repetition is a powerful force. Repeated events have higher rates of recall, particularly when the events are spaced out over time (Madigan, 1969; Melton, 1970). Repeated exposure to severe, traumatic stressors is associated with depressive symptoms (Herman, 1992) and more varied and severe PTSD symptoms (Briere et al., 2008; Cloitre et al., 2009). Even mild stressors, when repeated, can lead to depression-like effects in animal models, including learned helplessness and decreased pleasure-seeking behaviors (Cabib & Puglisi-Allegra, 1996; Willner, 1997). In human adults, daily hassles are associated with the persistence of clinically significant distress (Depue & Monroe, 1986). Here, we used CMR3 to test novel predictions about how spaced repetitions of mild stressors might perpetuate negative mood by amplifying the activation of memories for repeated stressors.

Simulation 5. Method. We repeated the methods in Simulation 4, including the emotional composition of lists in the developmental and neutral periods in three separate conditions: Low-, Moderate-, and High-Frequency stressors (see Simulation 4 for full methods). However, in this simulation, we selected one negative event to repeat across lists, at a frequency of every other list. As in Simulation 4, we followed each developmental period with a period of 100% neutral events. The absence of any new emotional learning during this period allowed us to isolate how the prior, repeated stressor affects subsequent mood and recall without confounds due to new emotional events.

Results. In this simulation, we sought to determine the predictions of retrieved-context theory for the effects of stressor repetition on emotion and memory outcomes. In particular, we tested the effects of repeating one stressor during the developmental period on mood and correct recall during the later neutral period, when the virtual subjects were encountering only

new, neutral events. During the neutral period, virtual subjects recalled the memory of the repeated-stressor, even though neither the repeated-stressor nor other negative events had been presented during the neutral period (Figure 7). In comparison, no stressors had intruded into the neutral period during the *Frequent Stressors* simulation (Simulation 4), in which all negative events were unique, and in which arousal was held constant for all events (for the role of high arousal during encoding on intrusive memories, see *Intrusive Memories,* Simulation 7). Thus, spaced repetitions predisposed the repeated-stressor event to intrude into memory in later time periods of the simulation.

Upon analyzing the patterns of intrusion, two groups emerged. During the neutral period, one group exhibited stressor-memory intrusions at a frequency greater than one intrusion per list (High-Intrusions Group, n = 12). The second group exhibited stressor-memory intrusions at a frequency near zero (Low-Intrusions Group, N = 85). The latter group exhibited a more favorable mood (Figure 7, Panel A) as well as greater responsiveness to the subsequent simulation of behavioral activation therapy (see Simulation 6: Behavioral Activation Therapy). That is, when the repeated-stressor intruded into memory during the neutral period, it evoked negative mood. Further, the degree of negative emotion that it evoked increased as the frequency of stressors in the surrounding context increased during the developmental period. This is evident in that the rate of stressor-memory intrusions was constant across conditions (Figure 7, Panel D), yet within both groups, virtual subjects still exhibited greater or lower levels of negative mood depending on which condition they were in, or which composition of "background events" had been present when the repeating stressor was encoded. Virtual subjects exposed to high, vs. moderate or low, rates of contextual stressor events had lower mood when the memory of the repeated stressor intruded (Figure 7, A). Thus, the stressor memory evoked greater negative emotion during intrusions when it had been encoded against multiple backdrops of other negative events.

Figure 7, Panel B, demonstrates that virtual subjects demonstrated similar patterns of correct recall regardless of High-Intrusions vs. Low-Intrusions group, although virtual subjects in the High-Intrusions group exhibited better probability of correct recall in the neutral period if their developmental period contained a low frequency of contextual stressors. In Figure 7, Panel C, the higher rate of repetitions for the High-Intrusions group during the developmental period reflects repeated memories of the repeated-stressor event. This is because during the developmental period, in which the repeated stressor was actually occurring, it is considered a correct recall.

Overall, CMR3 predicts that repeating a stressor has two effects. First, repetition associates the stressor with a wider variety of contexts, thus increasing the potential retrieval cues that can prompt recalling the stressor. This raises the probability of recalling the stressor and primes it for activation in contexts where it does not belong (i.e., as a memory intrusion). Second, the emotion evoked by the recall of a stressful event reflects not only the details of that event, but further, of the emotional backdrop of other contextual events that were occurring while the memory was encoded. That is, our simulations predicted that the ratio of negative to positive events in the surrounding context during the encoding of a repeated-stressor will influence the subsequent mood that it later evokes. If the repeated stressor is encoded over multiple negative contexts, it will evoke high levels of negative emotion when recalled; but if it is encoded over multiple neutral or positive contexts, then these other contexts will become associated with the memory of the event. As a result, even though the events' own features are negative, the model predicts that associations with more positive emotional contexts will reduce the degree of negative emotion evoked when the stressor-memory is recalled. Thus, retrieved-context theory can begin to provide a quantitative rationale for how factors like social support can contribute to reduced distress and resilience to psychopathology in the aftermath of a stressful event (Brewin, Andrews, & Valentine, 2000).

Behavioral Activation Therapy

In clinical psychology, a theory of psychopathology must account not only for risk and maintaining factors, but also for what interventions reduce symptoms and by what mechanisms. Thus, we sought to determine whether CMR3 could predict the efficacy of Behavioral Activation Therapy (BAT), which is an efficacious treatment for depression based on learning principles. In BAT, patients increase the rate of positive or reinforcing events that they experience (Beck et al., 1979; Jacobson et al., 1996; Lewinsohn, Sullivan, & Grosscap, 1980). This is theorized to help correct learned associations between neutral stimuli and punishing outcomes (Kanter et al., 2010) and to help re-engage with opportunities for positive experiences (Jacobson, Martell, & Dimidjian, 2001). Here, we tested the ability of retrieved-context theory to account for the efficacy of BAT by first indirectly manipulating mood to increasingly negative mood states, following the procedures in Simulation 4 (*Frequent, unique stressors)* and in Simulation 5 (*Repeated stressors)*. After the simulation resulted in neutral, negative, or very negative mood states across virtual subjects, we then simulated a course of BAT and examined the model-predicted effects on mood, both during the treatment period and during a recovery period afterward. Simulating a course of BAT as a supplemental simulation to both Simulation 4 and Simulation 5 allowed us to examine the model's predictions for the conditions under which BAT is efficacious: specifically, for its efficacy under conditions in which negative mood has generalized to neutral events (Simulation 4) and in which negative mood is being perpetuated by intrusive memories of negative events (Simulation 5).

Simulation 6. Method. We extended Simulation 4 (Persistent Negative Emotion) and Simulation 5 (Repeated Stressors) by adding two additional periods in the simulation. After each virtual subject experienced the given developmental period and acquired a persistent mood state during the neutral period, we introduced a *treatment period*. During the treatment period, we presented the model with 20 lists of BAT. We operationalized BAT as a series of 20 lists, each with a list length of 10 events, that contained an increased, 3:1 ratio of positive to negative

events (list composition: 30% positive, 10% negative, 60% neutral). After each treatment period, we presented each model with a series of 60 lists of a *recovery period*, in which the lists of encoding events were largely neutral with a small, balanced proportion of positive and negative events (80% neutral, 10% negative, 10% positive). This enabled us to assess whether the model could predict a sustained response to treatment, as well as to test the model's predictions regarding the conditions leading to relapse. As in the prior simulations, we tracked the model-predicted mood and recall patterns during the treatment and recovery periods.

Results. In this simulation, we sought to test CMR3's ability to predict efficacy of behavioral activation therapy for alleviating negative mood in major depressive disorder (Lewinsohn et al., 1980). During BAT, the model predicted improved mood for all patients in Simulation 4, in which all stressors were unique, for all frequencies of negative developmental events (Figure 6, Panel A). In the recovery period after the BAT period, all patients in Simulation 4 sustained recovery, which we operationalized as having an average mood (Positive – Negative context values) at or above 0.0 (as opposed to negative values). We considered an average mood level of 0.0 given that the recovery period consisted mainly of neutral events (80% of events), with a small and equal mixture of positive (10%) and negative (10%) events. CMR3 predicted that BAT would lift mood and that, in the absence of further high-frequencies of stressors, or of new repeated stressors, this mood improvement would be maintained over time (Figure 6, Panel A). In the CMR3 framework, this occurs because positive events evoke positive emotional context, which begins to alleviate the negative context lingering from prior stressors. Over time, new events bind to a progressively less-negative, and eventually positive, emotional context. New events thereby reactivate more positive emotional contexts when they are later recalled.

By contrast, in the *Repeated Stressors* simulation (Simulation 5), CMR3 predicted that BAT did produce sustained mood improvements in virtual subjects who continued to experience high rates of stressor-memory intrusions of the repeated event. Although the high rates of

positive events during the treatment period temporarily lifted mood (Figure 7, Panel A), intrusive memories of negative events continued to evoke negative context. As a result, subjects experienced blunted emotional responses to positive events in the long-term. Thus, retrieved-context theory predicts that repeated stressors can lead to intrusive memories and consistent activation of associated negative emotion that blunt positive emotional responding, and thus interfere with BAT for the treatment of depression if the intrusive memories are not themselves addressed. Treatments that directly target intrusive memories in depression have shown promising results (Brewin et al., 2009), as have treatments targeting perseveration about distressing memories in the context of rumination (Watkins et al., 2011). Our findings in this simulation provide independent corroboration that such approaches might facilitate the efficacy of cognitive-behavioral therapy when distressing memory-intrusions are prominent. Next, we examine the model's ability to characterize the mechanisms of intrusive memories of high-arousal negative stressors.

Intrusive Memories

Motivation. In the aftermath of extreme or traumatic stress, patients with PTSD report that memory intrusions constitute their most distressing symptom (Ehlers & Steil, 1995). For intrusive memories in PTSD, stimuli that share strong perceptual similarity to those that directly preceded the event serve as a strong cue of memory intrusions (Ehlers et al., 2004). Such stimuli, here called "trauma cues," often share little or no semantic similarity to the event itself (Ehlers & Clark, 2000). For example, a patient who was assaulted near a water fountain may experience flashbacks when hearing the sound of running water, even though there is no semantic association between water and assault.

Although the tendency of perceptual cues to activate trauma-memory intrusions has been attributed to strong stimulus-response associations due to fear conditioning (Foa & Kozak, 1986), patients with intrusive memories in PTSD often report emotions other than fear (e.g.,

sadness, anger, or shame) as their primary emotional response to the event. Further, there is a need to develop a theory of how such memory intrusions can occur so widely across disorders and stressor types, given that intrusive memories of painful events are found in numerous disorders beyond PTSD (Brewin et al., 2010; Hackmann et al., 2000; Muse et al., 2010) and often involve stressors not traditionally considered to be traumatic, such as social rejection or perceived failure (Williams & Moulds, 2007).

Here, we examined whether CMR3 can predict intrusive recall of a high-arousal stressful event as the result of encounters with associated trauma cues. For the sake of brevity, we will refer to a high-arousal stressful event as a “trauma event” in this simulation. However, note that this simulation does not distinguish between trauma vs. non-trauma events. Rather, this simulation modes the effects of high-arousal during encoding as a general principle of the episodic memory system, affecting memories quantitatively (that is, having the same effect for all memories, but with this effect amplified according to how much arousal is present during encoding) rather than qualitatively (that is, affecting only memories of a specific “trauma category”). Retrieved-context theory thus predicts that any high-arousal, negative event has the potential to lead to subsequent intrusive memories, with a higher probability of this outcome as the level of negative arousal during encoding increases. In addition, we sought to identify whether CMR3 could predict that approaching, rather than avoiding, trauma cues leads to a reduction in intrusive memories – the principle that forms the basis for exposure therapy for PTSD (Foa & Kozak, 1986; Powers, Halpern, Ferenschak, Gillihan, & Foa, 2010).

Simulation 7. Method. This simulation follows the same format as Simulation 4. First, we presented virtual subjects with a developmental period of 50 lists of 10 encoding events, with all developmental periods containing 40% neutral, 30% positive, and 30% negative events. During this developmental period, we selected one negative event at random to serve as a traumatic event, or “trauma event.” To model the traumatic event, we assigned a value greater than 1.0 to the arousal parameter, ϕ_{emot}. Having been introduced to retrieved-context models by Talmi et

al. (2019), ϕ_{emot} scales the strengths of the new context-to-event associations that form in M^{CF} during encoding (equation 4). If ϕ_{emot} is greater than 1.0, then it amplifies these associations; if it is less than 1.0, then it attenuates these associations. To test the effects of temporal context on cueing trauma-memory intrusions, we set ϕ_{emot} to 5.0, because this value resulted in good variability of trauma-memory intrusion rates across virtual subjects (Figure 8).

Whereas in prior simulations we varied the contents of the developmental period, here we kept the developmental period consistent across conditions. After each developmental period, we presented virtual subjects with a neutral period consisting of 110 lists of new, unique neutral events. This allowed for 10 lists of neutral-only events after the developmental period, followed by 100 lists of the intervention for each neutral period condition. During the neutral period, we simulated three conditions, during which virtual subjects could either approach or avoid trauma cues, modeled as the three neutral stimuli that had directly preceded the traumatic event during encoding. In the *Avoidance* condition, virtual subjects did not approach any trauma cues after the traumatic event. In the *Single Exposure* condition, virtual subjects approached trauma cues once, which we modeled by embedding the set of trauma cues as new encoding events during a single list in the neutral period. In the *Repeated Exposures* condition, virtual subjects approached trauma cues eight times, for the approximate number of weeks that patients are encouraged to engage in vivo exposure exercises during a course of prolonged exposure therapy (Foa, Hembree, & Rothbaum, 2007). We modeled this by embedding the three trauma cues as new encoding events at eight regular intervals. The Avoidance, Single Exposure, and Repeated Exposures conditions allowed us to test the model's ability to capture a dose-response relationship between the number of times a virtual patient approached versus avoided trauma cues.

Results. In this simulation, we sought to determine whether retrieved-context theory could predict the effects of "trauma cues," or perceptual stimuli similar to those that directly preceded a traumatic event, on cueing intrusive recall of that event (Ehlers et al., 2000). Figure

7 shows the rate of trauma-memory intrusions for each virtual subject during the neutral period that followed exposure to the traumatic event. At the selected level of arousal ($\phi_{emot} = 5.0$), the model predicted that all subjects would spontaneously recall the traumatic event for a short time after exposure. However, by the neutral period, two groups emerged. For one group, comprising 85 of the 97 virtual subjects (87.6%), trauma-memory intrusions had subsided to fewer than one intrusion per recall list by the start of the neutral-events period (Low-Intrusions Group). In the second group that emerged during our simulation, 12 virtual subjects (12.4% of the sample) continued to experience trauma-memory intrusions at least once per recall list by the start of the neutral-events only period (High-Intrusions Group). This is not far from rates of chronic PTSD after trauma exposure in real participants, which typically range from 5-10% but can be higher depending on the type and severity of the trauma (Bonanno, Westphal, & Mancini, 2011).

Next, we assessed the model-predicted effects of encounters with trauma cues, defined as the three neutral stimuli that had directly preceded the traumatic event during encoding. In the Avoidance condition, virtual subjects approached no trauma cues; in the Single-Exposure condition, virtual subjects approached trauma cues once; and in the Multiple-Exposures condition, virtual subjects approached trauma cues eight times. For each condition, we examined the simulated rates of trauma-memory intrusions (Figure 9), the probability of correct recalls (Figure 10) of non-trauma events, and mood (Figure 11) at each timepoint.

Recall. The resulting memory and mood patterns depended on whether the virtual subject was experiencing infrequent intrusions (Low-Intrusions Group) versus frequent intrusions (High-Intrusions Group). In the Low-Intrusions group, approaching the trauma cues in both Single-Exposure and Multiple-Exposure conditions resulted in a minimal increase in trauma-memory intrusions at the first exposure timepoint (T_1), shown in Figure 9, Panel A. Otherwise, there was no change in trauma-memory intrusions in any condition over the course of the eight timepoints ($T_{1\text{-}8}$). By contrast, the High-Intrusions group showed a dose-response relationship between the number of trauma-cue exposures and reductions in trauma-memory

intrusions. Trauma-memory intrusions persisted in the Avoidance and Single-Exposure conditions (Figure 10, Panel B). However, in the Multiple-Exposure condition, the initial elevation in trauma-memory intrusions during the early exposures (T_{1-2}) declined, by the third exposure, to rates comparable with those in the single-exposure group. By exposures 5-8, virtual subjects in the Multiple-Exposure condition achieved rates of trauma-memory intrusions comparable to those in the Low-Intrusions comparison group. As re-encountering the trauma cues associated them with more varied non-trauma contexts, the cues became associated with and began to evoke a wider variety of recall outputs, thereby decreasing recall of the trauma event, but also, decreasing recall of recent events from a specific target context when these cues were part of the retrieval context (Figure 10).

Mood. In the High-Intrusions group, mood showed little improvement in the Avoidance and Single-Exposure conditions. In the Multiple-Exposure condition, mood also showed little improvement until trauma-memory intrusions began to substantially decrease at timepoint T_4 . As the frequency of trauma-memory intrusions decreased more substantially at timepoints T_{4-5}, the model predicted a clear improvement in mood (Figure 11). The Low-Intrusions group also experienced an increase in positive mood as the result of multiple encounters with trauma cues. This occurred because prior to their association with the traumatic event, the neutral trauma cues had been associated with a mildly positive context evoked by an earlier, positive stimulus. Thus, as the trauma cues became decoupled from the traumatic event due to forming associations with new neutral and positive contexts, their tendency to reactivate the trauma-memory decreased, and they instead began to evoke the positive contexts with which they had previously been associated.

The CMR3 model equations produced this effect by re-associating trauma cues with new neutral or positive contexts, not by replacing or altering prior associations. By predicting symptom change as the result of adding new associations to the trauma memory, rather than replacing old ones, CMR3 provides independent corroboration of emerging findings that, even

after successful exposure treatment, patients' fear responses can reemerge when they encounter previously-feared stimuli in contexts where new learning did not take place (Craske et al., 2008). Retrieved-context theory aligns with these prior findings by proposing that the old responses associated with the distressing stimulus are still latent and can later reemerge if either (a) the patient encounters a contextual cue that is more strongly associated with the old responses than with new responses, or (b) the current context contains few elements of the context in which learning the new responses took place.

What is novel in our work is the proposed mechanism: namely, that the restructuring of memory-to-context associations is mediated by the episodic memory system, rather than by systems involved in fear-conditioning. This allows our theory to account for the finding that perceptually similar stimuli serve to cue trauma-memory intrusions (Ehlers et al., 2004). Further, our theory can account for the observations that avoiding trauma cues perpetuates symptoms, and that *in vivo* exposure to trauma cues can reduce the frequency of intrusive memories (Foa & Kozak, 1986). However, our theory goes one step beyond fear-conditioning theories to propose a novel account of how, by restructuring episodic association networks in memory, trauma survivors can not only habituate, but also, recover prior enjoyment of trauma-associated stimuli (e.g., as in the case of a survivor of sexual assault who recovers feelings of joy and affection in romantic relationships).

Prolonged Exposure Therapy

Motivation. Among multiple efficacious treatments for trauma-memory intrusions and other symptoms in PTSD (Gallagher & Resick, 2012; Tran & Gregor, 2016), one particularly efficacious treatment is *prolonged exposure therapy* (PE; Foa & Kozak, 1986; Powers et al., 2010). Exposure-based treatments match the efficacy of other treatments for PTSD and, when study quality is accounted for, slightly outperform them (Tran & Gregor, 2016). Prolonged exposure therapy has two core components. During *in vivo* exposure exercises, patients

systematically approach feared stimuli that they have been avoiding due to their associations with the traumatic event (see *Intrusive Memories*). Second, during *imaginal exposure* exercises, patients intentionally reactivate the traumatic memory and imagine it as though it is happening again in the here-and-now.

Despite its strong empirical support, the precise mechanisms by which PE improves PTSD symptoms remain an active area of research (Rupp, Doebler, Ehring, & Vossbeck-Elsebusch, 2017). Although symptom improvement has been attributed to principles of fear conditioning and extinction, symptom reduction is not associated with the duration or degree of habituation achieved *within* individual exposure exercises (Craske et al., 2008; Jaycox, Foa, & Morral, 1998). Rather, symptom reduction is associated with the degree of habituation achieved *across* sessions (Rupp et al., 2017). This finding has raised concerns about circularity: for example, that "those who habituate, habituate," or that habituation is incidental to the success of exposure-based treatments (Brewin, 2006). In addition, the degree of initial fear activation during an exposure exercise, a core tenet of fear-conditioning theories of how PE works, is not consistently associated with PTSD symptom improvement (Rupp et al., 2017). Here, we test the ability of CMR3 to account for the efficacy of PE via the mechanism of restructuring episodic context associations, and in particular, to account for a pattern of reduced negative emotion in response to trauma memories that occurs across, rather than within, treatment sessions.

Simulation 8. Method. In this simulation, we repeated the methods from Simulation 7 (Intrusive Memories). We presented a developmental period consisting of 50 lists of 10 encoding events. Each encoding list contained 40% neutral, 30% positive, and 30% negative events and was followed by a simulated free recall session. During the developmental period, we randomly selected a negative event to serve as the trauma event. This time, we amplified the strength of the encoded associations between the trauma event and its context by a factor of $\phi_{emot} = 50.0$. We selected this high value to ensure that all subjects would experience trauma-memory intrusions in the subsequent neutral period.

After the developmental period in each simulation, we presented a neutral period consisting of 110 lists of neutral events. In the neutral period, we presented one of four conditions: (1) the Baseline condition, (2) the Imaginal Exposure condition, (3) the Positive Events control condition, and (4) the Re-encoding control condition. During the *Baseline* condition, the neutral period consisted solely of new, neutral events. The Baseline condition served as a manipulation check to ensure that the chosen level of arousal resulted in high rates of trauma-memory intrusions for all virtual subjects. During the *Imaginal Exposure* condition, we simulated the re-encoding of traumatic memories in safe, therapeutic contexts. We simulated this by re-introducing the trauma-memory at five regular intervals during the neutral period as a new encoding event. Prior to the re-encoding of the trauma memory, we modeled a safe context by introducing a positively valent stimulus. Because patients' negative affect tends to increase in anticipation of the imaginal exposure exercise, we introduced the positive stimulus several stimuli prior to the re-encoding of the trauma memory, such that positive affect would decrease as the moment of re-encoding the trauma memory approached.

To clarify the mechanisms of any change observed in the Imaginal Exposure condition, we included two control conditions. The first control was a *Positive Events* condition, in which we introduced a new, positive event at identical intervals (i.e., every 20 lists) in the neutral period as the re-encoding of the trauma event during the Imaginal Exposure condition. The Positive Events condition served as a control to verify that the Imaginal Exposure condition did not reduce negative emotion simply due to the positive events that preceded re-encoding of the trauma memory. The second control was a *Re-Encoding* condition, in which we introduced the trauma memory as a new encoding event at the same 5 regular intervals in the neutral period. However, unlike in the *Imaginal Exposure* condition, we did not evoke a less-negative context prior to its encoding. This served to test whether simply re-encoding the memory in new contexts would promote mood recovery, regardless of the emotional content of those new

contexts. As in the prior simulation, we calculated the average mood within each period, the rate of correct recalls of non-trauma events, and the rate of trauma-memory intrusions.

Results. In this simulation, we sought to test whether CMR3 could predict the finding that exposure to the trauma memory reduces the negative emotion evoked by the memory when it is recalled. In addition, we sought to determine whether CMR3 could predict the finding that reduced negative emotion across, but not within, such exposure exercises is associated with symptom improvement (Jaycox et al., 1998; Craske et al., 2008).

Mood patterns. The results for an individual virtual subject are shown in Figure 12, and the aggregated effects across virtual subjects are shown in Figure 13. First, as a manipulation check, we confirmed that the trauma event evoked severe and persistent negative emotion during the developmental period at the selected level of arousal ($\phi_{emot} = 50.0$). In the Baseline Condition, the trauma event resulted in high rates of recurrent memory intrusions throughout the subsequent neutral period. Without intervention, the repeated memory intrusions evoked negative context, which resulted in negative mood that persisted for the remainder of the neutral period (Figure 12, Panel A; Figure 13), as occurred in Simulation 5 (Repeated Stressors). Next, in the Positive Control Condition, we presented positive events at 5 regular intervals. Each event briefly instated a positive context when it occurred, but continued intrusions of the trauma-memory returned mood to a negative baseline each time (Figure 12, Panel B; Figure 13). In the Re-encoding condition, we presented the trauma event for re-encoding at the same 5 regular intervals, which resulted in no mood improvements (Figure 13). Finally, in the Imaginal Exposure condition, we first induced less-negative contexts and then paired these with a subsequent presentation of the trauma memory as an encoding event. Over the course of the five re-encodings of the trauma memory, the trauma memory began to associate with the new less-negative contexts. After each intentional re-presentation of the trauma memory, the subsequent trauma-memory intrusions evoked progressively reduced negative emotion, until mood stabilized near a neutral value of 0.0 (Figure 12, Panel C; Figure 12).

Because the trauma memory reinstates its old associated context when first retrieved, CMR3 predicts that its new updated context will not be evoked until the *next* time the memory is retrieved. Thus, the phenomenon of reduced negative emotion, or apparent habituation, *across* rather than within sessions of imaginal exposure emerged naturally from the model. This effect can be observed in Figure 12 (Panel B), and in Figure 13. Whereas the mean level of emotion becomes less negative after each re-encoding of the trauma memory (i.e., after each virtual PE session) designated by the vertical lines, the degree of negative emotion evoked by the memory during its reactivations stays level within each session. That is, during the current session, the memory evokes the negative context that had previously been associated with it. Then, each time the trauma memory is re-encoded in association with a less-negative context, it updates its associated affect and thus will evoke less-negative context the next time it is retrieved, resulting in an across-session reduction in negative affect.

Recall patterns. The high value of the arousal parameter, $\phi_{emot} = 50.0$, which strengthened the context-to-event associations during encoding, resulted in high rates of trauma-memory intrusions. The rate of trauma-memory intrusions remained essentially constant throughout timepoints of the simulation, with minimal variability across virtual subjects. Therefore, we report the mean frequency of trauma-memory intrusions across the timepoints in each condition: the mean frequency was 13.8 (SEM = .003) for the Baseline and Positive-Events simulations and 14.0 (SEM = .020) for the Re-Encoding and Imaginal Exposure simulations. Due to the very high rates of intrusions induced by the high arousal level used in this simulation, the probability of correct recall was low, and also consistent across simulations. The mean probability of correct recall across timepoints was .020 (SEM = .0003) for the Baseline and Positive Events simulations and .010 (SEM = .0020) for the Re-Encoding and Imaginal Exposure simulations.

Thus, the simulations predict a slight increase in trauma-memory intrusions for the two conditions that involved a re-presentation of the traumatic memory, and concurrently, a slight

decrease in probability of recalling other events. This is consistent with findings that although imaginal exposure is efficacious for alleviating negative affect, the re-presentation of the traumatic memory may initially result in a slight elevation in trauma-memory intrusions in a minority of patients (Foa, Zoellner, Feeny, Hembree, & Alvarez-Conrad, 2002). In CMR3, a decreased ability to recall nontraumatic stimuli, together with greater activation of the trauma-memory, occurs because the activated trauma-related context is dissimilar to the context of nontraumatic or neutral events, and therefore is less able to cue recall of these stimuli.

General Discussion

Emotional disorders, particularly depression but also including anxiety and PTSD, are the leading source of disability worldwide (World Health Organization, 2017). Their high comorbidity has generated a search for transdiagnostic processes (Insel et al., 2010). Such processes, which contribute to negative emotion across disorders, can point to more efficacious treatments for patients with complex diagnoses. Learning and memory processes have long been implicated as one such mechanism. For example, research on classical conditioning has fueled a number of treatment protocols for depression and PTSD, including behavioral activation therapy (Beck et al., 1979) and prolonged exposure therapy (Foa & Kozak, 1986). However, many phenomena remain unexplained, including why patients sometimes experience negative emotion without any active negative cognitions, and why exposure therapies are so effective for treating PTSD yet are weakly associated with their proposed mechanisms (Rupp et al., 2017). In addition, it is an active area of research whether memory for high-arousal negative events is handled by the episodic memory system, which supports memory for other personally-experienced events.

In this article, we proposed a new theory of how episodic memory processes interact with mood to produce psychopathology, which we formalize in a computational model called CMR3. The model we develop benefits greatly from prior work on context maintenance and

retrieval (CMR) models, particularly the eCMR model developed by Talmi and colleagues (2019), who first conceptualized emotion as a component of memories and their contexts. In addition, we draw upon major advances by prior theories of memory and emotion in clinical disorders, particularly major depression and posttraumatic stress disorder (Bower, 1981; Brewin, 2006, 2014; Ehlers & Clark, 2000; Foa & Kozak, 1986; Rubin et al., 2008; Rubin et al., 2011). The theory we develop benefits from the insights of this work while introducing novel processes that account for both classic and previously unexplained phenomena of memory in emotional disorders.

Retrieved-context theory of memory in emotional disorders. Retrieved-context theories of memory propose two core operations. First, memories form associations with their contexts during encoding. Later, perceptual cues in new contexts activate these networks to retrieve memories that were associated with similar contexts. Once a memory is retrieved, it reactivates elements of its encoding context, which can in turn cue other memories. The memory, its old context, and its new context, are then re-encoded together as part of a recursive process that updates the memory to associate it with new information. Recently, Talmi et al. (2019) broke new ground by conceptualizing emotion as a component of memories and their contexts. However, the ability of eCMR to capture phenomena in clinical disorders is limited by two features. First, eCMR models emotion as a single-dimensional construct, without having positive or negative valence, such that the model cannot predict what differentially leads to the activation of negative versus positive emotional context. Second, eCMR limits the effects of prior learning to a single experimental list, such that emotional learning could not accrue over the course of a patient's lifespan. Further, eCMR cannot distinguish between a target vs. non-target prior context, a necessary condition to modeling intrusive memories of distressing events, a core phenomenon of interest in PTSD and other disorders.

In this paper, we introduced a new retrieved-context theory of memory in mood, which we formalized as a mathematical model called CMR3. In CMR3, we introduced emotional

valence as a property of both memories and mood states (Figure 1), and we introduced the multilist capabilities of CMR2 (Lohnas et al., 2015). This allows CMR3 to model how memories from a non-target context can intrude into the current context, as well as how emotional learning accrues and influences behavior throughout a patient's lifespan. To empirically test how to represent emotional valence in CMR3, we compared the fits of three retrieved-context models: (1) CMR2, which represents emotion as a type of semantic content; (2) eCMR, which additionally represents emotion as present or absent in memories and their contexts, but with no valence; and (3) CMR3, which represents emotion as multivalent. To enable a direct comparison between each model, we coded a version of eCMR that has multilist capabilities, thus allowing it to predict recall intrusions alongside CMR2 and CMR3. Further, to isolate the effects of the representation of emotional valence in fitting the data, we held the effects of arousal constant in the model, $\phi_{emot} = 1.0$. CMR3 provided the best fits to the aggregate behavioral responses (see *Modeling Emotional Memory*), indicating that emotional valence functions as a component of both memories and their contexts, not just as semantic content, and that it is important to include it in the model. It is also important to include emotional valence in the model for theoretical purposes, since a model of emotion in clinical disorders cannot function without the ability to differentiate between the effects of positive and negative experiences.

Transdiagnostic framework. In this paper, we have sometimes used the language of diagnostic categories, particularly major depression and PTSD. This approach has allowed us to guide our theory by drawing upon a wealth of literature examining memory and mood in specific disorders. However, retrieved-context theory reflects recent initiatives to use findings and methods from cognitive neuroscience to identify the basic processes that cut across disorders (Montague et al., 2012; Sanislow et al., 2010). This approach is consistent with findings that disorders characterized by persistent negative emotion are highly comorbid (Brady et al., 2000; Kessler, Chiu, Demler, & Walters, 2005), and that memory disturbances, such as intrusive

memories of painful events, occur widely across emotional disorders (Brewin et al., 2010; Day et al., 2004; Hackmann et al., 2000; Osman et al., 2004; Speckens et al., 2007). We propose a framework in which intrusive memories and negative mood result from basic processes of cognition and emotion, occurring more often under some conditions, such as patterns of environmental stress interacting with a person's individual characteristics, but not constrained to a specific DSM-5 disorder. We outline retrieved-context theory's framework and proposed processes below.

Mood-congruent and Emotion-state dependent recall. In characterizing memory processes in emotional disorders, we began by demonstrating how retrieved-context theory can account for the interplay of mood and emotion in psychopathology. Using the match between encoding and retrieval contexts, CMR3 accounted for mood-congruent recall (Simulation 2) and emotion-state dependent recall (Simulation 3). In mood-congruent recall, items whose valence matches that of the retrieval context will have introduced that valence into their encoding contexts, and thus will have a stronger activation to promote their retrieval. In emotion-state dependent recall, neutral items are better recalled if the emotional context present during encoding contexts matches the emotional context present during retrieval. Although emotion-state dependent learning is observed inconsistently in laboratory settings with healthy participants (Eich, 1995), its influence appears more strongly in clinical populations (Craske et al., 2008). Retrieved-context theory is consistent with findings that new learning in exposure-based treatments for anxiety disorders is enhanced when new learning occurs in the same emotional context in which it will later be retrieved (Craske et al., 2008).

In his prior spreading activation theory, Bower (1981) proposed that mood-congruent and emotion-state dependent memory occur because emotional valence serves as an node in semantic networks. A positive or negative emotion thus activates semantically associated memories, which tend to share the same emotional properties due to semantic relatedness. In CMR3, because emotional valence also serves as a part of a memory's episodic context, it can

become associated with stimuli that themselves are neutral and have no semantic relation with negative emotion. This aspect of our theory is important for allowing CMR3 to model the generalization of negative affect to neutral events (Simulation 4) and how stimuli that have little semantic association to a traumatic event, but that preceded the event, can later cue intrusive memories and trauma-associated affect (Simulation 7). We discuss our model's predictions of these phenomena below.

Persistent Negative Emotion. Retrieved-context theory predicts that emotional context can linger after an event and thus bind to new events, even if these new events are neutral (Simulation 4). Thus, when stressors occur and induce negative emotion, memories of the new events that follow are encoded in association with the lingering negative emotion. Whereas the subsequent events may be experienced as neutral or positive during their encoding, the emotion that they evoke when later retrieved becomes more negative due to the negative emotional context present during encoding. The ability of negative emotion to bind to later events can explain how a depressed patient might experience a positive event as enjoyable while experiencing it, but later recall the event as having been more negative, since memory of the event activates not only its positive properties but further, the negative context with which it had become associated. Further, this mechanism can account for the phenomenon of "affect before cognition," in which patients report experiencing negative emotion without having any active thoughts or memories about a stressor, especially when perceptual stimuli associated with prior negative events are present (Ehlers & Clark, 2000).

Repeated stressors. Retrieved-context theory predicts that repetition has powerful effects on memory. Spaced repetitions of stressors associate the event with a larger number of contexts that can later serve as cues to activate the memory for spontaneous retrieval (Lohnas, Polyn, & Kahana, 2011; Madigan, 1969; Melton, 1970). Retrieved-context theory proposes that repetition, either by experiencing the stressful event multiple times (Simulation 5), or through repeated rehearsal and re-encoding of the memory in many different contexts, will increase the

chance that the memory may be cued for retrieval into a context in which it does not belong. Indeed, rumination, which includes the repetitive rehearsal and elaboration upon negative memories, is associated with and can prompt intrusive memories in PTSD (Michael, Halligan, Clark, & Ehlers, 2007). Our simulations are consistent with a core tenet of the *Autobiographical Memory Theory* (AMT) of intrusive memories in PTSD, which proposes that repeated rehearsal of distressing memories can lead to intrusive memories by amplifying the memory's accessibility (Rubin et al., 2011). We extend the AMT by proposing a novel mechanism for this process: that repeated rehearsal associates the memory with multiple contexts, thus increasing the probability that future contexts will contain some element of the memory's associated cues and thereby activate it for recall in contexts where the memory otherwise does not belong.

Behavioral Activation Therapy. Retrieved-context theory predicts that increasing the ratio of positive to negative events, as in behavioral activation therapy (Beck et al., 1979; Jacobson et al., 1996; Lewinsohn, Sullivan, & Grosscap, 1980), results in sustained mood improvement (Simulation 6). However, the model predicts that increasing the ratio of positive events is most effective when a patient's negative emotion is the result of the generalization of negative emotion to neutral events. When negative mood is being maintained by intrusive negative memories, our simulations predict that patients will initially have a short-term positive response to positive events, but will experience blunted responsiveness to positive events in the long run as memory intrusions reactivate negative emotional context. As such, retrieved-context theory highlights the importance of addressing intrusive memories in major depression. Treatments targeting this symptom have shown promising results (Brewin et al., 2009).

Intrusive memories of high-arousal, negative events. Intrusive memories occur prominently in PTSD, but they also occur widely across other disorders (Berntsen, 2010; Day et al., 2004; Hackmann et al., 2000; Muse et al., 2010; Osman et al., 2004; Price et al., 2012; Reynolds & Brewin, 1999; Speckens et al., 2007) and for painful events not traditionally categorized as traumatic (Brozovich & Heimberg, 2008; Williams & Moulds, 2007). Prior

theories, such as the *Autobiographical Memory Theory* (Rubin et al., 2011), have proposed that high emotional intensity, or emotional arousal, strengthens the accessibility of a memory, thereby predisposing it to spontaneously activate. We extend this theory by proposing that high emotional arousal during encoding is the driving mechanism, serving to strengthen the associations between an event and its context during encoding, as introduced in the eCMR model (Talmi et al., 2019). In a retrieved-context framework, strengthened context-to-event associations potentiate the memory for retrieval, since lower levels of associated context is needed to cue these associations. Further, strengthened associations strengthen the probability that unrelated contexts will contain enough elements of associated context to trigger the memory (see Simulations 7 and 8).

To allow the model to distinguish between recalling an event from a target context (voluntary recall) versus from a non-target context (a memory intrusion), we combined the high-arousal mechanism of eCMR (Talmi et al., 2019) with the multilist capabilities of CMR2 (Lohnas et al., 2015). The new model, CMR3, can predict findings that encountering the perceptual stimuli that directly preceding the event (its temporal context) cue heightened rates of trauma-memory intrusions (Simulation 7), even if the stimuli have no semantic relationship to the event (Ehlers & Clark, 2000; Ehlers et al., 2004). Accordingly, CMR3 proposes that re-associating the temporal context of a traumatic memory with new neutral or positive contexts, as occurs during *in vivo* exposure exercises, will lead these cues to eventually evoke affect, cognitions, and memories that are not associated with the traumatic event (Simulation 7). This theory is consistent with theories that extinction occurs by encoding new associations that exist alongside the feared stimulus' prior associations (Foa & McNally, 1996). In addition, retrieved-context theory predicts that old associations can still be reactivated when the feared stimuli are encountered in contexts that are dissimilar to the contexts in which new learning has taken place (Craske et al., 2008).

Intrusive memories when fear is not the primary emotion. Patients with PTSD often report that primary emotional responses other than fear, such as shame or sadness, are associated with their traumatic event (Resick & Schnicke, 1992). In addition, patients with depression often report vivid memory-intrusions characterized by intense negative emotion other than fear (Kuyken & Brewin, 1994; Reynolds & Brewin, 1999). Retrieved-context theory predicts the development of intrusive memories when high-arousal negative emotion is present during encoding, without constraints for what type of emotion must be present.

Prolonged Exposure Therapy. Based in principles of fear conditioning and extinction, Prolonged Exposure therapy (PE) is a highly efficacious treatment for PTSD (Powers et al., 2010; Tran & Gregor, 2016). *Emotional processing theory* (EPT), the theory behind PE, proposes that exposure-based treatments work by updating a “fear network” (Lang, 1979), or the set of associations between the fear-memory and otherwise neutral stimuli that were formed during the traumatic event, and which are maintained by avoiding opportunities for disconfirming information after the event (Foa & Kozak, 1986). In our *Prolonged Exposure Therapy* simulation (Simulation 8), we demonstrate that episodic memory principles predict a decrease in negative affect due to the repeated reactivation and re-encoding of negative memories in new, safe or neutral contexts. This result adds independent confirmation to a recent neurobiological theory, which proposes that symptom improvement depends on updating and reconsolidating emotional memories (Lane, Ryan, Nadel, & Greenberg, 2015). Further, the finding that PTSD symptom reduction is associated with apparent habituation (i.e., a decrease in negative affect due to reactivating the traumatic memory) across, but not within, treatment sessions (Jaycox et al., 1998; Rupp et al., 2017) emerges naturally from our model (Simulation 8). In an episodic memory framework, treatment first activates the memory and then re-encodes it in association with a changed emotional context. The updated emotional context is only reactivated the *next* time the memory is retrieved, such that negative affect decreases across rather than within sessions (Simulation 8).

The new theory we propose builds upon the framework of EPT to propose that instead of a “fear network” mediated by fear-learning systems, there exists an “emotion network” in episodic memory. This approach is consistent with Lang’s (1979) bioinformational theory of emotion, which focused on fear but proposed that networks of associations between stimuli, responses, and cognitions exist in memory for emotion more broadly. Fear-extinction learning may still contribute to recovery, perhaps combining with episodic memory processes to enhance the efficacy of PE for patients who experience fear (vs. other emotions) associated with their trauma (Foa, Riggs, Massie, & Yarczower, 1995).

Retrieved-context theory’s prediction of PE’s efficacy distinguishes it from *dual-representation theory.* Dual-representation theory (DRT) constituted a major advance in theories of memory in PTSD by proposing that high emotional arousal during encoding influences the strength of the associations between traumatic memories and their temporal contexts (Brewin, 2014; Brewin et al., 2010). Specifically, DRT proposes that high arousal disrupts the associations between events’ perceptual features and their spatiotemporal contexts (Brewin, 2014). Findings of stronger associations between negative stimuli during recall, and weaker associations between negative and neutral stimuli, inside the lab support this account (Bisby, Horner, Bush, & Burgess, 2018). In *Modeling Emotional Memory* (Simulation 1), we show that retrieved-context theory similarly predicts the tendency of emotional items to selectively cluster together with same-valent items during a laboratory free recall task. However, the predictions of DRT and retrieved-context theory diverge with respect to clinical literature. DRT predicts that the same intense distress which disrupted episodic memories’ ability to bind to their encoding contexts while experiencing the event should do the same during treatments that involve high levels of negative emotion. Thus, DRT predicts that in vivo or imaginal exposure exercises in which high distress is present should be harmful for treating intrusive memories in PTSD (Nadel & Jacobs, 1996; Bisby et al., 2018).

However, prolonged exposure therapy is highly efficacious for treating PTSD (Powers et al., 2010; Tran & Gregor, 2016). Intense negative emotion during exposure to trauma memories is either associated with better treatment response (Jaycox et al., 1998) or else is not correlated with treatment response (Rupp et al., 2017). In anxiety disorders more broadly, high levels of negative emotion and physiological arousal during exposure exercises are associated with better treatment response, even when the emotional intensity does not reduce by the end of the session (Kircanski et al., 2012; Lang & Craske, 2000). The retrieved-context theory we develop in this paper predicts these effects (*Intrusive Memories,* Simulation 7; *Prolonged Exposure Therapy,* Simulation 8).

Efficacy of cognitive therapies for PTSD. Considering emotion as a type of episodic context also enables retrieved-context theory to predict the efficacy of treatments like *Cognitive Processing Therapy* for PTSD*,* which are equally efficacious as PE (Rizvi, Vogt, & Resick, 2009). Our theory predicts that modifying the negative emotional context associated with a traumatic event, as occurs through restructuring maladaptive cognitions in *cognitive processing therapy* (Resick & Schnicke, 1992), should result in the trauma memory being re-encoded with the new less-negative context. This will both result in an apparent habituation effect and also reduce the tendency for intrusive memories cued by a match between the current emotional state and the emotions that were present during encoding (see Mood-congruent Recall, Simulation 2). Overall, retrieved-context theory proposes two treatment routes for PTSD: modifying the context-to-memory associations in the emotion network, as in PE (Foa & Kozak, 1986) and modifying the contents of the emotional context associated with the traumatic memory, as in CPT (Resick & Schnicke, 1992).

Limitations. Although retrieved-context theory provides advantages over dual-representation theory in its ability to predict certain treatment effects, it is worth noting that dual-representation theory was developed to characterize the internal consistency of traumatic memories (Brewin, 2014), which CMR3 does not fully address. In CMR3, events are

represented either on the level of individual stimuli, or a larger event can be presented as a series of stimuli (a “list”), from which the model predicts the retrieval of multiple elements. In *Intrusive Memories* (Simulation 7), CMR3 demonstrates how high arousal negative emotion during encoding of a list of stimuli can predispose trauma-related stimuli for intrusive memories (Figure 9), while decreasing the probability of correct recall of other stimuli (Figure 10). However, even when conceptualizing an event as a set or list of stimuli, in CMR3 each stimulus is presented one at a time. Further, CMR3 does not differentiate among the modalities of each stimulus (i.e., whether they are visual, auditory, internal physiological sensations, and so forth). It may be the case that CMR3 provides a strong account of the associations between distinct stimuli and traumatic events, whereas dual-representation theory may provide a stronger account of the internal consistency of the different perceptual experiences that make up a memory across different modalities. Empirical testing of both models, beyond the simulation work presented in this paper to develop retrieved-context theory, will shed light.

A second limitation lies in one of our model’s strengths, which is the ability to predict a wide array of phenomena in emotional disorders without making strict assumptions about what type of emotion is being experienced. This allows it to account for patterns of comorbidity between mood and stress disorders, as well as how intrusive memories can occur for many types of distressing events. However, in real life, emotion is a complex, multivariate construct whose attributes extend beyond negative-positive valence and low-high arousal. For example, anger and fear are both negatively valent and high-arousal emotions, yet they have distinct subjective experiences, action tendencies (avoid vs. approach; Carver & Harmon-Jones, 2009), and neural patterns (Kragel & LaBar, 2015). In the current version of CMR3, the effects of fear and anger would be indistinguishable. In addition, not all negative emotional states induce high arousal: for example, disgust may be associated with reduced physiological arousal (Ekman, Levenson, & Friesen, 1983). Future research is needed to characterize the full dimensionality and role of emotion in the memory system.

Finally, CMR3 represents mood and memory outcomes as the result of different processes in a simulated mind that are constrained by sets of parameter values, which determine how strongly each process will operate for a given individual. For example, CMR3 can represent an individual with heightened emotional sensitivity by setting the value of β_{emot}, the rate at which emotional information enters and updates the cognitive system, to a higher value. However, the model does not explain what led to a particular set of parameter values in the first place. The future development of retrieved-context theory will benefit from insights from biological and developmental models of how patients acquire distinct cognitive parameters innately and throughout development.

Conclusion

Retrieved-context theory proposes a parsimonious account of episodic memory that generates persistent negative emotion, intrusive memories, and their co-occurrence. This paper presents the first transdiagnostic, computational model of memory and mood in emotional disorders as these phenomena unfold over time. By fitting the model to empirical data and conducting simulations, we can test whether the predictions of narrative theories, when constrained by their proposed operations, map on to empirical literature. Together with recent initiatives to use computational modeling to characterize clinical phenomena (Bennett, Silverstein, & Niv, 2019; Montague et al., 2012), we propose that developing a clinical theory will benefit from testing whether, given an appropriate set of inputs, a formal model generated from that theory can capture and predict behavioral patterns from experimental and clinical paradigms. The resulting model should be able to predict the efficacy of existing treatments. In addition, the model must do so via mechanisms that map on to our understanding of human cognition and neural architecture. In this paper, we accomplish the first two criteria and lay a foundation for empirical work to accomplish the third. We hope this new framework will be used to refine and challenge the retrieved-context model, thus ultimately improving our understanding of memory and mood in psychopathology.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. *Arlington, VA*. American Psychiatric Publishing.

Bates, D. M., Mäechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.

Benjet, C., Bromet, E., Karam, E. G., Kessler, R. C., Mclaughlin, K. A., Ruscio, A. M., … Koenen, K. C. (2016). The epidemiology of traumatic event exposure worldwide: Results from the World Mental Healthy Survey Consortium. *Psychological Medicine*, *46*(2), 327–343. https://doi.org/10.1017/S0033291715001981.The

Bennett, D., Silverstein, S. M., & Niv, Y. (2019). The Two Cultures of Computational Psychiatry. *JAMA Psychiatry*, *76*(6), 563–564. https://doi.org/10.1038/nn.4238

Berntsen, D. (2010). The unbidden past: Involuntary autobiographical memories as a basic mode of remembering. *Current Directions in Psychological Science*, *19*(3), 138–142.

Bevington, P. R., & Robinson, D. K. (2003). *Data Reduction and Error Analysis for the Physical Sciences* (Third). New York, NY: McGraw-Hill HIgher Education.

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative Emotional Content Disrupts the Coherence of Episodic Memories. *Journal of Experimental Psychology: General*, *147*(2), 243–256.

Bonanno, G. A., Westphal, M., & Mancini, A. (2011). Resilience to Loss and Potential Trauma. *Annual Review of Clinical Psychology*, *7*, 511–535. https://doi.org/10.1146/annurev-clinpsy-032210-104526

Bower, G.H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148.

Bower, Gordon H., & Mayer, J. D. (1985). Failure to replicate mood-dependent retrieval. *Bulletin*

of the Psychonomic Society, *23*(1), 39–42. https://doi.org/10.3758/BF03329773

Bradley, B. P., Mogg, K., & Williams, R. (1995). Implicit and explicit memory for emotion-congruent information in clinical depression and anxiety. *Behaviour Research and Therapy*, *33*(7), 755–770. https://doi.org/10.1016/0005-7967(95)00029-W

Brady, K. T., Killeen, T. K., Brewerton, T., & Lucerini, S. (2000). Comorbidity of psychiatric disorders and posttraumatic stress disorder. *The Journal of Clinical Psychiatry*, *61*, 22–32.

Brewin, C. R. (2006). Understanding cognitive behaviour therapy: A retrieval competition account. *Behaviour Research and Therapy*, *44*(6), 765–784.

Brewin, C. R. (2014). Episodic memory, perceptual memory, and their interaction: foundations for a theory of posttraumatic stress disorder. *Psychological Bulletin*, *140*(1), 69–97.

Brewin, C. R., Andrews, B., Rose, S., & Kirk, M. (1999). Acute stress disorder and posttraumatic stress disorder in victims of violent crime. *Am J Psychiatric*.

Brewin, C. R., Gregory, J. D., Lipton, M., & Burgess, N. (2010). Intrusive images in psychological disorders: characteristics, neural mechanisms, and treatment implications. *Psychological Review*, *117*(1), 210–232.

Brewin, C. R., Wheatley, J., Patel, T., Fearon, P., Hackmann, A., Wells, A., … Myers, S. (2009). Imagery rescripting as a brief stand-alone treatment for depressed patients with intrusive memories. *Behaviour Research and Therapy*, *47*(7), 569–576.

Brewin, Chris R, Andrews, B., & Valentine, J. D. (2000). Meta-Analysis of Risk Factors for Posttraumatic Stress Disorder in Trauma-Exposed Adults. *Journal of Consulting and Clinical Psychology*, *68*(5), 748–766. https://doi.org/10.1037//0022-006X.68.5.748

Brown, G. W., Harris, T. O., & Hepworth, C. (1994). Life Events and Endogenous Depression. *Archives of General Psychiatry*, *51*, 525–534.

Brozovich, F., & Heimberg, R. G. (2008). An analysis of post-event processing in social anxiety disorder. *Clinical Psychology Review*, *28*(6), 891–903.

Cabib, S., & Puglisi-Allegra, S. (1996). Stress, depression and the mesolimbic dopamine

system. *Psychopharmacology*, *128*(4), 331–342. https://doi.org/10.1007/s002130050142

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401–423.

Carver, C. S., & Harmon-Jones, E. (2009). Anger Is an Approach-Related Affect: Evidence and Implications. *Psychological Bulletin*, *135*(2), 183–204. https://doi.org/10.1037/a0013965

Craske, M., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. *Behavior Research and Therapy*, *46*, 5–27. https://doi.org/10.1016/j.brat.2007.10.003

Day, S. J., Holmes, E. A., & Hackmann, A. (2004). Occurrence of imagery and its link with early memories in agoraphobia. *Memory*, *12*(4), 416–427. https://doi.org/10.1080/09658210444000034

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*(4), 319–345.

Ehlers, A., Hackmann, A., & Michael, T. (2004). Intrusive re-experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory*, *12*(4), 403–415.

Eich, E. (1995). Searching for Mood Dependent Memory. *Psychological Science*, *6*(2), 67–75.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic Nervous System Activity Distinguishes among Emotions. *Science*, *221*, 1208–1210.

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: exposure to corrective information. *Psychological Bulletin*, *99*(1), 20–35.

Foa, E.B., Hembree, E., & Rothbaum, B. O. (2007). *Treatments that work: Prolonged exposure therapy for PTSD: Emotional processing of traumatic experiences. Therapist Guide.* New York, NY: Oxford University Press.

Foa, E.B., & McNally, R. J. (1996). Mechanisms of change in exposure therapy. *Current Controversies in the Anxiety Disorders*, 329–343.

Foa, E.B., Riggs, D. S., Massie, E. D., & Yarczower, M. (1995). The Impact of Fear Activation and Anger on the Efficacy of Exposure Treatment for Posttraumatic Stress Disorder. *Behavior Therapy*, *26*, 487–499.

Foa, Edna B., Zoellner, L. A., Feeny, N. C., Hembree, E. A., & Alvarez-Conrad, J. (2002). Does Imaginal Exposure Exacerbate PTSD Symptoms? *Journal of Consulting and Clinical Psychology*, *70*(4), 1022–1028. https://doi.org/10.1037//0022-006X.70.4.1022

Goenjian, A. K., Molina, L., Steinberg, A. M., Fairbanks, L. A., Alvarez, M. L., Goenjian, H. A., & Pynoos, R. S. (2001). Posttraumatic Stress and Depressive Reactions Among Nicaraguan Adolescents After Hurricane Mitch. *American Journal of Psychiatry*, *158*(5), 788–794.

Hackmann, A., Clark, D. M., & McManus, F. (2000). Recurrent images and early memories in social phobia. *Behaviour Research and Therapy*, *38*, 601–610.

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, *143*(2), 575–596.

Herman, J. L. (1992). Complex PTSD: A Syndrome in Survivors of Prolonged and Repeated Trauma. *Journal of Traumatic Stress*, *5*(3), 377–391.

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*(1), 85–98.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *The American Journal of Psychiatry*, 748–751.

Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral Activation Treatment for Depression : Returning to Contextual Roots. *American Psychological Association*, (c), 255–270.

Jaffee, S. R., Moffitt, T. E., Caspi, A., Fombonne, E., Poulton, R., & Martin, J. (2002).

Differences in Early Childhood Risk Factors for Juvenile-Onset and Adult-Onset Depression. *Archives of General Psychiatry*, *59*, 215–222.

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.

Johnson, S. L., & Roberts, J. E. (1995). Life Events and Bipolar Disorder: Implications from Biological Theories. *Psychological Bulletin*, *117*(3), 434–449.

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(2), 92–87.

Kahana, Michael J, Aggarwal, E. V, & Phan, T. D. (2018). The Variability Puzzle in Human Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1857–1863.

Kanter, J. W., Manos, R. C., Bowe, W. M., Baruch, D. E., Busch, A. M., & Rusch, L. C. (2010). What is behavioral activation? A review of the empirical literature. *Clinical Psychology Review*, *30*, 608–620. https://doi.org/10.1016/j.cpr.2010.04.001

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1998). Stressful Life Events and Major Depression: Risk Period, Long-Term Contextual Threat, and Diagnostic Specificity. *The Journal of Nervous and Mental Disease*, *186*(11), 661–669.

Kendler, K. S., Kessler, R. C., Walters, E. E., MacLean, C., Neale, M. C., Heath, A. C., & Eaves, L. J. (1995). Stressful Life Events, Genetic Liability, and Onset of an Episode of Major Depression in Women. *American Journal of Psychiatry*, *152*(July), 833–842. https://doi.org/10.1176/foc.8.3.foc459

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*, *62*, 617–627.

https://doi.org/10.1037/a0034733.Comorbidity

Kircanski, K., Mortazavi, A., Castriotta, N., Baker, A. S., Mystkowski, J. L., Yi, R., & Craske, M. G. (2012). Challenges to the traditional exposure paradigm: Variability in exposure therapy for contamination fears. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(2), 745–751. https://doi.org/10.1016/j.jbtep.2011.10.010

Koren, D., Arnon, I., & Klein, E. (1999). Acute stress response and posttraumatic stress disorder in traffic accident victims: a one-year prospective, follow-up study. *American Journal of Psychiatry*, *156*(3), 367–373.

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*(11), 1437–1448. https://doi.org/10.1093/scan/nsv032

Kuyken, W., & Brewin, C. R. (1994). Intrusive memories of childhood abuse during depressive episodes. *Behavior Research and Therapy*, *32*(5), 525–528.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects. *Journal of Statistical Software*, *82*(13), 1–26.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*, e1, 1–64. https://doi.org/10.1017/S0140525X14000041

Lang, A. J., & Craske, M. G. (2000). Manipulations of exposure-based therapy to reduce return of fear: A replication. *Behaviour Research and Therapy*, *38*(1), 1–12. https://doi.org/10.1016/S0005-7967(99)00031-5

Lang, P. J. (1979). A Bio-informational Theory of Emotional Imagery. *Psychophysiology*, *16*(6), 495–512.

Lang, P. J. (1995). The Emotion Probe: Studies of Motivation and Attention. *American Psychologist*, *50*(5), 372–385.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual Variability in Free Recall.

Journal of Memory and Language, *64*(3), 249–255.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337–363.

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin & Review*, *22*, 1328–1333. https://doi.org/10.3758/s13423-014-0791-2

Madigan, S. A. (1969). Intraserial Repetition and Coding Processes in Free Recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 828–835.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.

Mathews, A., Mogg, K., May, J., & Eysenck, M. (1989). Implicit and Explicit Memory Bias in Anxiety. *Journal of Abnormal Psychology*, *98*(3), 236–240. https://doi.org/10.1037/0021-843X.98.3.236

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, *12*(2), 227–255. https://doi.org/10.1016/0272-7358(92)90116-P

Melton, A. W. (1970). The Situation with Respect to the Spacing of Repetitions and Memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596–606.

Michael, T., Halligan, S. L., Clark, D. M., & Ehlers, A. (2007). Rumination in Posttraumatic Stress Disorder. *Depression and Anxiety*, *24*, 307–317.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 3111–3119.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, *16*(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic Memory and Beyond:

The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, *67*, 105–134. https://doi.org/10.1146/annurev-psych-113011-143733

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488.

Muse, K., Mcmanus, F., Hackmann, A., Williams, M., & Williams, M. (2010). Intrusive imagery in severe health anxiety: Prevalence, nature and links with memories and maintenance cycles. *Behaviour Research and Therapy*, *48*, 792–798.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods*, *36*(3), 402–407.

Osman, S., Cooper, M., Hackmann, A., & Veale, D. (2004). Spontaneously occurring images and early memories in people with body dysmorphic disorder. *Memory*, *12*(4), 428–436.

Paunovic, N., Lundh, L. G., & Öst, L. G. (2002). Attentional and memory bias for emotional information in crime victims with acute posttraumatic stress disorder (PTSD). *Anxiety Disorders*, *16*, 675–692.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.

Powers, M. B., Halpern, J. M., Ferenschak, M. P., Gillihan, S. J., & Foa, E. B. (2010). A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clinical Psychology Review*, *30*(6), 635–641.

Price, K., Veale, D., & Brewin, C. R. (2012). Intrusive imagery in people with a specific phobia of vomiting. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, 672–678. https://doi.org/10.1016/j.jbtep.2011.09.007

Rauch, S., & Foa, E. (2006). Emotional Processing Theory (EPT) and Exposure Therapy for PTSD. *Journal of Contemporary Psychotherapy*, *36*, 61–65. https://doi.org/10.1007/s10879-006-9008-y

Reynolds, M., & Brewin, C. R. (1999). Intrusive memories in depression and posttraumatic stress disorder. *Behaviour Research and Therapy*, *37*(3), 201–215.

Rizvi, S. L., Vogt, D. S., & Resick, P. A. (2009). Cognitive and affective predictors of treatment outcome in cognitive processing therapy and prolonged exposure for posttraumatic stress disorder. *Behaviour Research and Therapy*, *47*(9), 737–743.

Rubin, D. C, Berntsen, D., & Bohni, M. K. (2008). A Memory-Based Model of Posttraumatic Stress Disorder : Evaluating Basic Assumptions Underlying the PTSD Diagnosis. *Psychological Review*, *115*(4), 985–1011.

Rubin, David C, Dennis, M. F., & Beckham, J. C. (2011). Autobiographical memory for stressful events: The role of autobiographical memory in posttraumatic stress disorder. *Consciousness and Cognition*, *20*(3), 840–856. https://doi.org/10.1016/j.concog.2011.03.015

Rupp, C., Doebler, P., Ehring, T., & Vossbeck-Elsebusch, A. N. (2017). Emotional Processing Theory Put to Test: A Meta-Analysis on the Association Between Process and Outcome Measures in Exposure Therapy. *Clinical Psychology and Psychotherapy*, *24*, 697–711.

Sanislow, C. A., Pine, D. S., Quinn, K. J., Kozak, M. J., Garvey, M. A., Heinssen, R. K., … Cuthbert, B. N. (2010). Developing Constructs for Psychopathology Research: Research Domain Criteria, *119*(4), 631–639. https://doi.org/10.1037/a0020909

Siddiqui, A. P., & Unsworth, N. (2011). Investigating the role of emotion during the search process in free recall. *Memory & Cognition*, *39*, 1387–1400. https://doi.org/10.3758/s13421-011-0125-9

Speckens, A. E. M., Hackmann, A., Ehlers, A., & Cuthbert, B. (2007). Imagery special issue: Intrusive images and memories of earlier adverse events in patients with obsessive compulsive disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 411–422.

Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A Retrieved Context Model of the Emotional

Modulation of Memory. *Psychological Review*, *126*(4), 455–485.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, *88*(3), 248–257. https://doi.org/10.1037/0021-843X.88.3.248

Usher, M., & Mcclelland, J. L. (2001). The Time Course of Perceptual Choice : The Leaky , Competing Accumulator Model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037//0033-295X.108.3.550

Watkins, E. R., Mullan, E., Wingrove, J., Rimes, K., Steiner, H., Bathurst, N., … Scott, J. (2011). Rumination-focused cognitive – behavioural therapy for residual depression: Phase II randomised controlled trial. *The British Journal of Psychiatry*, *119*, 317–322.

Watkins, P., Mathews, A., Williamson, D. A., & Fuller, R. D. (1992). Mood-Congruent Memory in Depression : Emotional Priming or Elaboration ? *Journal of Abnormal Psychology*, *101*(3), 581–586. https://doi.org/10.1037/0021-843X.101.3.581

Williams, A. D., & Moulds, M. L. (2007). Cognitive avoidance of intrusive memories: Recall vantage perspective and associations with depression. *Behaviour Research and Therapy*, *45*, 1141–1153. https://doi.org/10.1016/j.brat.2006.09.005

Willner, P. (1997). Validity, reliability and utility of the chronic mild stress model of depression: A 10-year review and evaluation. *Psychopharmacology*, *134*(4), 319–329. https://doi.org/10.1007/s002130050456

World Health Organization. (2017). Depression and Other Common Mental Disorders: Global Health Estimates.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 792–804.

Table 1.

Transition probabilities regressed on Initial-Item Valence and Next-Item Valence.

		β	*SE*	*t*	LR $\chi^2(df)$
Step 0					
	Intercept	.070***	.001	52.42	
Step 1					87.16 (4) ***
	Negative Initial Item	.013***	.001	9.00	
	Positive Initial Item	.001	.001	0.96	
	Negative Next-Item	.010***	.001	6.78	
	Positive Next-Item	.006***	.001	4.01	
Step 2					1,147.9 (4) ***
	Negative Initial Item x Negative Next-Item	.057***	.002	35.14	
	Negative Initial Item x Positive Next-Item	.000	.002	-0.22	
	Positive Initial Item x Negative Next-Item	.001	.002	0.81	
	Positive Initial Item x Positive Next-Item	.021***	.002	12.77	

Note. SE: Standard errors of model coefficients for fixed effects. *LR:* Likelihood ratio. *t*-values calculated via Satterthwaite's method for linear mixed effects models. ***$p < .001$.

Table 2.

Best-fitting parameters for each model, averaged across individual-subject fits.

	CMR2		eCMR		CMR3	
	Mean	SEM	Mean	SEM	Mean	SEM
β_{enc}	.298	.008	.295	.008	.286	.007
β_{rec}	.713	.013	.717	.013	.701	.012
$\beta_{distract}$	.458	.018	.481	.020	.453	.020
β_{post}	.672	.021	.675	.025	.733	.020
β_{emot}	--	--	.369	.018	.322	.016
γ_{FC}	.799	.010	.800	.010	.813	.009
γ_{CF}	.872	.009	.845	.010	.869	.009
γ_{emot}	--	--	.394	.025	.417	.026
ϕ_S	1.413	.074	1.009	.064	.957	.072
ϕ_D	.611	.042	.747	.037	.728	.037
κ	.257	.012	.279	.013	.293	.012
η	.184	.010	.181	.010	.189	.010
λ	.118	.010	.115	.011	.125	.010
s	1.044	.041	1.049	.039	1.117	.049
ω	22.821	.383	22.702	.406	22.605	.382
α	.934	.006	.958	.006	.954	.006
c_{thresh}	.339	.018	.991	.038	.992	.038

Note. β_{enc} = drift rate for temporal context during encoding. β_{rec} = drift rate for temporal context during recall. $\beta_{distract}$ = drift rate for temporal context during a distractor item or shift between mental tasks. β_{post} = drift rate for temporal context at the end of a recall session. β_{emot} = drift rate for emotional context. γ_{FC} = learning rate for item-to-temporal context associations. γ_{CF} = learning rate for temporal context-to-item associations. γ_{emot} = learning rate for emotional context-to-item associations. ϕ_S scales the primacy effect and ϕ_D determines its rate of decay over the presentation of new items. κ = rate of evidence decay in the leaky accumulator due to the passage of time. η = width of the normal distribution of noise values for each item's recall evidence in the leaky accumulator. λ = degree of lateral inhibition between items in the leaky accumulator. *s* = scales the strength of inter-item semantic associations. ω scales the raised threshold for item repetition, and α scales the decay in that threshold as subsequent items are recalled. c_{thresh}= threshold of similarity between an item's encoding and retrieval contexts needed in order to consider the item a correct recall.

Table 3.

RMSE's for individual behavioral analyses.

	CMR2	eCMR	CMR3
Total	.064	.061	.058
Emotional Clustering	.007	.008	.006
SPC	.025	.028	.027
PFR	.059	.038	.037
Left CRP	.036	.030	.031
Right CRP	.043	.050	.053
Semantic CRP	.084	.076	.076
ELI's	.325	.322	.325
PLI's	.098	.098	.098

Note. Total: All data points. Emotional Clustering: Emotional clustering effect. SPC: Serial position curve. PFR: Probability of first recall. Left CRP: Negative lag values of the Lag-Conditional Response Probability graph. Right CRP: Positive lag values of the Lag-Conditional Response Probability graph. Semantic CRP: Semantic conditional response probability graph. ELI's: Frequency of extra-list intrusions per list. PLI's: Frequency of prior-list intrusions per list.

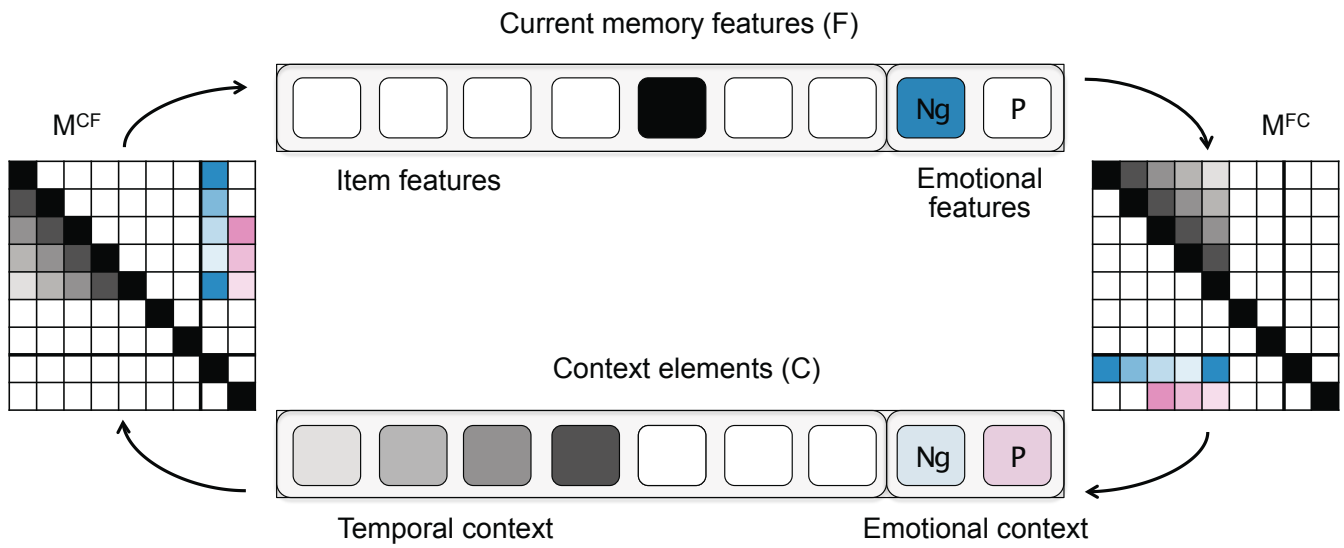

Figure 1. Schematic of how CMR3 models memory encoding and retrieval. The vector F represents event features comprising both item and emotional content. Item features are either present (have a value of 1.0, shown in black) or absent (have a value of 0.0, shown in white). Valence is encoded by setting either the Ng cell (where values are shown in blue) or the P cell (where values are shown in pink) to 1.0, indicating negative and positive valence, respectively. Values of 0.0 for both cells, shown in white, indicate neutral valence. The vector C represents context and comprises recent event features (temporal context) and previously evoked emotions (emotional context). The shaded colors represent the decay of context features, which can take any value between 0.0 and 1.0. The F and C vectors interact through two weight matrices, M^{CF} and M^{FC}, which store the strengths of associations between event features and context elements. These associations are also shown in shaded colors to indicate values between 0.0 and 1.0. Context cues memories for recall by activating context-to-item associations stored in M^{CF}. New events update context with their own features, and recalling a memory reactivates its associated context, via item-to-context associations stored in M^{FC}. See text for details.

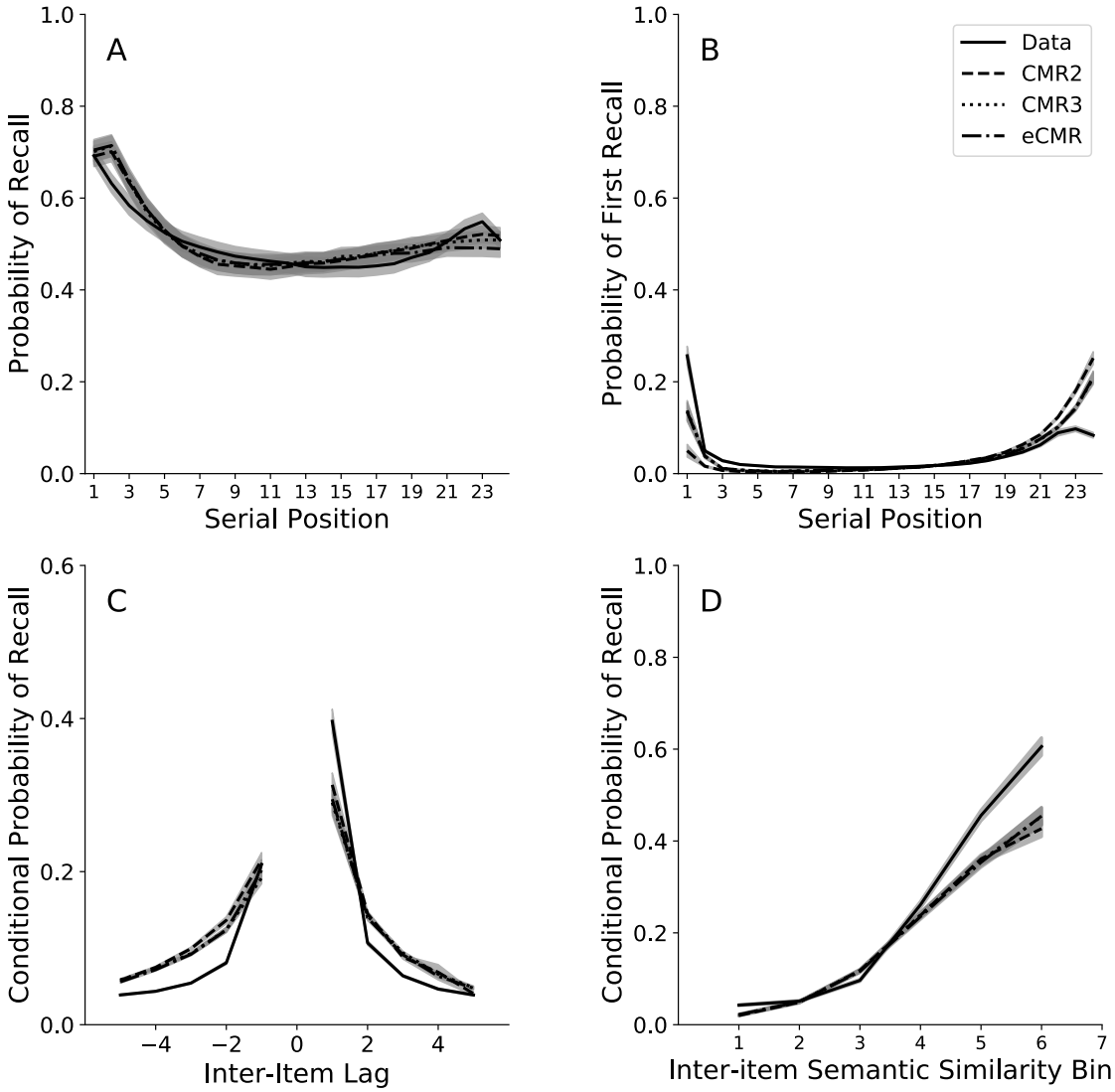

Figure 2. Modeling the organization of memories during free recall. We fit CMR2, eCMR, and CMR3 to serial position effects and to temporal, semantic, and emotional clustering in a delayed free recall task. **A.** The *serial position curve*, which plots recall probability as a function of items' presentation order during encoding, illustrates the mnemonic benefits for early and late list positions (primacy and recency). **B.** The *probability of first recall* curve plots the probability that an item will be recalled first as a function of an items' presentation order during encoding. **C.** The *lag conditional response probability* graph represents the tendency of items to cluster during recall according to their proximity to one another during encoding. For an example, the value at a lag of +2 is the probability that, given the recall of item A, the item that is recalled directly after item A was encoded 2 positions after item A during study. **D.** The *semantic conditional response probability* curve represents the tendency of items to cluster during recall according to their semantic similarity to one another. The curve depicts the probability that, given the recall of item A, the item recalled right after item A has a certain level of similarity with A, plotted against the median value of each bin.

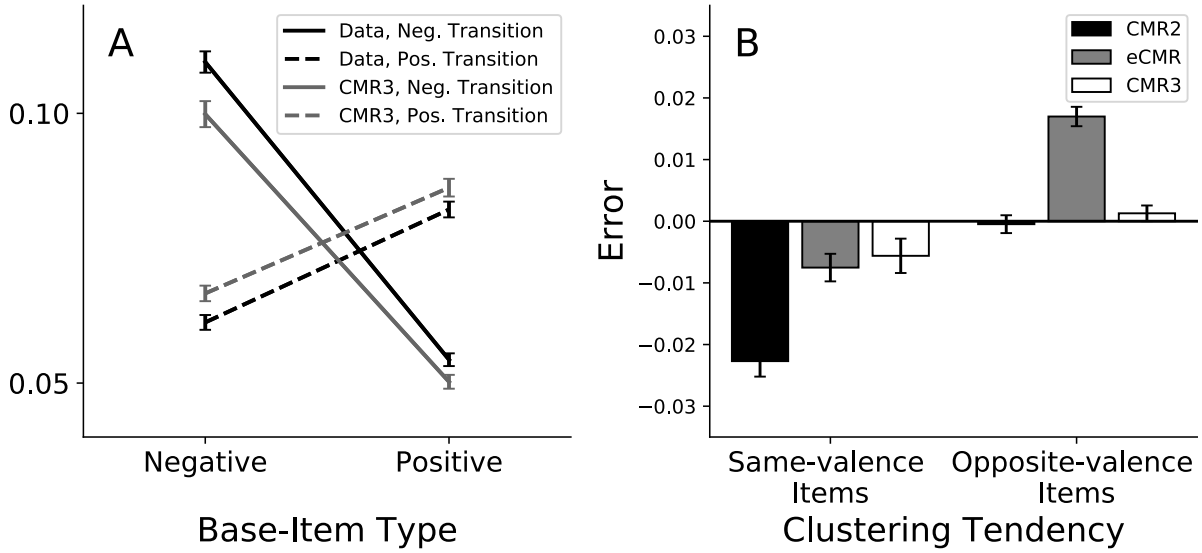

Figure 3. Modeling the organization of memories during free recall, continued. The emotional clustering effect represents the tendency of items to cluster during recall according to their emotional valence. **A.** The fit of CMR3 to the emotional clustering effect in the data. The graph plots the conditional probability of recalling an item, given that a negative item vs. a positive item has just been recalled (i.e., the "Base-item type" of the transition). Solid lines represent the tendency to transition to a negative item, and dashed lines represent the tendency to transition to a positive item. **B.** The error of model fits for CMR2 (black bars), eCMR (gray bars), and CMR3 (white bars) in predicting the tendency of same-valent items to cluster together during recall (left bars) and the tendency of opposite-valent items to not cluster together during recall (right bars).

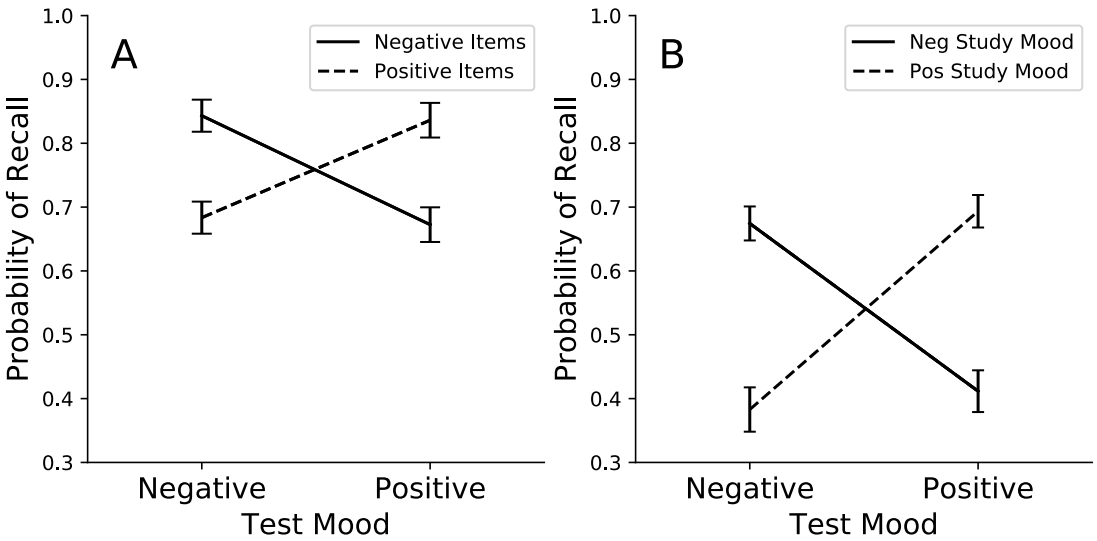

Figure 4. Simulating Mood-congruent and Emotion-state Dependent Recall. Using parameters obtained from the behavioral fits in Simulation 1, we used CMR3 to predict recall outputs based on two factors: (1) the match between an item's valence and the emotional state during retrieval (mood-congruent recall), and (2) the match between the emotional state during encoding and retrieval, regardless of the item's valence (emotion-state dependent recall). **A.** Mean probability of recalling a negative item (solid lines) or a positive item (dashed lines), depending on the valence of the recall context, which is marked on the horizontal axis. **B.** Mean probability of recalling a neutral item depending on whether it was encoded and retrieved in matching versus non-matching emotional states. The valence of emotion during retrieval, or *Test mood,* is marked on the horizontal axis. Solid lines represent negative mood during study, and dashed lines represent positive mood during study. Error bars represent +/- the standard error of the mean, calculated across all parameter sets' predictions.

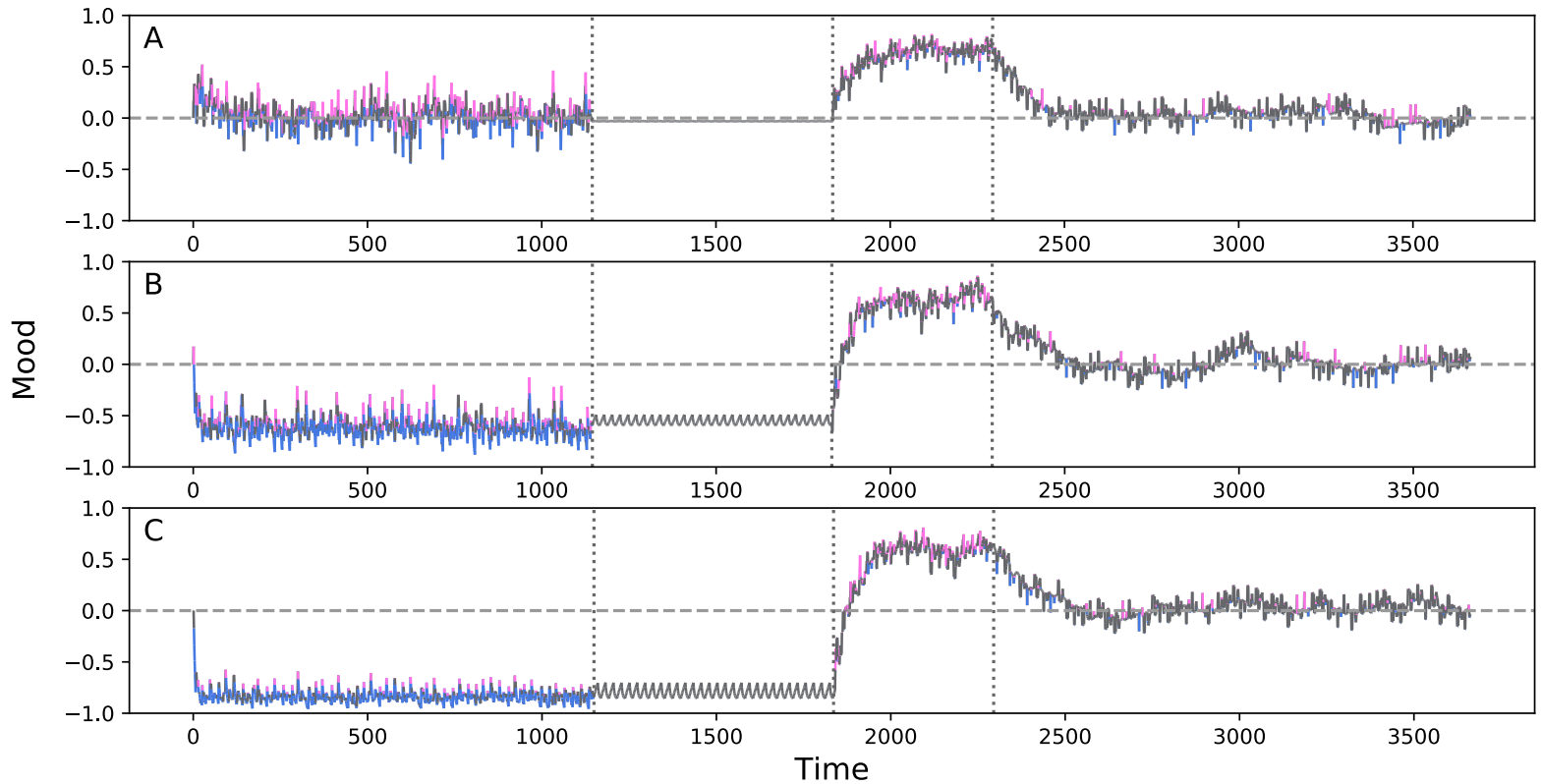

Figure 5: *Simulating the effects of stressor frequency.* Each panel illustrates the evolution of model-simulated mood across four periods, marked by vertical lines: (1) A developmental period, in which a virtual subject experienced a mixture of both emotional and neutral events, (2) A neutral period, in which a virtual subject experienced solely unique, neutral events, (3) A simulated course of behavioral activation therapy, in which a virtual subject experienced a heightened rate of positive events (30% positive, 60% neutral, 10% negative), and finally (4) A recovery period comprised of predominately neutral events (80% neutral, 10% negative, 10% positive). The beginning of each period is marked by a vertical dotted line. Across panels A-C we varied the ratio of positive to neutral events during the developmental period, with 30% negative and 30% positive in panel A, 60% negative and 20% positive in panel B, and 70% negative and 10% positive in panel C. The height on the graph represents the level of model-simulated mood (minimum of -1.0, maximum of 1.0), and the color on the graph represents the emotional valence of the encoded or retrieved event that evoked that level of mood (pink = positive, blue = negative, gray = neutral).

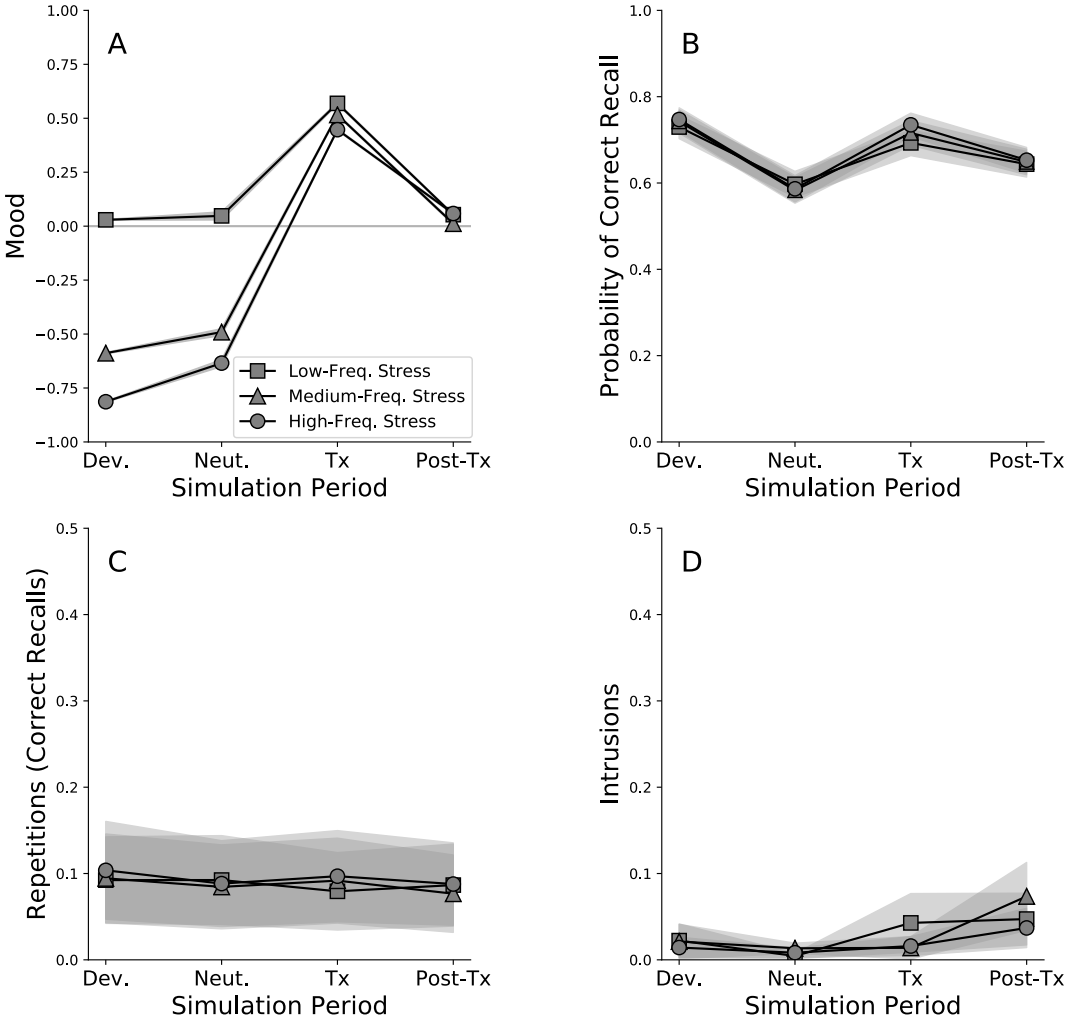

Figure 6. Aggregated results for the simulation of stressor frequency effects. This figure presents the model-predicted patterns of mood and recall for Simulation 4, aggregated across the results for all virtual subjects (i.e., parameter sets). In the first condition (square markers), each virtual subject experienced an even mixture of negative and positive events during the developmental period (30% negative, 30% positive events, 1:1 ratio). In the second condition (triangle markers), the virtual subject experienced a moderate frequency of stressors during the developmental period (60% negative, 20% positive events, 3:1 ratio). In the third condition (circle markers), the virtual subject experienced a high frequency of stressors during the developmental period (70% negative, 10% positive events, 7:1 ratio). After the developmental period, all virtual subjects experienced a neutral period (100% neutral events), followed by a "treatment period" containing a simulation of behavioral activation therapy (10% negative, 30% positive events). Finally, the simulation concluded with a post-treatument period (10% negative, 10% positive events). **A.** Virtual subjects' mood at each stage of the simulation, averaged within each stage. **B.** The probability of correct recall at each stage of the simulation, averaged within each stage. **C.** The number of repetitions per list at each stage of the simulation, averaged within each stage. **D.** The number of intrusions per list at each stage of the simulation, averaged within each stage. Error regions represent +/- the standard error of the mean, calculated across all parameter sets' predictions. Dev. = Developmental period. Neut. = Neutral period. Tx = Treatment period. Post-Tx = Post-treatment period.

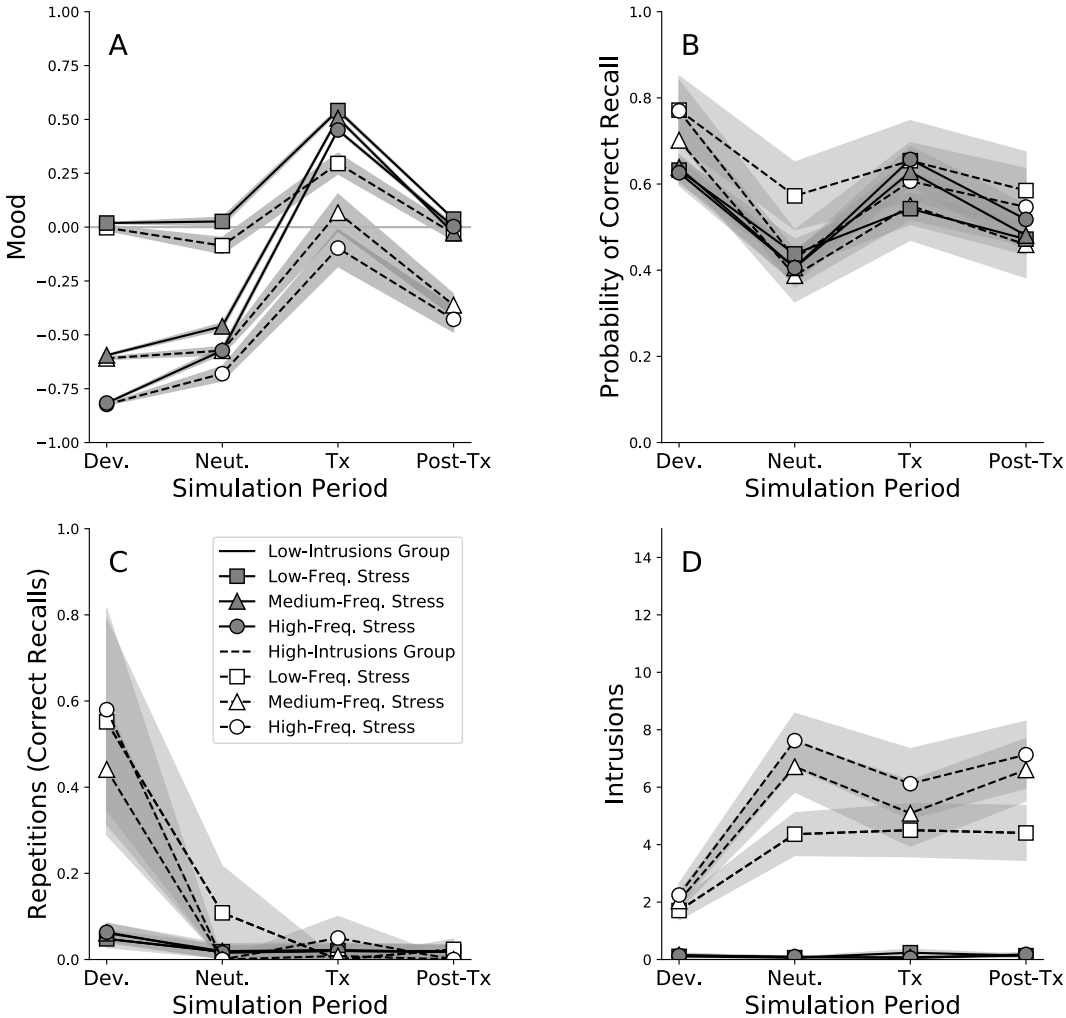

Figure 7. Aggregated results for the simulation of stressor repetition effects. This figure presents the model-predicted patterns of mood and recall for Simulation 5, aggregated across the results for all virtual subjects (i.e., parameter sets). In the first condition (square markers), each virtual subject experienced an even mixture of negative and positive events during the developmental period (30% negative, 30% positive events, 1:1 ratio). In the second condition (triangle markers), the virtual subject experienced a moderate frequency of stressors during the developmental period (60% negative, 20% positive events, 3:1 ratio). In the third condition (circle markers), the virtual subject experienced a high frequency of stressors during the developmental period (70% negative, 10% positive events, 7:1 ratio). In each condition, one negative event was selected to repeat every 10 lists. After the developmental period, all virtual subjects experienced a neutral period (100% neutral events), followed by a "treatment period" containing a simulation of behavioral activation therapy (10% negative, 30% positive events). Finally, the simulation concluded with a post-treatment period (10% negative, 10% positive events). **A.** Virtual subjects' mood, averaged within each stage of the simulation. **B.** The probability of correct recall at each stage of the simulation. **C.** The number of repetitions per list, averaged within each stage. **D.** The number of intrusions per list, averaged within each stage of the simulation. Solid lines show virtual subjects with low counts of intrusions (Low-Intrusions Group) and dashed lines show virtual subjects with high counts of intrusions (High-Intrusions Group). Error regions represent +/- the standard error of the mean, calculated across all parameter sets' predictions. Dev. = Developmental period. Neut. = Neutral period. Tx = Treatment period. Post-Tx = Post-treatment period.

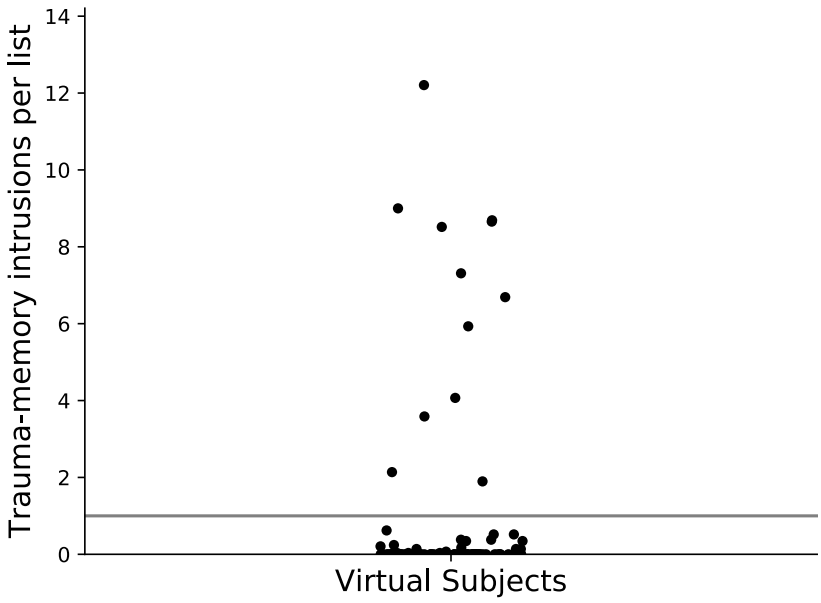

Figure 8. Rate of trauma-memory intrusions per list for each virtual subject. The modal virtual subject had fewer than 1.0 such intrusions per list during the period of neutral events following their traumatic event (Low-Intrusions Group), whereas 12 virtual subjects had more than 1.0 such intrusions per list during the period of neutral events following their traumatic event (High-Intrusions Group). The horizontal line is a reference line, height equal to 1.0.

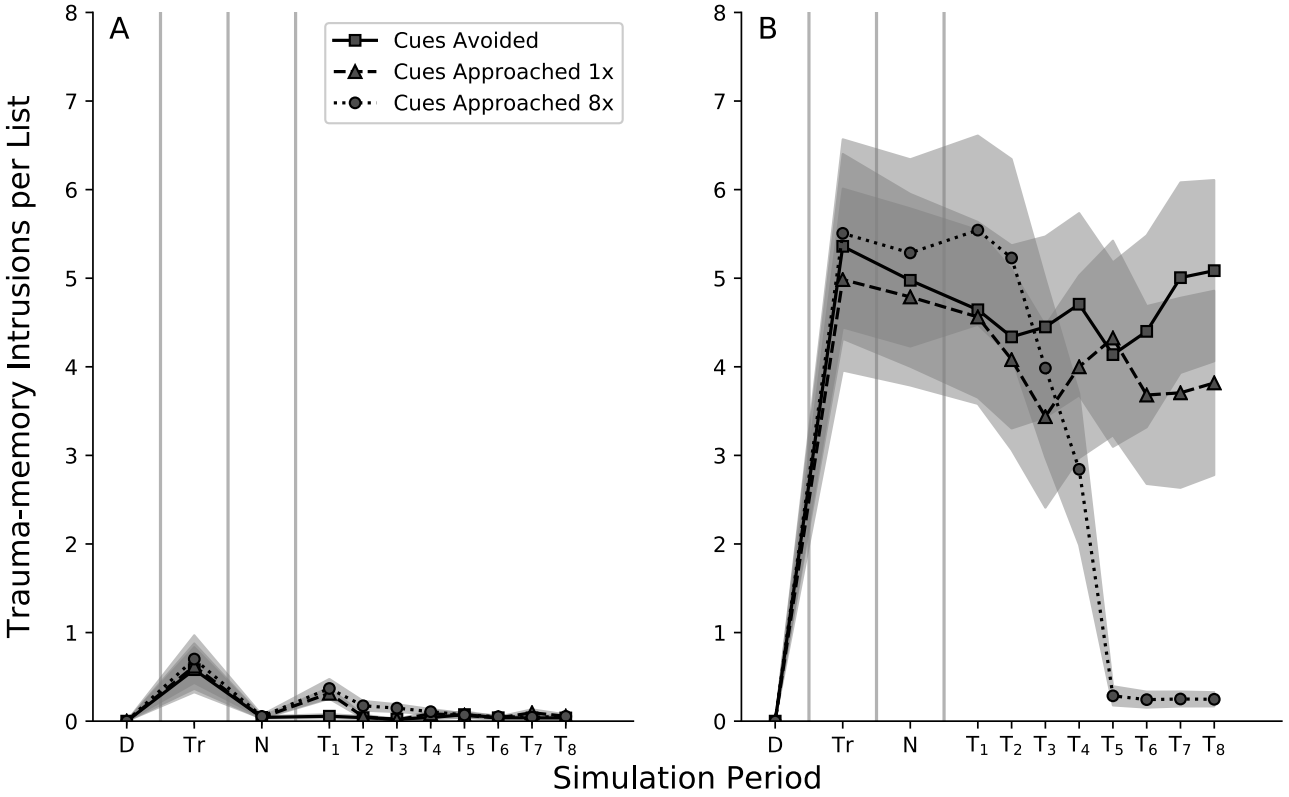

Figure 9. Aggregated results for simulating the effects of temporal context on trauma-memory intrusions. This figure presents the model-predicted patterns of trauma-memory intrusions resulting from encoding a traumatic event, aggregated across the results for all virtual subjects (i.e., parameter sets). Panel A shows results for the "Low-Intrusions Group," or virtual subjects who had low rates of trauma-memory intrusions after their traumatic event. Panel B shows results for the "High-Intrusions Group," or virtual subjects who had high rates of trauma-memory intrusions after their traumatic event. The markers represent one of three conditions. Condition 1 (square markers) is a baseline condition, in which the neutral period consists solely of unique, neutral events. Condition 2 (triangle markers) is a single exposure to a set of three trauma cues, or the nonnegative stimuli that preceded the traumatic event during encoding the neutral period. Condition 3 (circle markers) is eight exposures to the same set of three, nonnegative trauma cues. D = Developmental Period. Tr = The portion of the developmental period after the traumatic event occurs. N = The neutral period prior to the re-encounters with trauma cues. $T_{1\text{-}8}$ = each of the eight time-points at which a virtual subject has an opportunity to either approach or avoid trauma cues.

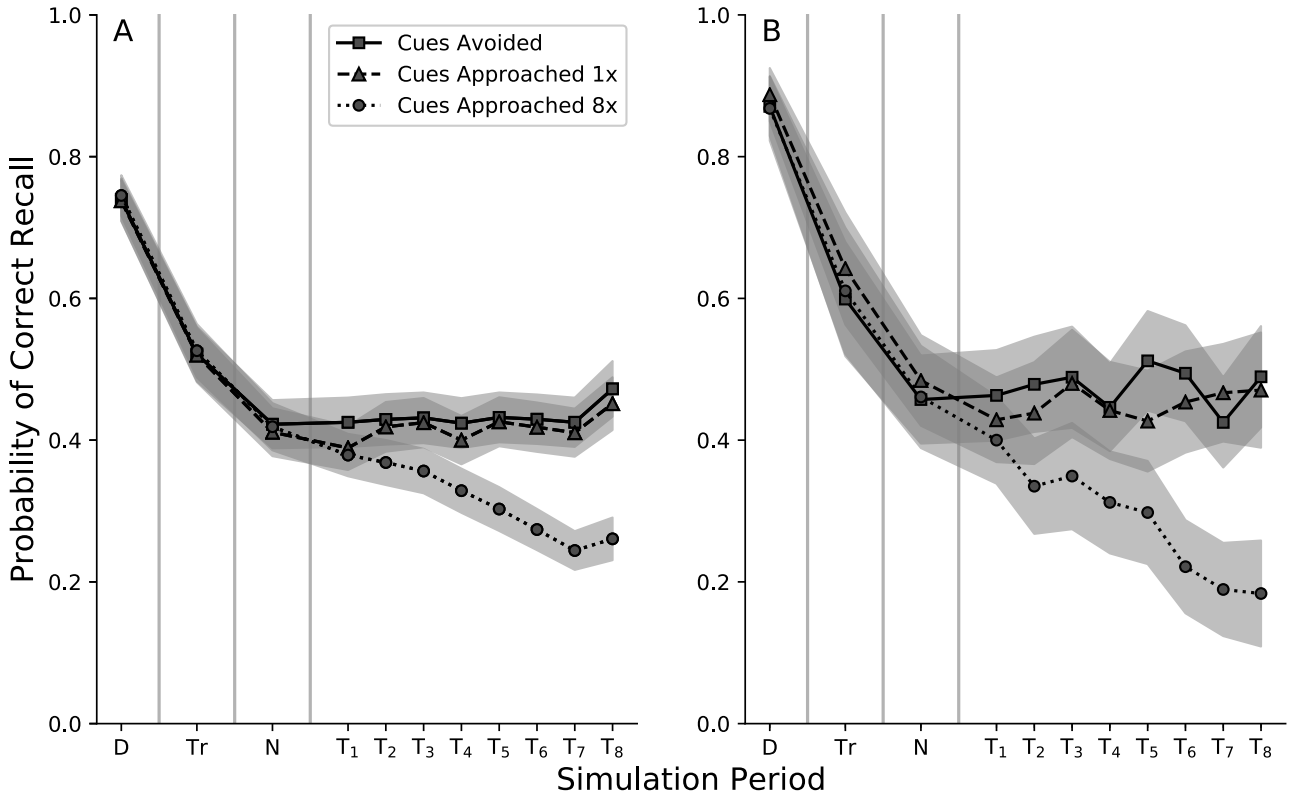

Figure 10. Aggregated results for simulating the effects of temporal context on correct recall. This figure presents the model-predicted patterns of mood resulting from encoding a traumatic event, aggregated across the results for all virtual subjects (i.e., parameter sets). Panel A shows results for the "Low-Intrusions Group," or virtual subjects who had low rates of trauma-memory intrusions after their traumatic event. Panel B shows results for the "High-Intrusions Group," or virtual subjects who had high rates of trauma-memory intrusions after their traumatic event. The markers represent one of three conditions. Condition 1 is a baseline condition (square markers), in which the neutral period consists solely of unique, neutral events. Condition 2 is a single exposure to a set of three trauma cues, or the nonnegative stimuli that preceded the traumatic event during encoding the neutral period (triangle markers). Condition 3 is eight exposures to the same set of three, nonnegative trauma cues (circle markers). D = Developmental Period. Tr = The part of the developmental period after the traumatic event occurs. N = The neutral period prior to the re-encounters with trauma cues. $T_{1\text{-}8}$ = each of the eight time-points at which a virtual subject has an opportunity to either approach or avoid trauma cues.

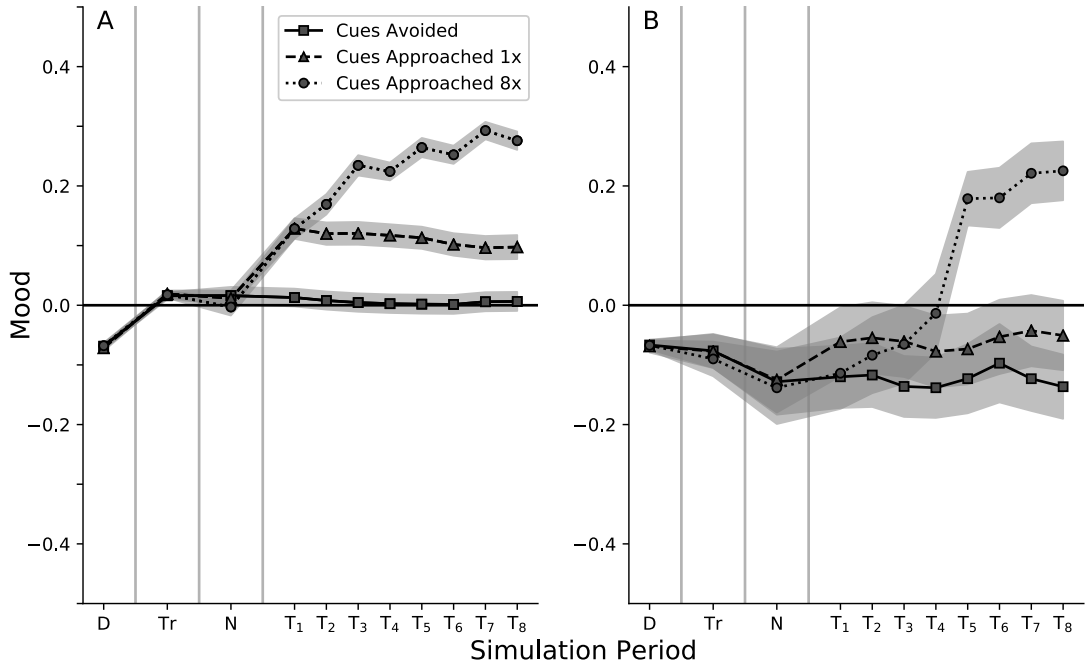

Figure 11. Aggregated results for simulating the effects of temporal context on mood. This figure presents the model-predicted patterns of mood resulting from encoding a traumatic event, aggregated across the results for all virtual subjects (i.e., parameter sets). Panel A shows results for the "Low-Intrusions Group," or virtual subjects who had low rates of trauma-memory intrusions after their traumatic event. Panel B shows results for the "High-Intrusions Group," or virtual subjects who had high rates of trauma-memory intrusions after their traumatic event. The markers represent one of three conditions. Condition 1 (square markers) is a baseline condition, in which the neutral period consists solely of unique, neutral events. Condition 2 (triangle markers) is a single exposure to a set of three trauma cues, or the nonnegative stimuli that preceded the traumatic event during encoding the neutral period. Condition 3 (circle markers) is eight exposures to the same set of three, nonnegative trauma cues. D = Developmental Period. Tr = The part of the developmental period after the traumatic event occurs. N = The neutral period prior to the re-encounters with trauma cues. $T_{1\text{-}8}$ = each of the eight time-points at which a virtual subject has an opportunity to either approach or avoid trauma cues.

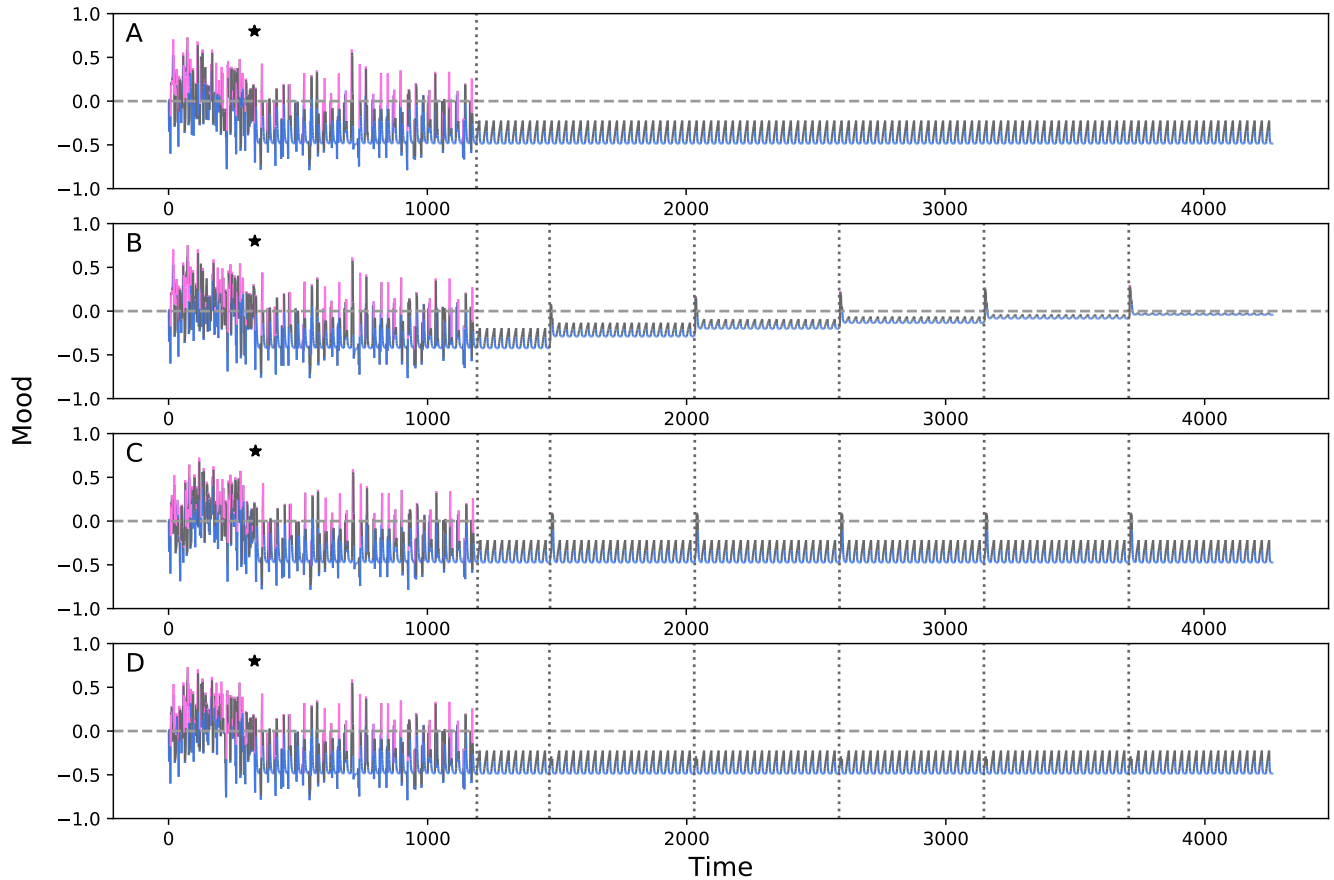

Figure 12: *Simulating the effects of trauma and imaginal exposure therapy on intrusive memories.* Each panel illustrates the evolution of model-simulated mood across two periods: (1) A developmental period, in which a virtual subject experienced an even mixture of negative and positive life events (30% negative, 30% positive), with one high-arousal negative event ($\phi_{emot} = 50.0$) at the location of the asterisk (*), followed by (2) A neutral period, in which a virtual subject experienced solely unique, neutral events (100% neutral). In panel A, the neutral period consisted solely of new, neutral events with no intervention (Baseline). In panel B, the neutral period simulated 5 sessions of imaginal exposure therapy: a less-negative context was evoked by introducing a positive event every 20 lists, this time followed immediately by a re-presentation of the trauma-event as an encoding event (Full treatment)**.** In panel C, the neutral period introduced 5 positive events every 20 lists, marked by vertical dotted lines (Positive-Context Only). In panel D, the neutral period reintroduced the high-arousal negative event 5 times, every 20 lists, marked by vertical dotted lines, without any change in emotional context prior to reencoding (Reencoding Only). The height on the graph represents the level of model-simulated mood (minimum of -1.0, maximum of 1.0). The variability of model-simulated mood in the neutral period reflects the degree of negative emotion evoked by each memory intrusion. The color on the graph represents the emotional valence of the encoded or retrieved event that evoked that level of mood (pink = positive, blue = negative, gray = neutral).

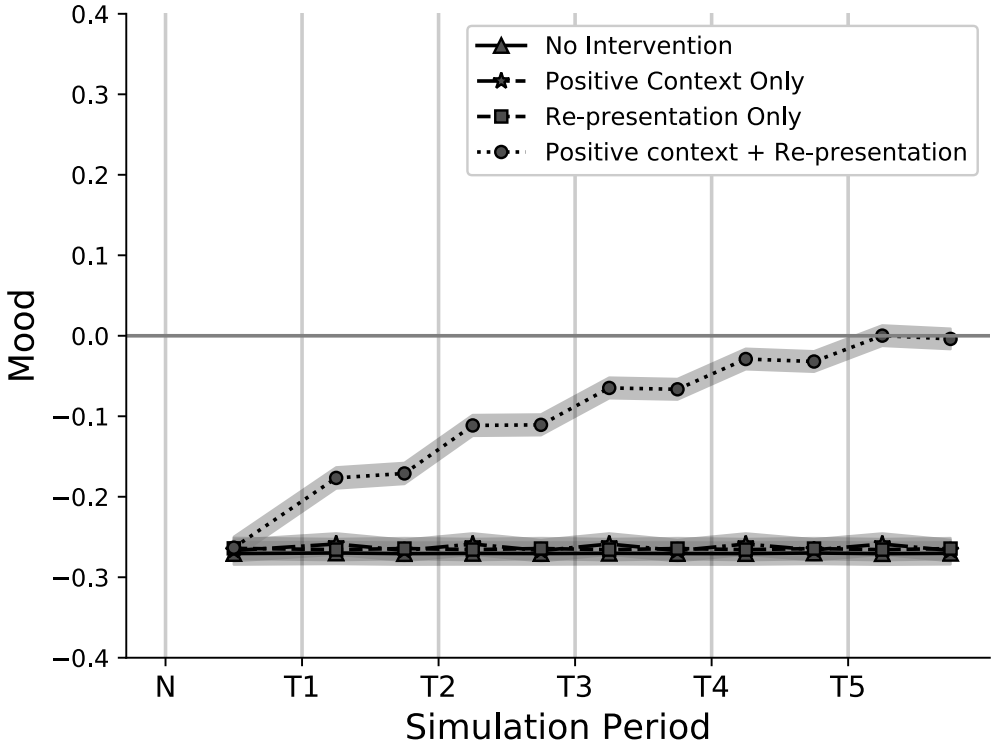

Figure 13. Aggregated results for the simulation of trauma exposure and imaginal exposure therapy. This figure presents the model-predicted patterns of mood for each of four conditions, aggregated across the results for all virtual subjects (i.e., parameter sets). Each virtual subject experienced a developmental period, consisting of an even mixture of negative and positive life events (30% negative, 30% positive), with one high-arousal negative event ($\phi_{emot} = 50.0$), and (2) A neutral period, which consisted solely of unique, neutral events (100% neutral). In Condition 1 (triangle markers), a baseline condition, the neutral period consists solely of unique, neutral events. In Condition 2 (star markers), the neutral period additionally introduces 5 positive events every 20 lists. In Condition 3 (square markers), the neutral period instead introduces 5 re-encodings of the traumatic memory every 20 lists, with no changes to the surrounding emotional context. In Condition 4 (circle markers), the neutral period simulates 5 sessions of imaginal exposure therapy: less-negative context is again evoked by introducing a positive event every 20 lists, this time followed by a re-presentation of the traumatic event for re-encoding**.** The beginning and end of each simulated treatment session are marked by vertical lines. Between these lines, the first data point represents the mean level of mood present during the first half of the session, whereas the second data point represents the mean level of mood present during the second half of the session. N = The neutral period after the traumatic event has occurred. T1-T5 = each of the five timepoints at which a virtual subject receives the given intervention for that condition.