# A memory-based theory of emotional disorders

Rivka T. Cohen

University of Pennsylvania

Michael J. Kahana

University of Pennsylvania

## Author Note

National Institutes of Health grant MH55687 to MJK provided support for this research. We thank Ayelet Ruscio, Lily Brown, Sudeep Bhatia, Nora Herweg, Dilara Berkay, Christoph Weidemann, Dan Schonhaut, Ethan Solomon, Paul Wanda, David Halpern, Nick Diamond, Noa Hertz, Alice Healy, Katherine Hurley, Jessie Pazdera, Effie Li, Nicole Miller, Anne Park, and Michael Barnett for helpful conversations and comments. Portions of this work were presented at the Context and Episodic Memory Symposium in 2018 and the Society for Mathematical Psychology in 2019. Correspondence may be addressed to either Rivka T. Cohen (rivkac@sas.upenn.edu) or Michael J. Kahana (kahana@psych.upenn.edu). Data and code are publicly available on the Computational Memory Lab website: http://memory.psych.upenn.edu/Publications

Abstract

Learning and memory play a central role in emotional disorders, 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

A memory-based theory of emotional disorders

## Introduction

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). 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 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). 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 a diagnostic criterion and 50-75% of patients have co-occurring MDD (K. T. 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. Here we develop a transdiagnostic, computational model of memory in emotional disorders. This model describes how distressing events interact with the memory system, as well as the mechanisms that produce the negative mood and intrusive memories that cut across both depression and PTSD.

Research implicates learning and memory as transdiagnostic processes underlying emotional disorders (Beck, Rush, Shaw, & Emery, 1979; Bower, 1981; Brewin, 2006). Patients with PTSD experience intrusive memories of traumatic events (Ehlers, Hackmann,

& Michael, 2004), but patients with depression, social anxiety, and other emotional disorders also report high rates of vivid, intrusive memories of painful events (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). In addition, patients with anxiety and depression tend to recall negative events more frequently than positive events, implicating mechanisms of *mood-congruent recall* (Matt, Vazquez, & Campbell, 1992; P. C. Watkins, Mathews, Williamson, & Fuller, 1992). Here, mood-congruent recall refers to better recall for memories whose emotional valence matches the current emotional context. Further, during anxiety treatment, patients benefit most when therapeutic interventions take place in an emotional context that will match subsequent retrieval contexts, suggesting processes of *emotion-state dependent recall* (M. Craske et al., 2008).

Despite these findings, current theories of learning and memory in emotional disorders do not describe whether and how mood-congruent recall, emotion-state dependent recall, and intrusive memories of painful events influence one another (Brewin, 2014; Rubin, Berntsen, & Bohni, 2008; Brown, Zandberg, & Foa, 2019). Leading theories of memory in PTSD and other emotional disorders have proposed that the arousal and valence of an experience influence its tendency to result in intrusive memories (Brewin, 2014; Rubin et al., 2008; Rubin, Dennis, & Beckham, 2011). Valence refers to the attractive (“positive") or repellent (“negative") properties of a stimulus, and arousal refers to the emotional intensity or degree of activation of the sympathetic nervous system due to a stimulus (Barrett, 1998; Russell, 1980; M. M. Bradley, Greenwald, Petry, & Lang, 1992; T. F. Brady, Konkle, Alvarez, & Oliva, 2008), in human memory and in emotional disorders.

However, these theories differ regarding whether high negative arousal during encoding results in intrusive memories by impairing (Brewin, 2014) or enhancing (Rubin et al., 2008, 2011) a new memory’s association to its encoding context. Further, these

approaches disagree on whether it is specifically *negatively-valent* emotional arousal that results in intrusive memories, by disrupting memory-context associations during encoding (Brewin, 2014), or whether it is the intensity of emotional arousal that leads to intrusive memories, regardless of the event's negative or positive properties, in addition to factors like the frequency of rehearsal and centrality to the person's life story (Rubin et al., 2008, 2011). Resolving these questions is vital to improved understanding of how painful memories operate in and contribute to emotional disorders, and thereby, for generating new treatment avenues.

Computational models can help answer these questions. By building a model of the human memory system, we can test how our theory of human memory behaves under different simulated circumstances. If our model can produce the behavior we observe in the real world, then this contributes evidence in support of the theory. However, if our model cannot produce patterns that match real-world behavior, this sheds light on whether, and how, we need to change the theory. Computational modeling is especially useful when developing a theory of unobservable processes, such as what cognitive operations could generate the memory patterns that human subjects and researchers *can* observe. For example, memory researchers used computational models to develop the theory that associations between events and their contexts guide recall in human *episodic memory*, or memory for personally experienced events (Howard & Kahana, 2002a; Lohnas, Polyn, & Kahana, 2015; Healey & Kahana, 2016; Polyn, Norman, & Kahana, 2009; Talmi, Lohnas, & Daw, 2019; Moscovitch, Cabeza, Winocur, & Nadel, 2016). In the resulting *retrieved-context theory* (Howard & Kahana, 2002a; Lohnas et al., 2015; Healey & Kahana, 2014; Polyn et al., 2009; Talmi et al., 2019), the brain creates networks of associations between the contents of a memory and its context, or the internal or external surroundings (the time or place, the person's emotional or physiological state) that were present just before or while the event took place. These contextual cues then reactivate their associated memories, and in turn, recalling a memory reactivates a mental representation of its

associated contextual elements.

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 the mutual influence of memory and mood in emotional disorders. We then use this model to simulate the human memory system and test whether our updated theory of emotion in episodic memory can account for the persistent negative mood and intrusive memories observed in clinical settings.

## From list events to life events

We sought to develop a model of memory for life events grounded in the established scientific literature on memory for list events. Early memory scientists embraced the list-learning technique as it afforded precise experimental control over the conditions prevailing during encoding and retrieval, as well as item properties and their relations (Crowder, 1976; Kahana, 2012). Free recall of word lists constitutes one of the most widely studied list-learning paradigms — one that has generated textbook phenomena and fueled the development of major theories. In each trial, subjects study a list of sequentially presented words which they subsequently attempt to recall after a delay. This task earns the moniker "free" as the experimenter neither supplies specific retrieval cues nor imposes any requirement as to the order of recall. Because subjects must recall a specific list of items, they must direct their memory search to the items encoded in the target list-context. This context includes both the time-window in which the item occurred (temporal context) and the semantic properties shared by the items in that particular list. Each list item constitutes a mini-event, or episode, in the subject's life, and the subject's task is to search memory for the episodes occurring in a particular temporal context (the list). Tulving (1972) referred to recall of such episodes, bound to a specific temporal

context, as *episodic memory*.

We suggest that the free recall task can serve as a building block in a model of autobiographical memory. In the real world, a complex event consists of a list of stimuli, encoded against the backdrop of other recent internal and external stimuli. We often reflect back on our experiences, either recalling them to ourselves, or to our friends. Such reminiscence is akin to free recall, and recording of these reminiscences in memory will recursively shape the way we encode and retrieve related experiences in the future. Occasionally, memories arise spontaneously, or reflecting on a given event will conjure up memories of other related events. In a free recall task, a subject will sometimes remember an item that had not occurred on a target list. Such memory intrusions, and their mnemonic sequelae, will figure prominently in our modeling of non-voluntary, unwanted memories in PTSD.

To develop our model, which we call CMR3, we used a list-learning paradigm. The list-learning paradigm is especially useful for modeling intrusions due to its ability to capture two apparently distinct processes that both share the episodic memory system: voluntary memory (strategic, in response to a targeted memory search) and involuntary memory (spontaneous and unintended). Some theorists have proposed that each type of memory should take place through distinct mechanisms; however, it is increasingly understood that both voluntary (strategic) and involuntary autobiographical memories operate through the same episodic memory system (Berntsen, 2010). Specifically, whereas people may generate additional mental context cues to guide the intended retrieval of memories, their voluntary recall is still determined by their current context. Autobiographical memories that may be experienced as “spontaneous” are also reactivated by the same process of encountering an associated spatial, temporal, emotional or other related contextual cue. Accordingly, when patients with PTSD report experiencing apparently-spontaneous memory intrusions, typically further investigation reveals the presence of one or more associated temporal, spatial, perceptual or other context cues

associated to the trauma that cued the memory (Ehlers & Clark, 2000).

The externalized free recall paradigm is well suited for modeling intrusive involuntary recall because it operationalizes how both voluntary (strategic) and involuntary (spontaneous) recall outputs occur when the episodic memory system is engaged. During an externalized free recall task, subjects are instructed to report all words that enter their mind, regardless of whether or not they are trying to recall the words. In addition to the recall outputs that they report during a typical delayed free recall task, here subjects also report all other spontaneous recalls that occur during retrieval, even when subjects are aware that this spontaneous background noise is not from the intended target context for recall. Thus, intrusions are actually the result of a spontaneous, unintended process that arises as a byproduct during the subject's intentional retrieval search. In this paradigm, intrusions occur as an unintended and undesired process that subjects must try to censor as they maintain their focus on their desired mental content (i.e., the correct results of the recall search). The two main types consist of prior-list intrusions (PLI's), which are items the subject encountered in a prior list and does not wish to recall, and extra-list intrusions (ELI's), which are items the subject never encountered in the current laboratory session, but has encountered previously in their life prior to the lab, and which they also do not wish to recall. In this way, ELI's can also be conceptualized as a type of prior-list intrusion, just stemming from a prior "list" that took place prior to the laboratory session.

The list-learning paradigm is deceptively simple. Yet, it has generated a wealth of classic findings about human memory, including the primacy and recency effects (Murdock, 1962), temporal contiguity (Kahana, 1996) and semantic clustering effects (Romney, Brewer, & Batchelder, 1993; Bousfield, 1953). After their discovery in list-learning tasks, researchers have observed and replicated these classic findings in human memory for real-world events (Jansari & Parkin, 1996; Loftus & Fathi, 1985; Moreton & Ward, 2010; Uitvlugt & Healey, 2019; Healey, Long, & Kahana, 2019). As with any model, the goal of list-learning tasks is not to perfectly replicate the complexity of real-world events, but

rather, to create a tractable version, or representation, of such events that draws upon the same processes that occur in memory for real-world events. This enables the experimenters to begin the process of model design.

**Retrieved-context and episodic memory**

In retrieved-context models, the context that is encoded in association with a memory guides the later activation of that memory and determines whether it will be consciously retrieved (Healey & Kahana, 2016; Howard & Kahana, 2002a; Lohnas et al., 2015; Polyn et al., 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.

Talmi et al. (2019) extended earlier retrieved-context models to account for the influence of arousal on recall performance. Their eCMR model conceptualized emotion as having a single dimension, the presence or absence of emotional arousal. In eCMR, the presence or absence of emotion is a component of memories and their contexts, and the presence of arousal 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 as 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**

Here we propose a retrieved-context model of emotion and memory (CMR3) that considers the role of intrusions in the dynamics and persistence of affective states across a lifetime of simulated memories. 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 negative events on negative mood (Simulation 4) and clarify the model's predictions that repeated negative events will have a greater likelihood of becoming activated as intrusive, involuntary memories due to being associated with a wider variety of cueing contexts (Simulation 5). Then, we show the model's ability to capture the effectiveness of positive event scheduling, a core component of behavioral activation therapy for depression (Simulation 5, Treatment and Recovery Periods). We demonstrate CMR3's ability to predict that high arousal during encoding will lead to the development of intrusive memories (Simulation 6), as well as heightened nowness and vividness of such memories (Simulation 6), and then show CMR3's ability to capture the efficacy of prolonged exposure therapy for alleviating intrusive memories of high-arousal negative events (Simulation 7). We conclude with a discussion of how retrieved-context theory generates novel predictions and relates to current theories of memory in emotional disorders.

## 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. In addition, context includes a mental representation of the perceptual features of an event, which are integrated into a composite with other contextual features. Thus, the CMR family of models supports both recall stemming from the direct perceptual features that comprised an event, such as seeing the face of someone who was present during the event, as well as other perceptual and contextual features that become cues of that memory by virtue of their shared contexts. 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 new mechanisms that improve on prior models as follows. 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 (spontaneous, non-voluntary) 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 $\mathbf{f}_t$ 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 $\mathbf{c}_t$). 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 of operationalizing intrusive, non-voluntary memories, a symptom of great interest in PTSD and other emotional disorders. Within the list-learning paradigm described above, memory-intrusions are modeled as instances of recalling an item from an undesired, non-target list context (Zaromb et al., 2006; Lohnas et al., 2015). A model that resets the memory system at the end of each individual list cannot differentiate between target (intended) and non-target (unintended, undesired) recall contexts, and thus cannot model recall of a memory-intrusion from a non-target recall context. We call this updated model CMR3 and present a mathematical description below.

**Item Representation.** In CMR3, each element of $\mathbf{f}_t$ represents whether that item is present or absent in the current memory episode, and each element of $\mathbf{c}_t$ represents the extent to which prior memories or cognitive states are still active during encoding. In addition to the representations of item features in $\mathbf{f}_t$ and temporal context in $\mathbf{c}_t$, 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 $\mathbf{f}_t$, one cell indicates whether the item is negative, the second whether it is positive, and neutral items have no content in either cell. In $\mathbf{c}_t$, 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; P. J. 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 by the presentation of 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 weight matrices, as is the case during the presentation of true items. The Hebbian weight matrices store the associations between item features and context features that form during the experiment (see *Encoding* section, below). 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:

$$\mathbf{c}_{t+1} = \rho \mathbf{c}_t + \beta \mathbf{c}^{IN} \tag{1}$$

As such, context becomes a recency-weighted sum of past context states. The features from $\mathbf{f}_t$ that will enter into the new context, represented by the term $\mathbf{c}^{IN}$, are determined by multiplying the weight matrix of item-to-context associations, which we call $M^{FC}$, by $\mathbf{f}_t$, the vector of current features, and then norming the resulting vector, such that $\mathbf{c}^{IN} = \frac{M^{FC}\mathbf{f}_t}{\|M^{FC}\mathbf{f}_t\|}$. In equation 1, $\rho$ scales the magnitude of prior context so that $\mathbf{c}$ does not

grow without bounds (see Howard & Kahana, 2002a). The $\beta$ 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 ($\beta_{enc}$), retrieving a new memory ($\beta_{rec}$), registering a shift between list contexts ($\beta_{post}$), or experiencing a context shift due to a distractor task ($\beta_{distract}$). Emotional context is stored in the source region of the context vector, and it updates according to the same equation that governs updating for temporal context, but at the rate of $\beta_{emot}$. The parameter $\beta_{emot}$ is the parameter termed $\beta_{source}$ in Talmi et al. (2019), renamed here to emphasize that it governs the updating of emotional source context.

**Emotional context and the calculation of "Mood".** As described above, CMR3 represents emotional context in the source region of the context vector. As in item features (see subsection "Item representation," above), one cell accumulates negative emotional context, and one cell accumulates positive emotional context. The model updates both types of emotional valence at the same rate of $\beta_{emot}$, but the values in each cell can vary independently of one another, depending on both the emotional valence of new events and on the emotional context reactivated by recalling a memory. Thus, the model produces independent indices and separate activations of positive and negative emotion. However, to simplify graphical representation of the model's predicted emotional responding in our simulations, we created a composite measure of "mood." We calculate "mood" as the difference between the degree of positive context and negative context that is present on a given timestep ($mood_t = c_{pos,t} - c_{neg,t}$. We used a simple difference metric due to keep values on an intuitive scale between -1 (completely negative emotion) and 1 (completely positive emotion), with 0 representing equivalent levels of positive and negative emotional responding.

**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, $\mathbf{f}_t$ and $\mathbf{c}_t$, 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 $\mathbf{f}_t$ and $\mathbf{c}_t$ 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, $\mathbf{f}_{items}$, and temporal (or non-emotional) context elements, $\mathbf{c}_{temp}$. 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, $\mathbf{c}_{emot}$, as below:

$$M^{CF} = \begin{pmatrix} \mathbf{f}_{items}\mathbf{c}_{temp}^{\top} & \mathbf{f}_{items}\mathbf{c}_{emot}^{\top} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{2}$$

$$M^{FC} = \begin{pmatrix} \mathbf{c}_{temp}\mathbf{f}_{items}^{\top} & \mathbf{0} \\ \mathbf{c}_{emot}\mathbf{f}_{items}^{\top} & \mathbf{0} \end{pmatrix} \tag{3}$$

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 $(1 - \gamma_{CF})$ and $M^{CF}$ by $(1 - \gamma_{FC})$, respectively. The parameters $\gamma_{CF}$ and $\gamma_{FC}$ 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 (Word2Vec; Mikolov, 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:

$$\Delta M^{CF} = \mathbf{f}_t \mathbf{c}_t^\top \tag{4}$$

$$\Delta M^{FC} = \mathbf{c}_t \mathbf{f}_t^\top \tag{5}$$

The rate at which these associations integrate into the $M^{FC}$ and $M^{CF}$ weight matrices is determined by three parameters: $\gamma_{FC}$, $\gamma_{CF}$, and $\gamma_{emot}$, which is the parameter termed $L_{sw}^{CF}$ in Talmi et al. 2019. The parameters $\gamma_{CF}$ and $\gamma_{FC}$ are the learning rates of new context-to-item and item-to-context associations, respectively. When updating associations in $M^{CF}$, $\gamma_{emot}$ 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).

The $M^{FC}$ matrix handles the degree to which recalled memories reactivate their associated temporal and emotional (source) contexts. For simulating mood in emotional

disorders, we were primarily interested in the overall reactivation of emotional context during recall, as an operationalization of mood, rather than in the relative activation levels of temporal and emotional context elements. Therefore, for simplicity and to conserve free parameters, we allowed associations between item features and both types of context in $M^{FC}$ to form at the same rate, $\gamma_{FC}$. However, in future work, a more complex version of the model could allow these associations (between item features and temporal context versus emotional context) to accumulate at separate rates in $M^{FC}$. Because source features in F do not currently contribute to the recall process, the upper- and lower-right quadrants of $M^{FC}$ are set to 0.0 (Polyn et al., 2009).

Thus, before $\Delta M^{FC}$ and $\Delta M^{CF}$ integrate into $M^{FC}$ and $M^{CF}$, we scale each $\Delta M$ elementwise by its respective matrix of learning rates, $L^{FC}$ and $L^{CF}$:

$$L^{CF} = \begin{pmatrix} \gamma_{CF} & \gamma_{emot} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{6}$$

$$L^{FC} = \begin{pmatrix} \gamma_{FC} & \mathbf{0} \\ \gamma_{FC} & \mathbf{0} \end{pmatrix} \tag{7}$$

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 $\phi_S$ scales the overall level of the primacy effect and $\phi_D$ determines the rate of decay in this primacy effect, as the $i$th 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:

$$M^{CF} = M^{CF}_{Pre} + L^{CF} \sum_t \phi_t \mathbf{f}_t \mathbf{c}_t^\top \tag{8}$$

$$M^{FC} = M^{FC}_{Pre} + L^{FC} \sum_t \mathbf{c}_t \mathbf{f}_t^\top \tag{9}$$

Drawing upon work by Talmi et al. (2019), we additionally implement a parameter,

$\phi_{emot}$, which modulates the strength of context-to-item associations when an emotional item evokes arousal during encoding. For items that evoke no arousal, $\phi_{emot}$ takes a value of 1.0, and for items that evoke emotional arousal, $\phi_{emot}$ 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, $\mathbf{c}_R$. 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:

$$\mathbf{a} = M^{CF}\mathbf{c}_R \tag{10}$$

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 $\mathbf{a}$. On each step of this leaky accumulator process (Usher & McClelland, 2001), the vector of evidence, $\mathbf{x}$, evolves according to the following equation:

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

That is, the level of evidence on the current step equals the level of evidence on the last step ($\mathbf{x}_{n-1}$), 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 $\lambda$ and the time-decay scalar, $\tau$), plus incoming item activations, and plus noise, $\epsilon$. When the evidence in $\mathbf{x}_n$ for an item $i$ passes a certain threshold, $\theta_i$, that item emerges as a candidate for recall. For computational efficiency, we limit this evidence accumulation race to the items with the highest activations ($N_{items}$ = list length x 2), since the other items are very unlikely to emerge as successful candidates for recall (Lohnas et al., 2015).

This item then updates the $\mathbf{f}_t$ vector with its features, meaning that the $\mathbf{f}_t$ vector once again holds a 1 in the cell corresponding to the retrieved item, and 0's in the other items' cells. The item's associated context is then retrieved by multiplying $M^{FC}$ by the newly updated $\mathbf{f}_t$ vector. If the recalled item's evoked context, $\mathbf{c}^{IN}$, is sufficiently similar to the current context ($\mathbf{c}^{IN} \cdot \mathbf{c}_t > \mathbf{c}_{thresh}$), with $\mathbf{c}_{thresh}$ being a threshold scalar), then that item is reinstated into $\mathbf{f}_t$. Since the CMR3 model context vectors represent two subregions, each of which is updated and normed separately, the total contextual similarity value ranges between 0.0 and 2.0. 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 $\mathbf{c}_t$ 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 $\omega$ raises the absolute level of the threshold, and $\alpha$ decays this newly raised threshold each time the given item is reactivated. All dynamics of this part of the recall process (the reinstatement of items after the leaky accumulator) follow the model equations defined in CMR2 (Lohnas et al., 2015).

**Correspondence between CMR3 and prior retrieved-context models.** The model we formalize above, CMR3, builds on earlier retrieved-context models of episodic memory (Howard & Kahana, 2002a; Polyn et al., 2009; Lohnas et al., 2015; Kahana, 2020). The CMR model extended the original temporal context model (Howard & Kahana, 2002a; Sederberg, Howard, & Kahana, 2008) by including semantic and source information into

the latent representation of context (Polyn et al., 2009). Talmi et al. (2019) added (negative) arousal as an additional contextual attribute, and additionally allowed arousal to modulate the learning rate for context-to-feature associations. The CMR2 model (Lohnas et al., 2015; Healey & Kahana, 2016) extended the basic model to account for learning and interference effects across multiple lists, as well as providing a post-retrieval recognition mechanism that plays an important role in monitoring potentially intrusive memories. CMR3 extends CMR2 by adding a multivalent representation of emotion, allowing for neutral, positive, negative, or mixed emotional states. CMR3 extends eCMR by incorporating the multilist capabilites of CMR2.

## Experiment 1: The role of emotional valence in memory

To understand how episodic memory can promote negative versus positive emotion in emotional disorders, it is crucial to understand how emotional valence is represented in and evoked by human memory. When freely recalling lists of studied items, subjects tend to recall clusters of negative, positive, and neutral items (Long, Danoff, & Kahana, 2015). This *emotional clustering effect* suggests that not only the presence of emotion, but also its valence, guides the organization and retrieval of items in memory. Here, we replicate the emotional clustering effect in an independent dataset (Kahana, Aggarwal, & Phan, 2018; Aka, Phan, & Kahana, 2020) (described as Experiment 1, below).

Subsequently, we evaluate the ability of three variants of retrieved context theory to model this phenomenon: one in which there is no representation of emotional information other than as part of the memory's semantic content (Lohnas et al., 2015), one in which a single binary feature indicates whether a memory does or does not possess emotional information (Talmi et al., 2019), and a third in which two binary attributes separately indicate whether a memory possesses positive or negative emotional attributes (the current CMR3 model). We fit each of these models to the emotional clustering effect along with a set of other benchmark recall phenomena, and to the emotional in free recall (see

Simulation 1).

## Methods

Ninety-seven young adults (Mean age = 22; 51.5% female) completed delayed free recall of 576 unique lists, each comprising 24 unique words drawn from a subset of 1638 nouns from the University of South Florida word pool (D. L. Nelson, McEvoy, & Schreiber, 2004). Following list presentation and before the start of the recall period, subjects performed a 24 second arithmetic distractor task intended to attenuate recency sensitive retrieval processes. During the 75s recall period, subjects attempted to vocally recall as many words as they could remember, in whatever order they came to mind. Because subjects must recall words from the most recent list, 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. For full details of the experiment stimuli and methods see Kahana et al. (2018).

To assess the valence of each word, we used ratings from a prior norming study (Long et al., 2015). In this rating 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.

## Results

This experiment's aim was both to replicate the emotional clustering effect reported by Long et al (2015) and to obtain individual subject parameter sets for use in Simulations 2-8. By having each subject contribute 24 sessions of trials (totalling 576 lists) the present study provided sufficient data to allow for stable modeling at the individual subject level.

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. 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. We also assessed five benchmark recall phenomena unrelated to the emotional valence of items: serial position and output order effects, contiguity effects, semantic clustering effects, and patterns of prior- and extra-list intrusions (Kahana, 2020). We explain each of these effects as we present the results below.

The serial position curve (Figure 2A) illustrates the probability of recalling items as a function of their ordinal position within the study list. As is typical of delayed recall experiments, this analysis revealed a strong primacy effect – superior recall of early list items – but little or no evidence of recency – i.e., enhanced recall of final list items. The primacy effect seen in the serial position curve largely reflects subjects increased tendency to initiate recall with early list items, as seen in the probability of first recall (PFR) curve shown in Figure 2B. The lag-conditional response probability (Lag-CRP) shows the probability of successively recalling items as a function of their separation (lag) in the study list. This curve illustrates the contiguity effect, wherein subjects exhibit a strong tendency to successively recall items studied in neighboring list positions (small values of interitem lag) and they make these transitions with a forward bias (see Figure 2C.

We used a similar conditional probability analysis to measure subjects' tendency to successively recall semantically related list items (*semantic clustering*; Howard & Kahana, 2002b; Long et al., 2015). To represent semantic relations among the words, we used Word2Vec, which is a method of estimating the representations of words in vector space, or word "embeddings" (Mikolov et al., 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 1,638 noun word pool, from which we computed six bins of similarity values (-.18 to .02, .02 to .22, .22 to .41, .41 to .61, .61 to .80, .80 to 1.0), following procedures in Howard and Kahana (2002b) and Long et al. (2015). Figure 2D shows that subjects exhibited a semantic clustering effect, successively recalling semantically related items with far greater probabilities than unrelated items.

In addition, 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). As expected based on previous research (Kahana, Dolan, Sauder, & Wingfield, 2005; Zaromb et al., 2006), our healthy young adult subjects produced low rates of PLI's and ELI's per session, with means of 0.10 (SEM = .01) PLI's and 0.34 (SEM = .03) ELI's across subjects.

Lastly, 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 (Figures 3). To test the significance of this effect, we conducted a three (Initial Item Valence: Negative, Positive, or Neutral) $\times$ three (Transition-Item Valence: Negative, Positive, or Neutral) linear mixed effects model predicting transition probabilities (lme4 R package; Bates, Mächler, 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, $\beta = .057$, $t(768) = 35.1$, $p < 0.001$, and positive items were significantly more likely to be recalled together with other positive items than with non-positive items, $\beta = .021$, $t(768) = 12.8$, $p < 0.001$. 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 organization of memories

To model the effect of emotional valence on the organization of memories, as seen through the dynamics of the free recall process, we fit CMR2, eCMR, and CMR3 to the behavioral effects observed in Experiment 1. 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, $\phi_{emot}$ (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 to the pre-experimental $M^{CF}$ matrix that represents associations between items and contexts.

## Simulation 1

**Method.** Using a particle swarm optimization method, we fit each of the three CMR variants to data from each individual subject. To obtain the best-fitting parameters for each model, fit to each subject, we minimized the $\chi^2$ error between each model's predictions and the actual data for that subject (see Appendix). This produced a set of 97 best-fitting parameters sets and predicted data points for each model (Table A1).

Fitting 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. The resulting parameters enable simulating an entire 24-session experiment, using the actual list items that a given subject experienced to determine the sequence of items presented to the model. The model then generates a set of simulated recall sequences, which we analyze to estimate the same behavioral patterns shown in Figure 2. For each model's simulated dataset of "virtual subjects," we repeated the full behavioral analyses that we ran on the empirical data. This enabled us to evaluate and compare each model's ability to fit the aggregate behavioral data.

**Simulation Results**

To assess each model's ability to fit the aggregate data, we calculated three measures of fit: (1) the same $\chi^2$ 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 (see Appendix). The resulting $\chi^2$ 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. 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.

Next, we examined which behavioral effects distinguished each model's ability to fit the aggregate data. We calculated RMSE values for each behavioral analysis (Table 2). The CMR3 model provided the smallest total 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 Figure 2 confirms that, with some variation, all three models predicted the shape of the serial position curve and the pattern of recall initiation, which indicates that subjects tend to initiate recall with items from the beginning or the end of the list (Figures 2a,b). All three models also captured subjects tendency to successively recall items studied in neighboring list positions (the temporal contiguity effect) (Figure

2c) and their tendency to successively recall semantically similar items (the semantic clustering effect, (Figure 2d).

The models differed, however, in their ability to capture the emotional clustering effect, the key phenomena of interest in this simulation (Figure 3). Figure 4b 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.

We have shown that to capture the emotional clustering effect, a model should represent emotion as having positive and negative valence (as in CMR3). One of the strengths of this dataset is its ability to show how emotional valence influences recall in the absence of significant arousal. A limitation is that we did not orthogonally manipulate arousal and valence. Further, as we limited our stimuli to low-arousal items in order to isolate the effects of valence, Experiment 1 did not produce an emotionally-enhanced recall effect, which eCMR was designed to capture (Talmi et al., 2019). Testing CMR3 and the updated eCMR (expanded here to have multilist capabilities) on experiments with high-arousal items can reveal the relative influences of emotional valence and arousal in episodic memory, and would indicate whether separate modeling assumptions are required to capture enhanced recall of high-arousal items (Talmi et al., 2019).

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

Having shown that CMR3 can account for the emotional clustering effect without sacrificing its fit to benchmark recall phenomena, we next examined CMR3's ability to account for mood-congruent and emotion-state dependent recall. We consider these two classic phenomena of memory and emotion as a prelude to our main objective, which is the development of a new theory that extends the retrieved context framework to account for persistent negative mood and the production of intrusive, non-voluntary memories. We see these as potential transdiagnostic processes underlying 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 & Russell, 1983). Yet, once recovered, this patient may tend to recall positive memories and have difficulty remembering the events that occurred while they were depressed. Heightened recall for memories whose emotional attributes match the person's current emotional state, or *mood-congruent recall*, has been observed widely across both clinical and healthy populations: e.g., for sadness and loss-related stimuli in patients with depression (B. P. Bradley, Mogg, & Williams, 1995; Teasdale & Fogarty, 1979; P. C. 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. C. Watkins et al., 1992). Talmi et al. (2019) have previously demonstrated the ability of retrieved-context models (eCMR) to capture mood-congruent recall for emotional versus non-emotional items. Here, we test whether retrieved-context theory can account for mood-congruent recall within distinct types of emotional states (decreased recall of positive events when in a negative mood state, and vice-versa).

**Simulation 2**

**Method.** We tested CMR3's ability to predict mood-congruent recall. Specifically, we simulated the results from a series of free recall experiments in which mood-congruent recall 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 from a normal distribution ($\mu = .2, \sigma = 0.5$). 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, with no emotional content in the subject's context vector. After each encoding list, we simulated a mood induction prior to free recall, by setting the virtual subject's emotional context to contain negative (20 lists), neutral (20 lists), or positive (20 lists) emotional content. The model then predicted the subject's recall responses during free recall, 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 virtual subject: that is, 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.** Using the sets of parameters obtained in Simulation 1 (see Table A1), we simulated a set of 97 virtual subjects, by specifying the CMR3 model with each set of parameters and obtaining each model-predicted output for a virtual recall task. For an overview of which parameters we manipulate in each simulation, see Table 3.

Then, 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 5a). 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 et al., 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.

In CMR3, we operationalized neutral mood as the reduced or relative absence of emotional context, rather than as a distinct mood state. This produces mood-congruent recall for positive and negative mood states, and enhanced recall of emotional material in a neutral mood state, consistent with empirical findings (Long et al., 2015; Siddiqui & Unsworth, 2011; Talmi et al., 2019). Enhanced recall of emotional material in a neutral mood state emerges due to emotional context adding additional contextual elements that can cue the retrieval of these memories. However, an alternate way to operationalize neutral mood would be to consider it a “mood state” in and of itself, rather than the

absence of negative or positive emotional content. In this design, the model would produce enhanced recall for neutral items in neutral mood states.

## 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 (M. Craske et al., 2008). 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 (M. Craske et al., 2008). 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 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 (M. Craske et al., 2008).

## Persistent negative mood

Negative mood that persists for two weeks or longer is a core feature of major depressive disorder (MDD; American Psychiatric Association, 2013). Stressful life events can trigger MDD, with the likelihood of a depressive episode increasing with the frequency of recent negative events (Kendler, Karkowski, & Prescott, 1998). 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 go beyond

modeling emotional effects on list memory to simulate an individual's life history, and evaluate how prior and recent negative life events can lead to the development and persistence of negative mood through the mechanisms of our memory model.

**Simulation 4**

Going beyond the classic list memory paradigm, we imagine human experience as unfolding in a sequence of events. We consider each event as analogous to a list composed of sequentially experienced stimuli, each defined by their features, emotional valence, and associations to other stimuli. Mental life involves both experiencing these events (akin to learning a list) and thinking about them, or remembering them, at future moments (akin to recalling the list). These simplifying assumptions allow us to take a first step in generalizing the rich literature on list memory paradigms to humans' experience at much longer time scales. Nonetheless, we recognize that our approach to modeling human experience necessarily simplifies the much more complex structure of real-life experiences.

In this and subsequent simulations, we evaluate different assumptions regarding a simulated individual's history with emotional experiences for that individual's contemporaneous and subsequent mood. We then simulate the acquisition of these experiences using the learning rule embodied in the CMR3 model and the model's assumptions regarding the evolution of context. We assume that individuals not only experience events, but also think about them, and simulate these internal reflections as a memory retrieval process akin to the free recall task. We model internal thinking as retrieval of stimuli, cued by the model's continuously evolving representation of context.

We designed these simulations to address four key components of a theory of clinical disorder: how the disorder develops, what factors maintain it, the mechanisms of treatment, and what factors lead to sustained improvement versus relapse. In our simulations, these correspond to four distinct periods: A developmental period, a neutral period, a treatment period, and a recovery period. Before describing how we simulated

each period, we first introduce our basic methods for simulating human individuals who have variable memory processes and unique sets of experiences.

**Method.** *Virtual Subjects.* As in our simulations of mood-congruent and mood-state-dependent recall, we create a sample of virtual subjects (N = 97). The CMR3 model equations govern the memory processes of each virtual subject, and each parameter set defines the extent to which each process affects encoding and retrieval for a unique subject. We obtained these model-parameters from our initial detailed model-fits to the behavioral data in simulation 1.

*Simulated Events.* For each virtual subject, we simulated a life-course of events, each presented within a "list," or sequences of stimuli, that comprise a larger episode in a virtual subject's life. At the end of each such episode, or list, we allowed the model to engage in a moment of "free recall," to model the spontaneous recall of these prior events, based on the literature that the same episodic memory system which supports recall of voluntary (intended) memories also supports recall of involuntary (spontaneous, unintended) memories. We simulated four periods of life events, each distinguished by their compositions of neutral, negative, and positive events: (1) a Developmental Period, (2) a Neutral Period, (3) a Treatment Period, and (4) a Recovery Period.

The *Developmental Period* tests the model's ability to predict different outcomes for virtual subjects depending on their history of early lifetime emotional events. During the Developmental Period, virtual subjects experienced 50 sets, or "lists," of 10 encoding events. We designed three such developmental periods for virtual subjects to experience. 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. Using a simulation approach allowed us to run each condition (Low, Moderate, and High frequencies of negative events during the Developmental Period) for each virtual subject.

After the Developmental Period, we introduced a *Neutral Period*, in which virtual subjects experienced 30 lists of 10 new, unique encoding events, all of which had neutral valence. The Neutral Period allows us to isolate and identify the effects of prior learning during the three distinct Developmental Periods on subsequent mood and memory, without needing to filter out the effects of new emotional events.

A theory of emotional disorder must account not only for risk and maintaining factors, but also for what interventions reduce symptoms. After the Neutral Period, we presented each virtual subject with a simulated *Treatment Period*, to determine whether CMR3 could predict the efficacy of positive event scheduling. In positive event scheduling, which is a core component of Behavioral Activation Therapy (BAT) for depression, patients increase their frequency of and re-engage with positive and meaningful activities (Beck et al., 1979; Jacobson et al., 1996; Lewinsohn, Sullivan, & Grosscap, 1980). Each simulated course of positive event scheduling consisted of 20 new lists of 10 events. Each list had a heightened ratio of positive to negative events (3:1), having a list composition of 30% positive, 10% negative, and 60% neutral events.

Finally, we sought to test the conditions under which the model predicted relapse versus a sustained response to treatment. During the *Recovery Period*, which followed the Treatment Period, we presented each virtual subject with 60 lists of predominantly neutral events intermixed with a small proportion of positive and negative events (80% neutral, 10% negative, 10% positive events).

*Event Properties.* For each event in the periods described above, we assigned different properties, including the type of emotional valence and the event's semantic relatedness to other events. Although the model allows events to have mixed valence, for simplicity, we allowed each event to have only one valence: negative, positive, or neutral. To simulate

event properties and semantic associations, we drew the values for all inter-event similarities, regardless of events' emotional valence, at random and without replacement from the same normal distribution ($\mu = 0.2$, $\sigma = 0.5$). For interpreting these inter-item similarities, positive values indicate similarity and negative values indicate dissimilarity. Thus, all events share the same distribution of inter-item similarities both within a list of events, and across lists of events.

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 have an inter-item similarity value of 1.0, or perfect similarity, to ensure that all simulated events are unique. In addition, we set $s$, which scales these inter-item similarity values and thus controls how much semantic associations guide memory retrieval, to a value of 0.15. Drawing inter-item 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.

*Simulation outputs.* As each virtual subject experienced the simulated encoding events, we tracked the model-predicted levels of mood evoked by each event. As described above (see Model Description), the model produces separate activations for the negative and positive emotion evoked by each event. For conciseness in graphical depiction of the model results, we calculated a composite measure of mood that consists of the difference between the levels of positive and negative affect present in the context vector at that time.

During the session of free recall after each list, as each virtual subject's memory system spontaneously responds to the current context as a retrieval cue for encoded memories, we tracked the occurrences of *voluntary memories*, or memories from an intended (target) life-context (for simplicity, operationalized as the most-recent list of events), as well as *involuntary memory intrusions*, or memories from an undesired and

non-intended (non-target) life context (operationalized as a list preceding the most-recent list of events). For consistency with the list-learning literature, we refer to the former as "correct recalls" and the latter as "prior-list intrusions." The *correct recall* metric refers to the probability of correctly recalling a target item, and it is calculated independently of the number of prior-list intrusions that occur in that list.

**Results and Discussion.** In this simulation, we tested how different compositions of emotional events during a virtual subject's early development influence their subsequent mood state and memory patterns. In Figure 6, we illustrate the dynamics of emotion for one virtual subject, to demonstrate the simulation outputs in fine detail. Each panel depicts the model-predicted mood for the same virtual subject as they experience each of three early Developmental Periods: Low-Frequency, Medium-Frequency, and High-Frequency negative events. Since this is a simulation, we were able to run the simulation multiple times for the same subject, holding the subject's unique individual properties constant, while varying their early Development Period among each of these three conditions. As time unfolds along the horizontal axis, this figure then displays the virtual subject's mood during the subsequent Neutral, Treatment, and Recovery periods (see Method for full details). This allowed us to assess how the principles of retrieved-context theory predict that a virtual subject may develop negative mood, maintain it once prior negative events have abated (the Neutral Period), what factors contribute to recovery (the Treatment Period), and under what conditions the virtual subject relapses or sustains that recovery (Recovery Period).

In Figure 6, the vertical axis tracks the model-predicted mood for this virtual subject, analogous to a heart monitor. The graph's height represents the positivity or negativity of model-predicted mood, calculated as the difference (subtraction) between the level of positive emotion and negative emotion present in the virtual subject's emotional context 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. In each run of the simulation, the number of events that a virtual subject encounters is always the same. However, because CMR3 models recall as a noisy process, this allows the number of recalled memories to vary in each simulation run; therefore, the length of each simulation differs across panels. Finally, Figure 7 displays the aggregated patterns of model-predicted mood and memory for all virtual subjects, averaged within each period of the simulation (Developmental, Neutral, Treatment, or Post-Treatment).

**Mood.** As the frequency of negative events during the developmental period increased (i.e., in the Low, Moderate, and High-Frequency negative events conditions during the Developmental Period), 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 7, Panel A). The emotional context that accrued during the developmental period persisted into the neutral period, such that virtual subjects who experienced high rates of prior negative events developed persistent negative mood, even though they rarely recalled events from the developmental period itself (Figure 7, Panel B). Thus, the model predicted that negative mood can perpetuate from prior negative experiences even when a patient is not actively recalling them. Rather, recalling recent neutral events that had become associated to negative background context during their encoding, evoked the emotional context with which they had been encoded. Thus, the negative emotion from recent negative events generalized to neutral events, causing them to reactivate negative mood even though these neutral events had only neutral properties in and of themselves.

After each type of developmental period (low, medium, and high frequency negative events), positive event scheduling produced an improvement in mood that persisted during the recovery period (Figure 6 and Figure 7a; but see the subsequent "Repetition Effects" simulation for the conditions under which the model predicts that this intervention is less or not effective). The persistence of negative mood in the neutral period prior to this

intervention arose because, although neutral events in these simulations do not hold emotional content, neutral events will be bound to a variable emotional context. When bound to a relatively neutral or positive context, recalling such neutral events will wash out a strong negative context as time goes by, since they are associated to neutral context. Thus, neutral events can attenuate the extremity of mood.

However, this does not occur if the neutral events have been associated to negative contexts, as in the Medium-Frequency and High-Frequency conditions. In this case, when neutral events are recalled, they will reactive their associated negative contexts, thus causing negative emotional context to persist. Thus, encountering new neutral or positive events against a nonnegative emotional context backdrop is essential for mood to improve. Overall, the model predicts that it is the combination both of the valence of an event's features, and the valence of its associated emotional context, that together contribute to the current mood state. Neutral events temporarily improve mood when they occur, but then what mood they reinstate when the subject recalls the event depends on what their associated emotional context was.

**Recall.** In Figure 7, we present the mean levels of types of recall outputs within each period (Panels B-D), averaged across virtual subjects. For consistency with list-learning paradigms, we use "correct recall" to refer to instances in which the virtual subject has recalled a memory that occurred within the intended (target) life-context (i.e., "list"). In our simulations, this corresponds to voluntary (strategic, or targeted) autobiographical recall (Berntsen, 2010). Also drawing from list-learning literature, we use the term "prior-list-intrusion" to refer to instances in which the virtual subject has spontaneously recalled an undesired memory from an unintended (non-target) life-context (i.e., a "prior list"). Because such memories intrude from a prior list, they are called *prior-list intrusions* (PLIs). During our simulations, such memories can arise spontaneously and intrude into the virtual subject's awareness as each modeled memory system responds freely to the current retrieval context.

During the neutral period, the model predicted decreased probability of recall for recent neutral events while virtual subjects were still experiencing persistent low mood, due to the prior stressors (Figure 7, Panel B). Recall for these events improved during the simulated treatment period (Figure 7, Panel B). Prior theorists accounting for reduced recall of neutral or positive events during depressed mood states have proposed that such reduced recall arises from impairment of the neural structures that support the memory system (Austin, Mitchell, & Goodwin, 2001; Bremner, Vythilingam, Vermetten, Vaccarino, & Charney, 2004). Our results propose an alternative, or perhaps complementary, explanation. In a negative mood state, mood-congruent recall processes (see Simulation 3) should decrease recall for neutral and positive stimuli. Since cognitive tests often use neutral stimuli, 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) memory system impairment. Importantly, our model is just one of many possible explanations for this finding. Higher-level cognitive processes not accounted for by our model, such as heightened self-referential thinking in depression (Clark & Beck, 1999; Mor & Winquist, 2002) and reduced attention to the experience of new events (E. Watkins & Teasdale, 2004), may also affect and decrease episodic memory recall.

## Repetition Effects

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, Kaltman, & Green, 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 negative events might perpetuate negative mood by amplifying the activation of memories for repeated negative events.

### Simulation 5

**Method.** We repeated Simulation 4 (see above for full method). However, during the developmental period, we selected one negative event to repeat across lists, every other list. As in Simulation 4, we followed each developmental period with a period of 100% neutral events.

**Results.** In this simulation, we sought to determine the model-predicted effects of repeated negative events on emotion and memory. Memory of the repeated negative-event intruded into memory during the neutral period, in which neither this repeated negative-event nor other negative events had occurred (Figure 8). In comparison, during Simulation 4, in which all stressors were unique, no memories of negative events intruded during the neutral period. Thus, spaced repetitions predisposed this stressful event to spontaneously reactivate during recall periods. Of clinical interest, two distinct groups emerged among the virtual subjects. One group experienced multiple stressor-memory intrusions per list (High-Intrusions Group, n = 12), whereas the second group exhibited almost zero stressor-memory intrusions (Low-Intrusions Group, N = 85) and more positive mood (Figure 8, Panel A). Thus, different memory system parameters predisposed some virtual subjects to experience heightened vulnerability to intrusive memories of negative events and corresponding negative mood.

In retrieved-context theory, repetition associates the negative event 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 negative event and potentiates it for activation in contexts where it does not belong (i.e., as a memory intrusion).

Compounding this effect, the principles of mood-congruent recall predict worsened mood for two reasons: reactivating negative context increases the probability of recalling subsequent negative memories, and decreases the probability of recalling positive memories that might otherwise lift mood (see Simulation 2).

Repeated negative events also affected the ability of positive event scheduling to lift depressed mood. During the Treatment Period, heightened rates of positive events during the simulated treatment temporarily lifted mood (Figure 8, Panel A), but for patients with high rates of intrusive memories of negative events, these negative-memory intrusions continued to evoke negative context. The resulting negative context not only worsened mood, but also, reduced recall of more recent positive events due to the process of mood-congruent recall (see *Mood-congruent recall*, Simulation 2), as observed in real depressed participants (R. E. Nelson & Craighead, 1977).

Our simulation highlights the importance of directly targeting intrusive memories of negative events in treatments for depression. Indeed, 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 (E. R. Watkins et al., 2011). Next, we examine the model's ability to characterize the mechanisms of intrusive, non-voluntary memories of high-arousal negative stressors.

## Intrusive Memories

### Motivation

In the aftermath of extreme or traumatic stress, patients with PTSD report that memory intrusions – spontaneous, and at times repetitive, involuntary memories of the traumatic event that patients have trouble filtering from their mind – constitute their most distressing symptom (Ehlers & Steil, 1995). For intrusive memories in PTSD, stimuli whose features strongly overlap with the stimuli 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. The tendency of contextual cues, including perceptual, temporal, and spatial elements of context, to activate trauma-memory intrusions is best known in PTSD. However, intrusive memories of painful events are found in numerous disorders beyond PTSD (Brewin, Gregory, Lipton, & Burgess, 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 (A. D. Williams & Moulds, 2007).

Leading treatments for trauma-memory intrusions in PTSD, such as prolonged exposure therapy (Foa & Kozak, 1986), involve repeated exposure to the contextual cues that trigger intrusive involuntary recall of the trauma memory, called *in vivo* exposures, as well as repeated *imaginal exposure* to the trauma-memory itself, which we examine in Simulation 8. Here, we examined whether retrieved-context theory's proposed mechanism, of associated context serving as a retrieval cue, can account for findings that neutral stimuli preceding a negative event can reactivate intrusive memories of that event. Further, we examined the model-predicted effects of avoiding such cues, versus repeatedly approaching these cues during *in vivo* exposures.

## Simulation 6

**Method.** This simulation follows the same format as Simulation 4. First, we presented all virtual subjects with the same developmental period of 50 lists of 10 encoding events, consisting of 40% neutral, 30% positive, and 30% negative events. Within 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, $\phi_{emot}$. Having been introduced to retrieved-context models by

Talmi et al. (2019), $\phi_{emot}$ scales the strengths of the new context-to-event associations that form in $M^{CF}$ during encoding (equation 4). If $\phi_{emot}$ 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 $\phi_{emot}$ to 5.0, because this value resulted in heightened occurrences of trauma-memory intrusions, and thus sufficient variability of trauma-memory intrusion rates across virtual subjects to conduct analyses (Figure 9).

To examine the effects of high-arousal negative events without "background noise" from the effects of high-arousal positive events, we set $\phi_{emot}$ to modulate context-to-memory associations only for negative events. However, setting $\phi_{emot}$ to also modulate context-to-memory associations for positive events would result in enhanced memory for and spontaneous intrusions of high-arousal positive memories, consistent with empirical findings (Berntsen, 1996; Berntsen & Rubin, 2002).

After the developmental period, we presented each virtual subject 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 given Neutral Period condition. We simulated three such 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.

Although subjects may experience voluntary (strategic) and involuntary (spontaneous) autobiographical memories as distinctive processes, literature suggests that both processes share and operate through the same mechanisms in episodic memory (Berntsen, 2010). For example, in PTSD, when patients report experiencing a spontaneous memory intrusion, further investigation typically reveals that the memory was in fact cued by a temporal, spatial, or other contextual cue associated to the trauma (Ehlers & Clark, 2000), just as contextual cues guide recall of other episodic memories. Additionally, both in the laboratory and in daily life, even healthy subjects spontaneously recall neutral memories that are unintended and undesired, highly vivid, and which they may not recognize as being an intrusion (Berntsen & Rubin, 2002; Berntsen, 2010; Staugaard & Berntsen, 2014; Zaromb et al., 2006; Kahana et al., 2005; Roediger & McDermott, 1995).

To examine such memory-intrusions in the laboratory, episodic memory research has used the *externalized free recall task* (Zaromb et al., 2006; Kahana et al., 2005; Lohnas et al., 2015), which we use as the basis for our simulation here. During an externalized free recall task, subjects report all items that enter their mind during the task, whether or not they intended to recall the item. Thus, subjects report both voluntary and involuntary (spontaneous) recalls. The latter arise in the background of the subject's intentional retrieval search, occurring as a spontaneous and undesired process that subjects must try to censor as they maintain their focus on their desired mental content. Externalized free recall literature shows that patients exhibit varying degrees of control over, and awareness of, prior-list and extra-list intrusions (Zaromb et al., 2006), analogous to spontaneous trauma memories in PTSD. Therefore, in this simulation, we operationalized memory-intrusions as spontaneous memories that a virtual subject recalls from an unintended and undesired prior life-context, or "list of stimuli," which we therefore term *prior-list intrusions* (PLIs).

**Results.** We tested whether retrieved-context theory could predict the effects of "trauma cues," or stimuli whose features are similar to those that directly preceded a traumatic event, on cueing intrusive involuntary recall of that event (Ehlers & Clark, 2000). Figure 9 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 resembles actual rates of chronic PTSD after trauma exposure which typically range from 5-10% but can be higher depending on the type and severity of the trauma (Bonanno, Westphal, & Mancini, 2011).

In each group, 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 each of three conditions. 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 frequencies of trauma-memory intrusions (Figure 10), virtual subject's ability to voluntarily recall non-trauma memory content (i.e., the probability of "correct recall"; Figure 11), and mood (Figure 12).

**Recall.** Across virtual subjects, the model predicted that an initial re-encounter with the neutral context that had preceded the trauma-event cued an increase in

memory-intrusions of the event, consistent with prior findings and theory that even neutral and semantically unrelated stimuli directly preceding a traumatic event can cue flashbacks in PTSD (Ehlers et al., 2004). Retrieved-context theory predicts this finding for the stimuli that precede rather than follow a traumatic event due to the *temporal contiguity effect*. As replicated in Experiment 1 (Figure 2, Panel C, and see Kahana, 1996, 2012), retrieved-context models propose that items in memory are organized according to the time in which they occurred, such that recalling one item cues recall of items that occurred close in time to it during encoding. However, this effect is "forward-asymmetric" (Kahana, 2012): prior stimuli are more likely to cue recall of the associated stimuli that follow than those that precede them because, as context evolves over time, each prior stimulus becomes part of the contextual cues for the subsequent stimuli.

The tendency of trauma-cues to reactivate trauma-memory intrusions differed according to whether a virtual subject had been experiencing a low vs. high frequency of trauma-memory intrusions after the event. Virtual subjects in the Low-Intrusions group experienced a minimal and temporary increase in trauma-memory intrusions upon their first encounter with the trauma-cues following the traumatic event (Figure 10, Panel A), after which the frequency of such intrusions returned to near-zero levels. Otherwise, these virtual subjects continued to have minimal to no trauma-memory intrusions, regardless of condition.

For virtual subjects experiencing high rates of trauma-memory intrusions (i.e., the High-Intrusions group), repeatedly approaching the trauma cues reduced trauma-memory intrusions in the High-Intrusion Group to a frequency comparable with that of the Low-Intrusion Group, i.e., nearly zero ("Multiple Exposures," Figure 10, Panel B). In retrieved-context theory, re-encountering the trauma cues associated these cues with more varied non-traumatic contexts. The cues became associated with and began to evoke a wider variety of recall outputs, thereby decreasing their tendency to cue recall of the trauma event. However, avoiding trauma cues, or encountering them only once, resulted in

the trauma-memory intrusions persisting at high levels (Figure 10, Panel B). Based on this simulation, we propose that restructuring context-to-memory associations in episodic memory is one process by which *in vivo* exposures in Prolonged Exposure Therapy (Foa & Kozak, 1986) reduce the rate of trauma-memory intrusions. We model *imaginal exposures*, the second core component of PE, in Simulation 8.

**Mood.** In the High-Intrusions group, multiple exposures to the trauma-cues allowed mood to improve as the rate of intrusive trauma-memories decreased, whereas mood showed little improvement in the Avoidance and Single-Exposure conditions (Figure 12). 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. Thus, retrieved-context theory proposes a novel account of how, by restructuring episodic association networks in memory, trauma survivors can not only habituate to trauma-associated stimuli, but also, recover prior enjoyment of such stimuli (e.g., as in the case of a survivor of sexual assault who not only habituates, but also recovers feelings of joy and affection, in romantic relationships).

By predicting symptom change as the result of associating new neutral or positive contexts to the trauma cues, rather than replacing old ones, our model 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 (M. Craske et al., 2008). Retrieved-context theory proposes 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. These predictions accord with principles of inhibitory learning (M. Craske et al., 2008); here, we show that episodic memory processes can produce this same pattern of findings, via restructuring the network of memory-to-context associations. This framework enables our model to predict the reemergence of trauma-relevant emotional context beyond fear, such as guilt, shame, sadness, or anger, when a trauma-memory is reactivated by a context in which new learning has not taken place.

**Emotion Dysregulation.**

Retrieved-context theory also accounts for how individual differences can moderate the frequency and emotional experience of intrusive memories. For example, emotion dysregulation, or difficulties understanding and regulating the intensity and valence of one's emotional experiences, can serve as a risk factor for the development of PTSD and can lead to more extreme responses to trauma cues (Tull, Barrett, McMillan, & Roemer, 2007; Tull, Berghoff, Wheeless, Cohen, & Gratz, 2018; Messman-Moore, Walsh, & DiLillo, 2010). In this follow-up analysis to Simulation 6, we assess how emotion dysregulation moderates the dynamics of intrusive memories.

**Method.** . We repeated Simulation 6, but this time altered the properties of virtual subjects' emotion and memory systems to create two groups: (1) virtual subjects with heightened emotion dysregulation (High Emotion Dysregulation group), and (2) virtual subjects with low emotion dysregulation (Low Emotion Dysregulation group). To create these groups, we operationalized emotion regulation ability as the rate at which emotional context updates for a given subject, $\beta_{emot}$, meaning how swiftly the emotional contents of new events enter and update the emotional components of the subject's internal context state. In the High Emotion Dysregulation group, we set $\beta_{emot} = .8$ for all virtual subjects,

whereas in the Low Emotion Dysregulation group, we set $\beta_{emot} = .3$ for all virtual subjects. Note that in this paradigm, although emotion enters and leaves the context vector more swiftly for patients with high emotion dysregulation, lasting associations still form between events and the higher-strength emotional context that these patients experience.

**Results.** Among virtual subjects with frequent trauma-memory intrusions, heightened emotion dysregulation led to higher rates of trauma-memory intrusions (Figure 13. In the High Emotion Dysregulation group, virtual subjects experienced a mean of 6.82 (SEM = 1.20) trauma-memory intrusions per list during the Neutral Period, compared to a mean of 4.92 (SEM = 1.10) trauma-memory intrusions per list in the Low Emotion Dysregulation group. The High Emotion Dysregulation group also experienced more severe negative emotional responses to trauma-memory intrusions during the Neutral Period, with a mean mood of -0.44 (SEM = 0.04) versus a mean mood of -0.13 (SEM = 0.05) among the Low Emotion Dysregulation group during this period.

Retrieved-context theory predicts this pattern of results because virtual subjects with heightened emotion dysregulation experience emotional context more strongly, and thus have heightened levels of negative emotional context associated to a trauma-event. A stronger degree of emotional context associated to events thus amplifies the role of emotional context in guiding the retrieval context. Thus, heightened emotion dysregulation results in a greater tendency to recall negative memories when experiencing a negative mood state (mood-congruent recall), as well as a greater tendency to recall even neutral memory content that was encoded during an emotional context that matches the subject's current emotional state (emotion-state dependent recall).

## Intrusive Memories: Vividness and Nowness

### Motivation

Unwanted and intrusive, vivid and highly visual memories of painful events are best known in PTSD but occur widely across emotional disorders, including depression

(Reynolds & Brewin, 1999), social phobia (Hackmann et al., 2000), obsessive compulsive disorder (Speckens et al., 2007), health anxiety (Muse et al., 2010), body dysmorphic disorder (Osman et al., 2004), and agoraphobia (Day et al., 2004). Intrusive trauma memories are characterized by a greater sense of "nowness," or a feeling that the memory is happening again in the here-and-now. Since the classic work of Tulving (1972), episodic memory has been conceptualized as a form of "mental time travel" due to the ability of even neutral memories to reinstate the context of prior events and thereby transport the rememberer back in time. Here, we test retrieved-context theory's prediction that heightened arousal during encoding predisposes a memory to be recalled more vividly and with a heightened sense of "nowness," or the perception that the memory sources from the immediate and most-recent context.

## Simulation 7

**Method.** *Vividness.* We operationalized vividness as how strongly a subject's recall context reactivates a memory's features. In CMR3, this corresponds to the item activation values that result from cueing the Hebbian weight matrices with the current recall context, $\mathbf{a} = M^{CF}\mathbf{c}_R$ (Equation 10).

*Nowness.* Recent work has conceptualized mental time travel, or the sense of being back in and reliving a memory, as the degree of similarity between current context and the context that was present at the time of encoding (Howard, Shankar, Aue, & Criss, 2015). Here, we operationalize the degree of perceived "nowness" as how closely a memory's reinstated context matches the subject's current temporal context. In prior retrieved-context models, this same similarity metric has also served to determine whether a subject can distinguish when a memory has been retrieved from the current target context, versus being recalled as an intrusion from a distant lure context (Lohnas et al., 2015).

In the model, this similarity value equals the dot-product similarity between the

subject's current context and the context reactivated by the trauma-memory during recall, $c_{sim} = \mathbf{c}^{IN} \cdot \mathbf{c}_t$). If the match between encoding and retrieval contexts exceeds the subject's threshold parameter, $c_{thresh}$, then the subject experiences the memory as belonging to the most recent context (as having occurred "just now"). The model parameter $c_{thresh}$ thus serves as an individual differences parameter that governs a subject's ability to exert cognitive control over the recall process, in order to successfully (or unsuccessfully) filter out unwanted memory intrusions and other extraneous mental content. A higher $c_{thresh}$ value corresponds to a greater degree of cognitive control.

*Arousal.* We measured how heightened negative arousal, represented in the model as the parameter $\phi_{emot}$, which scales the strength of context-to-item associations during encoding, affects vividness and nowness of trauma memories. To ensure a sufficient sample of trauma-memory intrusions for analyses, we varied $\phi_{emot}$ from 2.0 to 5.0, since at a level of 1.0 (i.e., a value determining that there is no effect of arousal on memory), the virtual subjects experienced zero trauma-memory intrusions to analyze. As in the prior intrusive involuntary recall simulation (Simulation 6), we randomly assigned 30% of items in each list of 10 items to be negative, 30% to be positive, and 40% to be neutral. We set the trauma item to occur in the middle of its list, at position 6 out of 10 positions, to minimize serial position effects.

**Results.** As negative arousal during the trauma increased, the trauma-memory's vividness (Figure 14, Panel A) and "nowness" (Figure 14, Panel B) increased. In our operationalized metrics, negative arousal during encoding subsequently increased both the magnitude of reactivation of a memory's features (vividness) and the match between the memory's reactivated context and its retrieval context (nowness).

Within the model framework, this occurs for three reasons. First, retrieved-context theory proposes that negative arousal strengthens associations between the memory and its context during encoding. Second, when heightened negative arousal potentiates the memory system's associative networks, trauma-memories require lower levels of associated

contextual cues to reactivate them, and the trauma-memory begins to intrude more frequently. Thus, new contexts already contain elements of trauma-associated context reactivated by prior trauma-memory intrusions. In addition, patients might introduce and consistently refresh the presence of trauma-associated context by ruminating about the trauma (Michael, Halligan, Clark, & Ehlers, 2007). Third, because traumatic events evoke strong negative emotional context, recalling a traumatic memory flavors the current context with negative features. The current context becomes more negative; thus, the emotional context present just prior to retrieval is a stronger match for the negative emotional context of the memory. In addition, because the current context becomes more negative, it is more likely to cue the reactivation of further negative memory-intrusions due to mood-congruent recall, which in turn flavor context with their own negative features, creating a snowball effect.

Because heightened arousal potentiates the associations between contextual cues and the trauma memory (Talmi et al., 2019), reduced levels of trauma-context are needed to cue spontaneous, involuntary recall of the trauma-memory. In this case – low levels of trauma-associated context being present – one might predict a low match between the current context and the context reactivated by the trauma-memory, which would correspond to a reduced "nowness" value in our simulations. However, even when external stimuli contain low levels of trauma-related perceptual features or spatiotemporal context, ruminating on the trauma memory can maintain internal representation of trauma-related context. In addition, one of the strengths of the eCMR and CMR3 models is the ability to account for a strong overlap in encoding and retrieval contexts in source contexts, such as emotional states, even when overlap in temporal contexts is reduced. Since patients with PTSD often experience a persistent sense of threat and other negative emotions (American Psychiatric Association, 2013), the negative emotional context that a trauma-memory reactivates will provide a strong match to current mood state, thus heightening the degree of contextual overlap.

## Prolonged Exposure Therapy

### Motivation

Among multiple efficacious treatments for trauma-memory intrusions and other symptoms in PTSD (Tran & Gregor, 2016), one particularly efficacious treatment is prolonged exposure therapy, or PE (PE; Foa & Kozak, 1986; Powers, Halpern, Ferenschak, Gillihan, & Foa, 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 Simulation 6). 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. 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 (M. Craske et al., 2008; Jaycox, Foa, & Morral, 1998).

### Simulation 8

**Method.** Prolonged exposure therapy includes two types of exposure: *in vivo* exposure, in which patients gradually and systematically approach feared trauma cues and contexts, and *imaginal exposure*, in which patients conduct exposures to the trauma memory. In PTSD, the trauma memory is itself a feared stimulus that patients with PTSD avoid. In Simulation 6, we modeled *in vivo* exposures, demonstrating how repeated exposures to feared cues and contexts associated to the traumatic event gradually reassociates those cues to new, safe stimuli. The model predicted that a sufficient number of *in vivo* exposures – in this simulation, four or more – will decrease the tendency of a given contextual cue to reactivate the trauma memory, reducing and ultimately eliminating

trauma-memory intrusions.

In addition to reducing the frequency of trauma-memory intrusions, the imaginal exposure component of PE decreases patients' distress in response to the trauma memory, both in cases when the patient recalls it either voluntarily or in which it intrudes spontaneously. We now present a simulation demonstrating how the model accounts for this phenomenon. Consistent with principles of *Emotional Processing Theory* (Foa et al., 2007), we model imaginal exposure as a process of reactivating the trauma memory, fully experiencing the memory and its emotional content as though the memory is happening again in the present, and then re-encoding the memory in association with the new, safe and supportive context (such as the therapy office).

Thus, in Simulation 8, we repeated the methods from Simulation 6 (Intrusive Memories), including calculating mood as a composite measure of both negative and positive emotion, which we consolidate for conciseness in graphing the results. 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. That is, other than the stimuli introduced as part of an intervention, all other stimuli were neutral. 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. To reduce the risk of "overly positive" results in the Imaginal Exposure condition, we set the positive event to have the same low intensity (i.e., level of arousal) as all other non-traumatic stimuli. Thus, this event had positive valence and $\phi_{emot} = 1.0$. 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 every 20 lists during the neutral period, to match exactly the intervals at which virtual subjects experienced the positive event and accompanying imaginal exposure exercise in the Imaginal Exposure condition (described above) 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 properties of the event that preceded each of the re-encodings – that is, to verify whether re-encoding the trauma memory was a necessary part of the treatment. Therefore, we matched the properties of the positive event in the Positive Events Control Condition to those of the positive event in the Imaginal Exposure condition, such that $\phi_{emot} = 1.0$, and all other parameters ($\gamma_{emot}$, $\beta_{emot}$, and so forth) consisted of the same parameters for each virtual subject across all prior simulations (and obtained in Simulation 2). 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, tested our model's predictions of how repeated exposure to the memory of a traumatic event influences the negative emotion that is evoked by that memory during recall. First, we presented virtual subjects with a high-arousal, negative event (trauma-event) during their Developmental Period. Then, we presented four conditions during the Neutral Period, to test whether the conditions under which retrieved-context theory predicts whether negative emotional responding to trauma memories will reduce: (1) spontaneously, with the passage of time and neutral events (Baseline Condition); (2) in response to new positive events (Positive Control Condition); (3) due to re-activating and re-encoding the memory, but without any changes in the patient's current emotional context (Re-encoding Condition); or (4) upon re-encoding the trauma memory in a safe emotional context, as might occur in a therapy session (Imaginal Exposure Condition).

The model-predicted mood for an individual virtual subject are shown in Figure 15, and the aggregated effects across virtual subjects are shown in 16. During the Baseline Condition, trauma-memory intrusions continued to evoke high levels of negative emotion throughout the neutral period (Figure 15, Panel A; Figure 16). During the Positive Control Condition, the new positive events briefly instated a positive emotional context when it occurred, but the trauma-memory continued to intrude and evoke its associated negative

emotion, thus returning mood to a negative baseline each time (Figure 15, Panel B; Figure 16). In the Re-encoding Condition, re-encoding the trauma memory without altering the contents of its associated emotional context also did not result in mood improvements (Figure 16). Finally, in the Imaginal Exposure condition, as virtual subjects reactivated the trauma memory and then reencoded it in association with less-negative contexts, each subsequent reactivation of the trauma memory evoked progressively reduced negative emotion, until mood stabilized near a neutral value of 0.0 (Figure 15, Panel C; Figure 16).

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. Thus, no condition decreased the frequency of trauma-memory intrusions, suggesting that imaginal exposure in PE may primarily serve to reduce negative affect associated with the trauma-memory, whereas in vivo exposure to trauma-cues may serve to reduce trauma-memories' intrusions, as found in Simulation 6.

In the Imaginal Exposure Condition, the classic finding of reduced negative emotion, or apparent habituation, across rather than within sessions of imaginal exposure (Jaycox et al., 1998; M. Craske et al., 2008) emerged naturally from the model. This effect can be observed in Figure 15 (Panel B), and in Figure 16. In retrieved-context theory, the trauma memory reinstates its old associated context when first retrieved, thus evoking strong negative emotion. Within the session, the memory forms associations to the new safe context, updating the trauma network in the patient's episodic memory. However, this updated context will not be evoked until the next time the memory is retrieved, resulting in reduced negative affect across, but not within, sessions. By proposing that this phenomenon occurs due to updating in the episodic memory system, retrieved-context theory provides a novel account for emerging findings that imaginal exposure in PE can

also reduce negative trauma-associated emotion across sessions for patients whose primary trauma-associated emotion is an emotion other than fear, such as guilt, shame, or grief (Brown et al., 2019).

## General Discussion

Learning and memory have long been implicated as core processes underlying emotional disorders (Beck et al., 1979). Here, we present a novel theory of how episodic memory processes contribute to memory and mood. Memories of life's events do not exist in isolation. Rather, our brain organizes memories based upon the time and place in which they occur, as well as their common perceptual, semantic, and emotional attributes. Theories of memory have long recognized the important organizational role of such spatiotemporal context in defining and organizing memories of personal experiences (Kahana, 2012). In retrieved-context theory, a person's context is a mental representation of recently experienced external and internal stimuli, which evolves over time in response to new experiences. Context associates with the features of new memories as they form, such that remembering an event will tend to reactivate its previously associated contexts. In turn, these reactivated internal contexts can cue retrieval of other memories. Context not only cues recall of memories, but further, incorporates the elements of reactivated memories, which then associate with subsequent stimuli in a *recursive process* (see, Kahana, 2020 for a review).

We propose a retrieved-context theory of memory in emotional disorders. Building upon Talmi et al's (2019) model of memory and emotion, we also assume that context includes the emotional attributes of an experience. However, rather than representing emotion as synonymous with arousal (Talmi et al., 2019), we model distinct mood states in emotional disorders by differentiating positive and negative emotions, as well as allowing the model to represent mixed emotional states. We further extend the model by enabling the accrual of prior experiences across a person's full lifespan, rather than in a single

experimental list (Lohnas et al., 2015). This enables our CMR3 model to capture the dynamics of unwanted, intrusive memories from long-ago life contexts, as in PTSD.

Consistent with efforts to ground clinical theory in basic cognitive processes (Insel et al., 2010; Sanislow et al., 2010), our model does not predict the development of a *category* of disorder, such as Major Depressive Disorder (MDD) or Posttraumatic Stress Disorder (PTSD). Rather, our model describes the underlying basic processes in episodic memory that generate negative mood and memory disturbances, as transdiagnostic symptoms.

**Mood-congruent and Emotion-state dependent recall.**

We first show how our CMR3 model predicts the *emotional clustering* of episodic memories. Subjects tend to recall emotional items in same-valence clusters, such that recalling a negative item leads to subsequently recalling another negative item, and recalling a positive item leads to subsequently recalling another positive item (see Experiment 1, and also Siddiqui & Unsworth, 2011; Long et al., 2015). In CMR3, clusters of same-valence memories arise during recall because, for example, recalling a negative item evokes negative emotional context, which then cues the recall of other items associated to negative contexts. This is an extension of the general principle in retrieved-context models, that similarity between encoding and retrieval contexts guides recall. Due to this mechanism, the CMR3 model also predicts both heightened recall of memories whose emotional properties match the current mood state (Mood-congruent Recall; Simulation 3), as well as heightened recall of even neutral memory material if it was encoded against an emotional backdrop that matches the person's current emotional state (Emotion-State Dependent Recall; Simulation 4). That is, neutral memory content will be better recalled if the emotional context present just prior to and during encoding matches the emotional context present during retrieval.

**Persistent Negative Mood**

Retrieved-context theory predicts that emotional context can linger after an event and bind to new events that follow, even if these new events are neutral (Simulation 4). Thus, negative emotional context can generalize throughout the memory network and reactivate when recalling neutral, or even positive, events. Increasing the ratio of positive to negative events, as in Behavioral Activation Therapy (Beck et al., 1979; Jacobson et al., 1996; Lewinsohn et al., 1980), evokes positive contexts that can also bind and generalize throughout the memory system, thus improving patients' mood (Simulation 4, Treatment Period).

However, for patients experiencing *intrusive* negative memories, our model predicted that positive event scheduling will briefly evoke positive emotion, but that the memory-intrusions will persistently reactivate their associated negative contexts, eventually blunting responsiveness to positive events (Simulation 5). Thus, our model corroborates the importance of targeting intrusive memories in treatments not only in PTSD, but also alleviating for negative mood in major depression (Brewin et al., 2009).

Encoding an event in multiple contexts associates its memory with a larger number of potential retrieval cues (Lohnas, Polyn, & Kahana, 2011; Madigan, 1969; Melton, 1970). Thus, repeatedly experiencing the same stressful event, or ruminating over the event in varied contexts, will increase the retrieval routes to that memory and thus its likelihood of spontaneous subsequent recall (Simulation 5). Indeed, rumination, which includes the repetitive rehearsal and elaboration upon negative memories, is associated with and can prompt intrusive memories in PTSD (Michael et al., 2007).

**Intrusive memories**

Intrusive memories of negative events 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; A. D. Williams & Moulds, 2007). Retrieved-context theory proposes a general mechanism for intrusive memories, in which encountering stimuli whose features resemble the perceptual, spatial, and temporal context that directly preceded a negative event will cue intrusive memories of that event (Simulation 6), even if the stimuli have no semantic relationship to the event, as is common in PTSD (Ehlers & Clark, 2000; Ehlers et al., 2004). Drawing on work by Talmi et al. (2019), we additionally propose that high emotional arousal during a traumatic event leads to stronger encoding of associations between the traumatic memory and its context. These strengthened contextual associations increase the probability that unrelated contexts will trigger the memory (see Simulations 7 and 9).

Trauma-memory intrusions often occur with vivid imagery and a sense of "nowness," i.e., that the memory is happening again in the here-and-now (Hackmann, Ehlers, Speckens, & Clark, 2004). Retrieved-context theory conceptualizes a memory's vividness as how strongly the retrieval context activates that memory's features (Simulation 7). Drawing on the concept of "mental time travel," in which recalling an event leads people to experience themselves being transported back to the time of the event's occurrence (Tulving, 1983), we model the "nowness" of a memory as the similarity between the memory's encoding context and the patient's current context during retrieval (Simulation 7). However, mental time travel is but one possible framework for perceived "nowness." A different interpretation is that nowness involves a fading of the current context into the background, not a sense that the intrusion belongs in the current context, as we propose.

**Treating Intrusive Trauma Memories.**

Retrieved-context theory, as instantiated in CMR3, suggests two treatment routes for trauma-memory intrusions. The first route is modifying the context-to-memory associations in the emotion network, as in Prolonged Exposure Therapy (PE; Foa & Kozak,

1986; Foa et al., 2007). Associating trauma cues to neutral or non-traumatic contexts, as in *in vivo* exposure, increases the probability that these stimuli will cue nontraumatic memory content, and reduces the frequency of negative memory-intrusions (Simulation 6). Further, reactivating and then reencoding the trauma-memory in association with new, safe or neutral contexts as in *imaginal exposure*, even if the trauma-memory itself evokes strong negative affect against the backdrop of these less-negative contexts, decreases negative affect across sessions (Simulation 8).

The second route for alleviating intrusive memories involves modifying a patient's current contextual state to reduce its overlap with the trauma-memory's encoding context. Reduced negative emotion due to cognitive restructuring would reduce the tendency of mood-congruent recall processes to reactivate the trauma-memory and other negative memories (Simulation 2). This might occur through restructuring maladaptive cognitions and thereby reducing shame and guilt, as in Cognitive Processing Therapy (CPT; Resick & Schnicke, 1992), as well as during processing in PE.

## Individual differences in risk and resilience

Retrieved-context theory proposes specific episodic memory processes that serve as risk or protective factors. Fitting our model to a subject's memory patterns can identify the parameters that determine how strongly a given memory process is in play for an individual subject.

*Emotional Valence.* The $\beta_{emot}$ parameter governs how quickly a subject's internal context updates to take in new emotional content. Swift emotional updating, as might occur in emotion dysregulation, increases the level of emotional context associated with new memories as they form. High values of $\beta_{emot}$ would also increase the levels of negative emotion present in a subject's daily context, increasing the probability of "cluster recalls" of one negative memory after another (the *emotional clustering effect*; see Experiment 1), as well as heightened rates of negative-memory intrusions due to mood-congruent recall

(Simulation 6).

*Emotional Arousal.* As proposed by Talmi et al. (2019), negative arousal strengthens context-to-memory associations during the encoding of negative memories. Thus, higher values of $\phi_{emot}$, the scalar that modulates these associations, will enhance recall of this material and predispose trauma- and other negative memories to intrude spontaneously into unrelated contexts (Simulations 6-8).

*General Parameters.* Model parameters that address general memory function can also help to explain individual differences that predispose a person to develop psychopathology. For example, $c_{thresh}$ controls the degree to which subjects can exert cognitive control to filter and censor unwanted memory content, such as trauma-memory intrusions.

**Relation to existing theories.**

**Emotional Processing Theory.** Emotional processing theory (EPT) has generated a highly efficacious treatment for intrusive trauma memories in PTSD called Prolonged Exposure therapy (PE; Foa & Kozak, 1986; Foa et al., 2007; Powers et al., 2010). In EPT, the trauma-memory exists in a network of associations with other stimuli, maladaptive cognitions, and fear responses, called the *fear network* or *emotion network* (Foa & Kozak, 1986; Brown et al., 2019; Lane, Ryan, Nadel, & Greenberg, 2015). Treatment reactivates the memory in order to restructure this network, by exposing the patient to the feared stimulus (in imaginal exposure, this is the memory itself). Retrieved-context theory shares several core principles with EPT: in CMR3, fully reactivating the trauma memory is necessary to restructure its network of contextual associations, thereby reducing the frequency of trauma-memory intrusions (Simulation 6) and the patient's distress upon recalling the trauma-memory (Simulation 8).

However, retrieved-context theory differs from EPT regarding the role of habituation. EPT arose from early models of fear conditioning, including the classic Rescorla-Wagner

model (Rescorla & Wagner, 1972). In the Rescorla-Wagner model, presenting the feared stimulus in the absence of the feared outcome weakens and ideally breaks the stimulus-response relationship (Rescorla & Wagner, 1972). Accordingly, EPT proposes that staying present with the stimulus until habituation occurs within the session weakens the association between the stimulus and the conditioned fear response, such that the memory no longer activates the fear response.

Yet, in PE, within-session habituation to the trauma memory has weak or no associations with symptom reduction (Rupp, Doebler, Ehring, & Vossbeck-Elsebusch, 2017; Jaycox et al., 1998). Rather, symptom reduction is most strongly associated with the degree of habituation achieved across sessions. That is, PE patients often experience extreme distress while reactivating the trauma-memory, find that their distress does not reduce within the session, and yet later experience reduced distress while reactivating the memory in a new session. This finding has raised concerns about circularity: reduced negative affect across treatment sessions upon recalling the trauma memory is proposed as a core treatment mechanism, yet reduced negative affect across treatment sessions upon recalling the trauma memory is also a primary treatment outcome. Perhaps "those who habituate, habituate," or habituation is incidental to the success of exposure-based treatments (Brewin, 2006).

To address these issues, recent updates to EPT incorporate principles of *inhibitory learning*, proposing that exposure-based treatments introduce new, competing CS-noUS associations, rather than breaking the preexisting CS-US association (Brown et al., 2019). Inhibitory learning and the effects of expectation violations enhance and powerfully contribute to new therapeutic learning when present (M. G. Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014; M. G. Craske, Liao, Brown, & Vervliet, 2012; Brown, LeBeau, Chat, & Craske, 2017).

Retrieved-context theory offers an alternative mechanism of restructuring the episodic memory network. In CMR3, recalling the memory first reactivates its old emotional

context, and then re-encodes the memory in association with a new, less-negative emotional context. The newly updated emotional context is only reactivated the next time that the patient retrieves the memory, such that negative affect associated to the trauma memory decreases across, rather than within, sessions. As a result, reduced negative emotion across, rather than within, exposure sessions emerges naturally from our model (Simulation 8).

Within an individual session of PE, the trauma memory may be rehearsed multiple times. In CMR3, repeating the trauma-memory within sessions will have a weaker effect on the trauma-memory network than repetitions across sessions, because context remains relatively constant within an individual session. This is a clinical application of the *spacing effect*, in which spaced rather than massed repetitions enhance learning by associating repeated experiences with more diverse contexts (Lohnas et al., 2011; Madigan, 1969; Melton, 1970).

A further consideration is that, although EPT is grounded in fear-conditioning principles of habituation and inhibitory learning, PE is still highly efficacious in treating PTSD for patients whose primary emotional response to their traumatic event involved an emotion other than fear, such as shame, guilt, or sadness (Brown et al., 2019), which is common for PTSD patients (Reynolds & Brewin, 1999; Resick & Schnicke, 1992). Our model proposes that in episodic memory, the emotional context associated with an episodic memory trace can incorporate an amalgam of different emotional responses, including but not limited to fear. Thus, reactivating and reencoding the trauma's episodic memory trace in association with novel contexts would still generate the clinical pattern of heightened negative emotion within a session, but reduced negative emotion in subsequent sessions, even for non-fear emotions. Accordingly, our model proposes a generalized framework, in which heightened negative arousal during encoding strengthens a memory's contextual associations, and thereby results in intrusive memories of negative events in patients with depression and other emotional disorders, not just PTSD (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).

Overall, retrieved-context theory generates several novel predictions for exposure-based treatments. First, even when patients confirm their expectations that, upon reactivating a trauma-memory, they will both experience extreme distress and even tolerate that distress poorly, repeated exposures will still result in symptom improvement. Second, the duration of within-session exposure to the memory does not affect the degree of subsequent reduction in negative affect, as long as the memory is fully reactivated for re-encoding, and the background context for reencoding is sufficiently "non-negative" emotionally (note that emotional content of the surrounding context for reencoding, such as the emotional tone of the therapist's office, is distinct from the emotion evoked by the trauma-memory). Third, PE therapy for PTSD will still be efficacious in treating trauma-memory intrusions even when the primary emotion associated with the trauma is not fear. Finally, the principles of reactivating and reencoding distressing memories, which PE pioneered for the treatment of trauma memories in PTSD, may also be effective for alleviating distress and memory-intrusions resulting from painful events in depression and other disorders.

**Reconsolidation.** *Reconsolidation* theories of change during imaginal exposure in PE propose that symptoms improve by updating and re-encoding the trauma memory to include new emotional elements (Lane et al., 2015). Our model does not propose a memory reconsolidation process. Retrieved-context theory proposes that reactivating and reencoding a memory does not change its contents, but rather, associates the memory to novel contexts. Accordingly, our model predicts that imaginal exposure will reduce negative affect evoked by a traumatic memory even if the contents of the memory itself do not change during treatment.

**Does PTSD reflect specialized memory processes?** Our model speaks to the longstanding controversy over whether memories of traumatic events operate according to specialized memory mechanisms (Brewin, 2014; Bisby, Horner, Bush, & Burgess, 2018;

Bisby, Burgess, & Brewin, 2020) or whether the unique qualities of trauma memories arise from the basic processes of a unitary memory system (Rubin et al., 2008, 2011).

Retrieved-context theory proposes that trauma memories and other autobiographical memories arise from common factors. As in the Autobiographical Memory Theory of PTSD (Rubin et al., 2011), our model proposes that emotional intensity, repetition, and rehearsal strengthen the accessibility of a memory and thereby predispose it to spontaneously activate. In our model, contextual associations in episodic memory serve as the core process underlying these principles – heightened emotional intensity strengthens context-to-item associations (Simulations 6-8, and see Talmi et al., 2019); repetition and rehearsal associates the memory to a wider array of contexts that can serve as retrieval cues (Simulation 5, and see Lohnas et al., 2011; Madigan, 1969; Melton, 1970); and the *semantic contiguity effect* generates heightened recall of memories that have strong semantic relations with other memories and with the current context (Simulation 1, and see Howard & Kahana, 2002b), potentially enhancing recall for memories that have meaningful relatedness to person's life story.

The dual-representation theory (DRT) of PTSD proposes that high-arousal negative emotion at encoding modulates the strength of associations between traumatic memories and their encoding contexts (Brewin, 2014; Brewin et al., 2010). Specifically, DRT proposes that high arousal negative emotion disrupts the associations between events' perceptual features and their spatiotemporal contexts (Brewin, 2014; Bisby et al., 2018). As a result, DRT proposes that trauma memories have unique properties that differ from those of other nontraumatic negative episodic memories. Further, DRT proposes that reunification of the memory trace - reassociating events' perceptual features to their contextual elements - is needed to reduce intrusive memories in PTSD (Brewin, 2014).

When high-arousal negative emotion occurs, particularly in the absence of dissociation which both theories predict would interfere with new learning, dual-representation and retrieved-context theories predict different outcomes for

exposure-based treatments. During exposure exercises, patients face feared stimuli, including feared memories, and are encouraged to fully experience their distress. Heightened levels of negative emotion and negative physiological arousal during exposures are either associated with better treatment outcomes (M. Craske et al., 2008; Kircanski et al., 2012; A. J. Lang & Craske, 2000; Jaycox et al., 1998), or else have no relation with treatment outcomes (Rupp et al., 2017). In retrieved-context theory, high-arousal negative emotion during exposure is beneficial, by more strongly binding the memory to the new, therapeutic context.

Dual-representation theory predicts that high-arousal negative emotion will disrupt the binding of items-to-contexts during new learning, even for non-personally relevant stimuli in a laboratory setting (Bisby et al., 2018). Accordingly, the negative emotion present during an even more emotionally evocative, imaginal exposure to the memory of prior trauma would be an unhelpful, or even harmful, byproduct of the treatment's true active ingredient: re-unifying the memory trace. Dual-representation theory predicts that Prolonged Exposure Therapy for PTSD and other exposure-based therapies will still be efficacious, as long as the high-arousal negative emotion is carefully modulated during session, thus preventing it from interfering with the goal of re-unifying the reactivated trauma memory (Bisby et al., 2018).

In further support of dual-representation theory, laboratory studies have found weaker associations during recall between negative and neutral stimuli than between pairs of neutral stimuli, which supports a dual-process account (Bisby & Burgess, 2014; Bisby et al., 2018). However, in Experiment 1, we replicated the reduced tendency of emotional items to cue subsequent recall of differently-valent items as part of the *emotional clustering effect*. We demonstrate that our model predicts this effect through the mechanism of emotional context similarity (Simulation 1).

In addition, these studies did not control for greater semantic relatedness among pairs of negative (vs. pairs of neutral) stimuli (Talmi & Moscovitch, 2004). Consequently,

this pattern of results may be accounted for by the classic “fan effect" (Anderson, 1974). Because neutral items have weaker semantic associations to other neutral items, as well as to items of negative or positive valences, the associations that neutral items form to other neutral items during the experiment will have less competition from prior, long-standing associations. As a result, negative items have stronger pre-experimental associations to other negative items that will compete for retrieval, and thus impair recall, of new associates learned during the experimental context. Nonetheless, this explanation is just one of many possibilities, and further empirical testing is needed to actually establish greater semantic relatedness of negative items as the underlying mechanism in Bisby and colleagues’ work.

**Spreading Activation Theory.** In *spreading activation theory*, Bower (1981) proposed that mood-congruent and emotion-state dependent memory occur because emotional valence serves as a 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. However, in retrieved-context theory, 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 allows our theory to model the generalization of negative affect to neutral events (Simulation 4). Emotional context also allows our model to account for how stimuli that have little semantic association to a traumatic event, but that preceded the event temporally, can later cue intrusive memories and trauma-associated affect (Simulation 6, and see Ehlers & Clark, 2000).

## Limitations

The utility of our model depends on whether list-learning experiments can be considered to be an adequate basis for a model of human memory. This is the core challenge for all modeling work: finding a paradigm simple enough to be tractable for

developing a model in its early stages, yet complex enough to ultimately extend its predictions to multidimensional, real-world experience. As in classic work by George Box, “All models are wrong, but some are useful" (Box, 1979). Certainly, learning and recalling lists of words in the lab, and testing subjects’ abilities to localize their recalls to a specific list context, lacks the complexity and multimodality of real-world emotional events.

At the same time, the list-learning paradigm has served as a powerful tool for isolating core principles of human memory, including primacy, recency, temporal and semantic contiguity effects (Kahana, 1996; Murdock, 1967; Howard & Kahana, 2002b; Sederberg, Miller, Howard, & Kahana, 2010), which have replicated in studies of memory for complex events in the real world (Jansari & Parkin, 1996; Loftus & Fathi, 1985; Moreton & Ward, 2010; Uitvlugt & Healey, 2019; Healey et al., 2019). We view the model’s current setup just as a needed starting point for more realistic models of human experience and suggest that future work based on ecological momentary assessment could lead to deeper insights into the mechanisms of memory for emotional events, and particularly, how this process unfolds in emotional disorders in patients’ natural environments.

Further, a core phenomenon shared by both major depression and PTSD is overgeneral memory, likely related to or perhaps a consequence of deficits in context discrimination abilities (J. M. G. Williams et al., 2007). In our model, experiencing repeated negative events, or multiple negative events that all share some common categorical information, across varied temporal contexts, will tend to form a web of associations across lists, much as in an experiment in which subjects experience repeated items from the same categories across lists (e.g., Postman, 1976; Anderson & Bower, 1972). Such manipulations cause high levels of associative interference. If the negative categories constitute the repeated items, reactivation of these categories will lead to higher levels of recall, whereas associative interference will lead to lower levels of event-specific recall. This is analogous to the effects of repetition in word lists (Zaromb et al., 2006; Miller, Weidemann, & Kahana, 2012). Although this mechanism could produce overgeneral

memory in depressed patients, a proper modeling exercise would require experimental data in which depressed patients study lists of category-exemplar pairs of varying valence/arousal and show improved memory for the categories but impaired memory for the examples. We suggest this as an interesting direction for future work.

Another limitation lies in our decision to model emotion as varying along dimensions of valence and arousal. This allows the model to account for patterns of comorbidity between anxiety, depression, and other emotional disorders, as well as how intrusive memories can occur for many types of distressing events. However, emotion is a complex, multivariate construct. For example, anger and fear are both negatively valent and high-arousal emotions, and thus would be indistinguishable in the current CMR3 model, yet they have distinct subjective experiences, action tendencies (avoid vs. approach; Carver & Harmon-Jones, 2009), and neural patterns (Kragel & LaBar, 2015). Future empirical work is needed to shed light on whether specific emotions, beyond dimensional features such as valence and arousal, have distinct influence on memory, to inform how best to represent emotion in this model.

In addition, not all negative emotions induce high arousal, which our model incorporates from eCMR (Talmi et al., 2019) and proposes as a core mechanism for intrusive memories. 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. Similarly, further research is needed to determine the precise relationship between arousal and the learning rate for new associations during encoding. One could easily introduce a non-linear relation between arousal and learning rate, as in the classic Yerkes-Dodson (1908) results, and perhaps this would be a fruitful direction for future work. Or, perhaps under extreme levels of arousal, there is a more systemic shutdown of the cognitive system, as suggested by dual-representation theorists.

Further, processes of cognitive control play an important role in autobiographical

memories. One of the key elements of the CMR3 model is its ability to use context-similarity as a filter to suppress memories that are contextually distant from the target of conscious awareness or attention, operationalized via the $c_{thresh}$ parameter. This modeling of cognitive control in the memory search process was first introduced into the CMR family of models in CMR2 (Lohnas et al., 2015) and is crucial to modeling the reactivation and filtering of trauma-memory intrusions. However, the model's implementation of this mechanism does not fully specify all possible sources of cognitive control that may be used by subjects in guiding their thought process. Further, the companion question, to what extent are people able to control this memory process, is not known and remains an important area for future research.

Similarly, the CMR3 model can incorporate mental content, such as background thoughts and emotions, as part of the context of an event and its subsequent memory representation. However, everyday thoughts and other forms of abstract cognition may operate according to different processes than episodic memory (Moscovitch, 1992). Determining how to operationalize that difference in retrieved-context models will be a promising direction for further research.

Finally, CMR3 represents mood and memory outcomes as the result of mental processes constrained by certain parameter values, which determine how strongly each process will operate for a given individual. However, the model does not explain what led a patient to have 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 inherit and acquire distinct cognitive parameters.

## Conclusion

Retrieved-context theory proposes a parsimonious account of episodic memory that generates negative mood, 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. Together with the emerging field of computational psychiatry (Bennett, Silverstein, & Niv, 2019; Montague, Dolan, Friston, & Dayan, 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 development of symptoms, their maintenance factors, and 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. Our simulations link clinical findings with cognitive operations and lay the groundwork for future work connecting these processes with neural mechanisms. We hope this framework will be used to both refine and challenge our retrieved-context model, thus ultimately improving our understanding of memory and mood.

References

Aka, A., Phan, T. D., & Kahana, M. J. (2020). Predicting recall of words and lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

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

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6(4)*, 451-474.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psycholgical Review*, *79*(2), 97-123. doi: 10.1037/h0033773

Austin, M. P., Mitchell, P., & Goodwin, G. M. (2001). Cognitive deficits in depression: possible implications for functional neuropathology. *The British Journal of Psychiatry*, *178*, 200-206.

Barrett, L. F. (1998). Discrete emotions or dimensions? the role of valence focus and arousal. *Cognition and Emotion*, *12*(4), 579-599.

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York, NY: 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 health survey consortium. *Psychological Medicine*, *46*, 327-343.

Bennett, D., Silverstein, S. M., & Niv, Y. (2019). The two cultures of computational psychiatry. *JAMA Psychiatry*, *76*, 563-564.

Berntsen, D. (1996). Involuntary autobiographical memories. *Applied Cognitive Psychology*, *10*(5), 435-454.

Berntsen, D. (2010). The unbidden past: Involuntary autobiographical memories as a basic

mode of remembering. *Current Directions in Psychological Science*, *19*, 138-142.

Berntsen, D., & Rubin, D. C. (2002). Emotionally charged autobiographical memories across the life span: The recall of happy, sad, traumatic and involuntary memories. *Psychology and Aging*, *17*(4), 636–652.

Bevington, P. R., & Robinson, D. K. (2003). *Data reduction and error analysis for the physical sciences* (3rd ed.). New York, NY: McGraw-Hill Higher Education.

Bisby, J. A., & Burgess, N. (2014). Negative affect impairs associative memory but not item memory. *Learning & Memory*, *21*(1), 760–766.

Bisby, J. A., Burgess, N., & Brewin, C. R. (2020). Reduced memory coherence for negative events and its relationship to posttraumatic stress disorder. *Current Directions in Psychological Science*, *29*(3), 267-272.

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*, 243-256.

Bonanno, G. A., Westphal, M., & Mancini, A. (2011). Resilience to loss and potential trauma. *Annual Review of Clinical Psychology*, *7*, 511-535.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*(2), 229-240.

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

Bower, G. H., & Mayer, J. D. (1985). Failure to replicate mood-dependent retrieval. *Bulletin of the Psychonomic Society*, *23*, 39-42.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. *Robustness in Statistics*, 201-236.

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*, 755-770.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering

pictures: pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 379.

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.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

Bremner, J. D., Vythilingam, M., Vermetten, E., Vaccarino, V., & Charney, D. S. (2004). Deficits in hippocampal and anterior cingulate functioning during verbal declarative memory encoding in midlife major depression. *American Journal of Psychiatry*, *161*, 637-645.

Brewin, C. R. (2006). Understanding cognitive behaviour therapy: A retrieval competition account. *Behavior Research and Therapy*, *44*(6), 765-784. doi: 10.1016/j.brat.2006.02.005

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

Brewin, C. R., Andrews, B., Rose, S., & Kirk, M. (1999). Acute stress disorder and posttraumatic stress disorder in victims of violent crime. *American Journal of Psychiatry*, *156*, 360-366.

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*, 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*, 569-576.

Briere, J., Kaltman, S., & Green, B. L. (2008). Accumulated childhood trauma and symptom complexity. *Journal of Traumatic Stress*, *21*, 223-226.

Brown, L. A., LeBeau, R. T., Chat, K. Y., & Craske, M. G. (2017). Associative learning versus fear habituation as predictors of long-term extinction retention. *Cognition and Emotion*, *31*(4), 687-698.

Brown, L. A., Zandberg, L. J., & Foa, E. B. (2019). Mechanisms of change in prolonged exposure therapy for ptsd: Implications for clinical practice. *Journal of Psychotherapy Integration*, *29*(1), 6-14.

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

Cabib, S., & Puglisi-Allegra, S. (1996). Stress, depression and the mesolimbic dopamine system. *Psychopharmacology*, *128*, 331-342.

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*, 401-423.

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: evidence and implications. *Psychological Bulletin*, *135*(2), 183-204.

Clark, D., & Beck, A. (1999). *Scientific foundations of cognitive theory and therapy of depression*. New York, NY: John Wiley & Sons.

Cloitre, M., Stolbach, B. C., Herman, J. L., Kolk, B. V. D., Pynoos, R., Wang, J., & Petkova, E. (2009). A developmental approach to complex ptsd: Childhood and adult cumulative trauma as predictors of symptom complexity. *Journal of Traumatic Stress*, *22*, 399-408.

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.

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

therapy. *Journal of Experimental Psychopathology*, *3*(3), 322-345.

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

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Lawrence Erlbaum and Associates.

Day, S. J., Holmes, E. A., & Hackmann, A. (2004). Occurrence of imagery and its link with early memories in agoraphobia. *Memory*, *12*, 416-427.

Depue, R. A., & Monroe, S. M. (1986). Conceptualization and measurement of human disorder in life stress research: The problem of chronic disturbance. *Psychological Bulletin*, *99*, 36-51.

Ehlers, A., & Clark, D. (2000). A cognitive model of posttraumatic stress disorder. *Behavior Research and Therapy*, *38*(4), 319-345. doi: 10.1016/S0005-7967(99)00123-0

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

Ehlers, A., & Steil, R. (1995). Maintenance of intrusive memories in posttraumatic stress disorder: A cognitive approach. *Behavioural and Cognitive Psychotherapy*, *23*, 217-249.

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., 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., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, *99*, 20-35.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology*, *66*, 325-331.

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*, 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.

Hackmann, A., Ehlers, A., Speckens, A., & Clark, D. M. (2004). Characteristics and content of intrusive memories in ptsd and their changes with treatment. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, *17*(3), 231-240.

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. doi: 10.1037/a0033715

Healey, M. K., & Kahana, M. J. (2016). A four–component model of age–related memory change. *Psychological Review*, *123*(1), 23-69. doi: 10.1037/rev0000015

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*, 92-104. doi: 10.3758/s13423-018-1537-3

Herman, J. L. (1992). Complex ptsd: A syndrome in survivors of prolonged and repeated trauma. *Journal of Traumatic Stress*, *5*, 377-391.

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. doi: 10.1006/jmps.2001.1388

Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic

retrieval? *Journal of Memory and Language*, *46*, 85–98.

Howard, M. W., Shankar, K. H., Aue, W. R., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, *122*(1), 24-53.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., 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., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., ... Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*, 295-304.

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.

Jansari, A., & Parkin, A. J. (1996). Things that go bump in your life: Explaining the reminiscence bump in autobiographical memory. *Psychology and Aging*, *11*, 85-91.

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*, 185-192.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. doi: 10.3758/BF03197276

Kahana, M. J. (2012). *Foundations of human memory.* New York, NY: Oxford University Press.

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, *71*, 107-138.

Kahana, M. 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. doi: 10.1037/xlm0000553

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic

recall: Age differences in editing of overt responses. *Journal of Gerontology: Psychological Sciences*, *60*(2), 92–97. doi: 10.1093/geronb/60.2.P92

Kahana, M. J., Zhou, F., Geller, A. S., & Sekuler, R. (2007). Lure similarity affects visual episodic recognition: Detailed tests of a noisy examplar model. *Memory & Cognition*, *35*(6), 1222–1232.

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*, 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*, 833-842.

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*, 745-751.

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*, 1437-1448.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *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.

Lang, A. J., & Craske, M. G. (2000). Manipulations of exposure-based therapy to reduce return of fear: A replication. *Behaviour Research and Therapy*, *38*, 1-12.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50*, 372-385.

Lewinsohn, P. M., Sullivan, J. M., & Grosscap, S. J. (1980). Changing reinforcing events: An approach to the treatment of depression. *Psychotherapy: Theory, Research, and Practice.*, *17*, 322-334.

Loftus, E. F., & Fathi, D. C. (1985). Retrieving multiple autobiographical memories. *Social Cognition*, *3*, 280-295.

Lohnas, L. J., Polyn, S. M., & Kahana, M. (2015). Expanding the scope of memory search: Intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337-363. doi: 10.1037/a0039036

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language*, *64*(3), 249-255. doi: 10.1016/j.jml.2010.11.003

Long, N. M., Danoff, M. S., & Kahana, M. J. (2015). Recall dynamics reveal the retrieval of emotional context. *Psychonomic Bulletin and Review*, *22*(5), 1328-1333. doi: 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. doi: 10.3758/s13428-011-0124-6

Mathews, A., Mogg, K., May, J., & Eysenck, M. (1989). Implicit and explicit memory bias in anxiety. *Journal of Abnormal Psychology*, *98*, 236-240.

Matt, G. E., Vazquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, *12*, 227-255.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Memory*, *9*, 596-606.

Messman-Moore, T. L., Walsh, K. L., & DiLillo, D. (2010). Emotion dysregulation and risky sexual behavior in revictimization. *Child Abuse & Neglect*, *34*(12), 967-976.

Michael, T., Halligan, S. L., Clark, D. M., & Ehlers, A. (2007). Rumination in posttraumatic stress disorder. *Depression and Anxiety*, *24*, 307-317.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781v3*.

Miller, J. F., Weidemann, C. T., & Kahana, M. J. (2012). Recall termination in free recall. *Memory & Cognition*, *40*, 540–550.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, *16*, 72-80.

Mor, N., & Winquist, J. (2002). Self-focused attention and negative affect: a meta-analysis. *Psychological Bulletin*, *128*, 638-662.

Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review*, *17*(4), 510–515.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience*, *4*(3), 257-267.

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.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482-488. doi: 10.1037/h0045106

Murdock, B. B. (1967). Recent developments in short-term memory. *British Journal of Psychology*, *58*(3-4), 421-433. doi: 10.1111/j.2044-8295.1967.tb01099.x

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, Instruments and Computers*, *36(3)*, 402-407. doi: 10.3758/BF03195588

Nelson, R. E., & Craighead, W. E. (1977). Selective recall of positive and negative

feedback, self-control behaviors, and depression. *Journal of Abnormal Psychology*, *86*(4), 379-388.

Osman, S., Cooper, M., Hackmann, A., & Veale, D. (2004). Spontaneously occurring images and early memories in people with body dysmorphic disorder. *Memory*, *12*, 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. doi: 10.1037/a0014420

Postman, L. (1976). Interference theory revisited. In J. Brown (Ed.), *Recall and recognition* (p. 157-181). Oxford, England: John Wiley and Sons.

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*, 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.

Rauch, S., & Foa, E. (2006). Emotional processing theory (ept) and exposure therapy for ptsd. *Journal of Contemporary Psychotherapy*, *36*, 61-65.

Rescorla, R., & Wagner, A. (1972). A theory of pavolvian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). New York: Appleton Century Crofts.

Resick, P. A., & Schnicke, M. K. (1992). Cognitive processing therapy for sexual assault victims. *Journal of Consulting and Clinical Psychology*, *60*, 748-756.

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

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 803-814. doi: 10.1037/0278-7393.21.4.803

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, *4*, 28-34.

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*, 985-1011.

Rubin, D. 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*, 840-856.

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.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178.

Sanislow, C., 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. *Journal of Abnormal Psychology*, *119*, 631-639.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6 (2)*, 461-464.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912. doi: 10.1037/a0013396

Sederberg, P. B., Miller, J. F., Howard, W. H., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, *38*(6), 689–699. doi: 10.3758/MC.38.6.689

Siddiqui, A. P., & Unsworth, N. (2011). Investigating the role of emotion during the search process in free recall. *Memory & Cognition*, *39*(8), 1387–1400.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonomic Bulletin and Review*, *8*(2), 203-220.

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.

Staugaard, S. R., & Berntsen, D. (2014). Involuntary memories of emotional scenes: The effects of cue discriminability and emotion over time. *Journal of Experimental Psychology: General*, *143*(5), 1939-1957.

Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, *126*, 455-485.

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, *32*(5), 742–751.

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*, 248-257.

Teasdale, J. D., & Russell, L. M. (1983). Differential effects of induced mood on the recall of positive, negative and neutral words. *British Journal of Clinical Psychology*, *22*, 163-171.

Tran, U. S., & Gregor, B. (2016). The relative efficacy of bona fide psychotherapies for post-traumatic stress disorder: A meta-analytical evaluation of randomized controlled trials. *BMC Psychiatry*, *16*(266).

Tull, M. T., Barrett, H. M., McMillan, E. S., & Roemer, L. (2007). A preliminary investigation of the relationship between emotion regulation difficulties and posttraumatic stress symptoms. *Behavior Therapy*, *38*(3), 303-313.

Tull, M. T., Berghoff, C. R., Wheeless, L. E., Cohen, R. T., & Gratz, K. L. (2018). Ptsd symptom severity and emotion regulation strategy use during trauma cue exposure among patients with substance use disorders: Associations with negative affect, craving, and cortisol reactivity. *Behavior Therapy*, *49*(1), 57-70.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* (p. 381-403). New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford.

Uitvlugt, M. G., & Healey, M. K. (2019). Temporal proximity links unrelated news events in memory. *Psychological Science*, *30*, 92-104.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Watkins, E., & Teasdale, J. D. (2004). Adaptive and maladaptive self-focus in depression. *Journal of Affective Disorders*, *82*, 1-8.

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. C., 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.

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.

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., &

Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*, 122-148.

Willner, P. (1997). Validity, reliability and utility of the chronic mild stress model of depression: A 10-year review and evaluation. *Psychopharmacology*, *134*, 319-329.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459-482.

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. doi: 10.1037/0278-7393.32.4.792

Table 1
*Transition probabilities regressed on initial-item valence and next-item valence.*

| | | $\beta$ | $SE$ | $t$ | LR$\chi^2(df)$ |
|---|---|---|---|---|---|
| Step 0 | | | | | |
| | Intercept | .070*** | .002 | 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
*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.

Table 3
*Parameter differences across simulations*

| | $\beta_{emot}$ | $\phi_{emot}$ |
|---|---|---|
| Simulation 1 | * | 1.0 |
| Simulation 2 | * | 1.0 |
| Simulation 3 | * | 1.0 |
| Simulation 4 | * | 1.0 |
| Simulation 5 | * | 1.0 |
| Simulation 6 | * | 5.0 |
| Simulation 6b (EDR) | .3 (Low EDR), .8 (High EDR) | 5.0 |
| Simulation 7 | * | 2.0-5.0 |
| Simulation 8 | * | 50.0 |

Note. This table displays values only for those parameters that we manipulate across simulations. Otherwise, the parameters for each virtual subject consist of the values obtained from the model fits to data from the individual real-life subjects in Simulation 1 (see Table A1). The entry * indicates that this parameter also held one of the parameter values from these 97 model fits to individual-subjects. EDR = Emotion dysregulation. $\beta_{emot}$ = drift rate for emotional context. $\phi_{emot}$ = a scalar that magnifies the strength of context-to-item associations during the encoding of a high-arousal negative item.

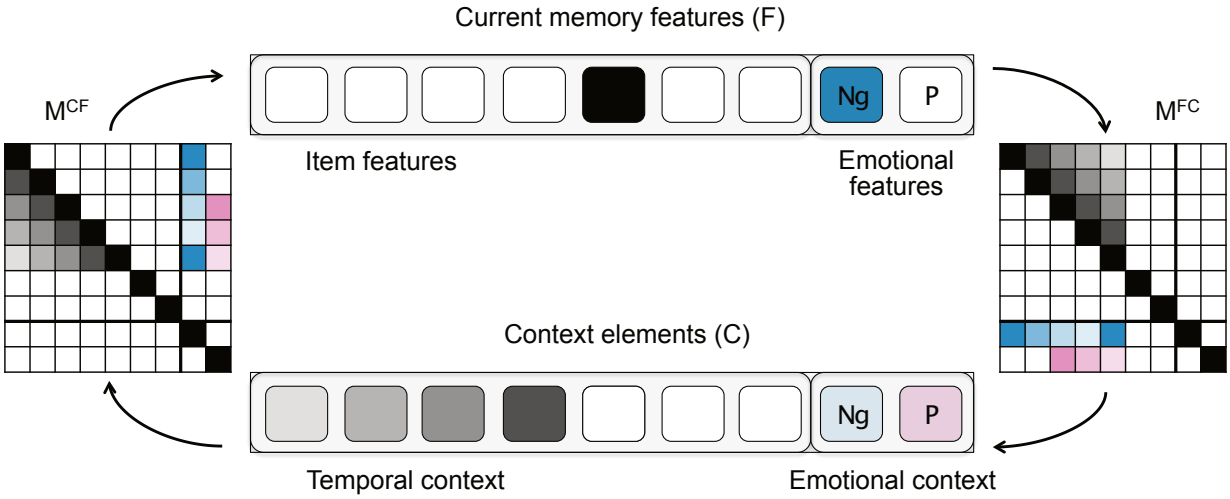


*Figure 1. Schematic of how CMR3 models memory encoding and retrieval.* The vector F represents event features comprising both item and emotional content. Item features are either present (have a value of1.0, shown in black) or absent (have a value of 0.0, shown in white). Valence is encoded by setting 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.

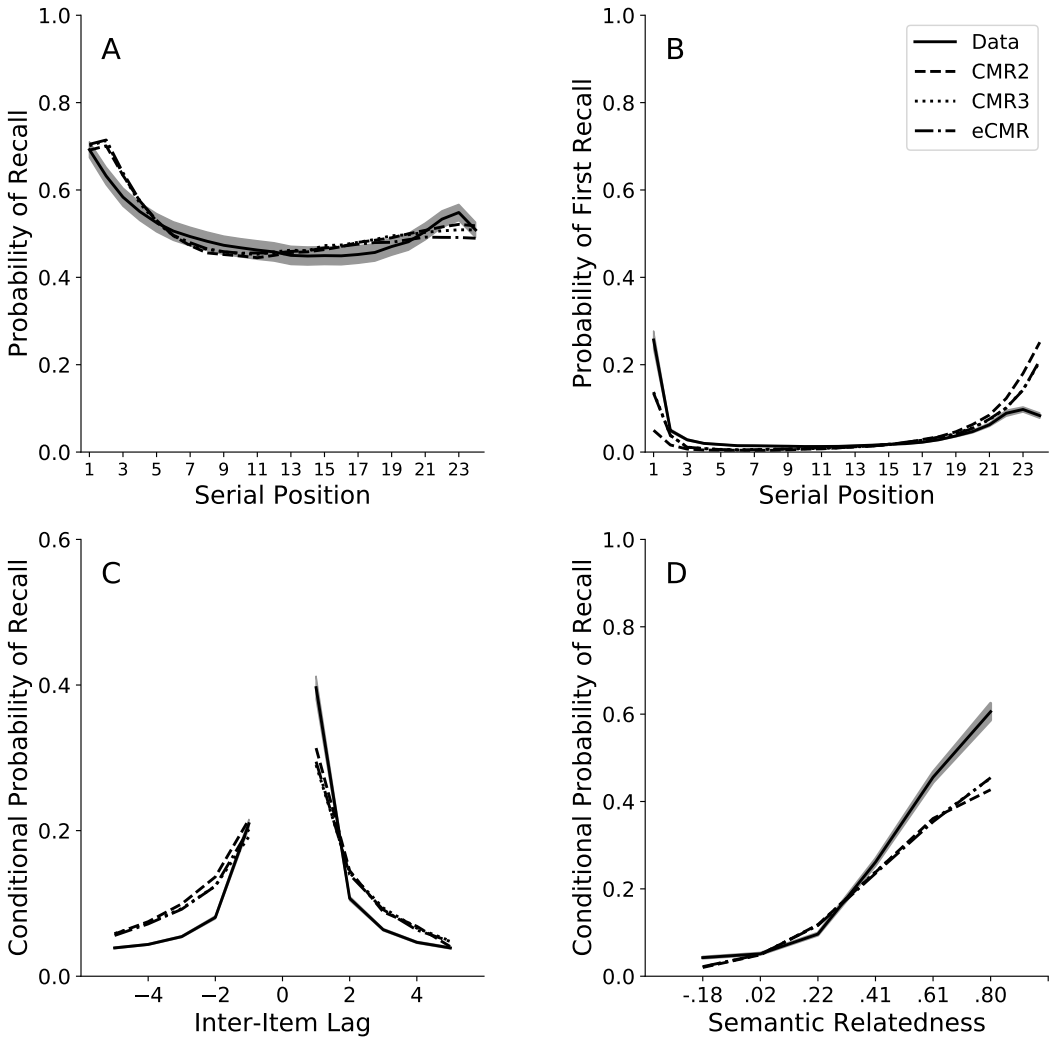


*Figure 2. Modeling the organization of memories during free recall.* We fit CMR2, eCMR, and CMR3 to serial position effects and to temporal, semantic, and emotional clustering in a delayed free recall task. Error regions represent +/- standard error of the mean, calculated across subjects (data). Error regions across parameter sets' simulated dataset values are omitted for visibility, but see Table 3 for RMSE fit indices. **A.** Recall probability as a function of items' presentation order during encoding. **B.** The probability that an item will be recalled first, as a function of an items' presentation order during encoding. **C.** The probability that items will cluster together 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 probability that items will 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 for six semantic word2Vec bins (-.18 to .02, .02 to .22, .22 to .41, .41 to .61, .61 to .80, .80 to 1.0).

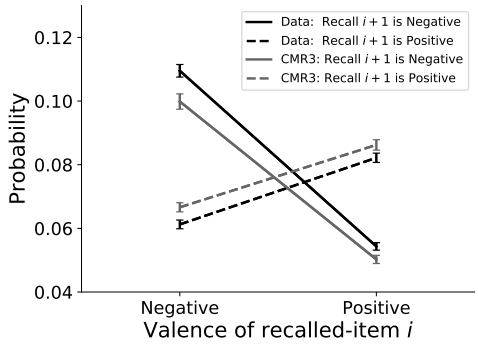


*Figure 3. The emotional clustering effect.* In a free recall task, subjects tend to successively recall items that share contextual features, including their emotional valence. Thus, after recalling a negative (positive) item they tend to recall other negative (positive) items. CMR3 provides a good fit to this effect, as seen in the conditional probability of transitions among items as a function of their valence. Solid lines represent the tendency to transition to a negative item, and dashed lines represent the tendency to transition to a positive item.

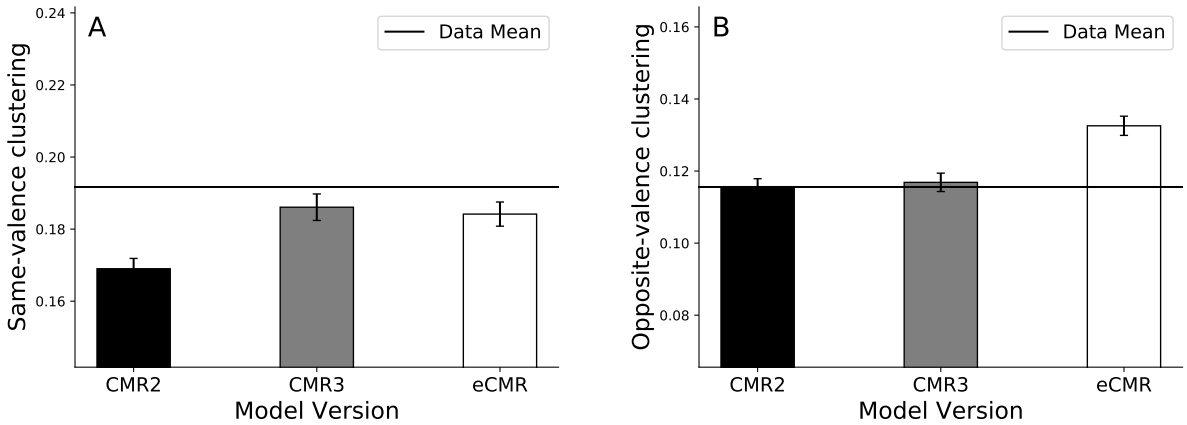


*Figure 4. Modeling the organization of memories during free recall, continued.* The emotional clustering effect represents the tendency of items to cluster during recall according to their emotional valence. This figure shows the model-predicted degree of clustering among **A.** same-valent and **B.** oppositely-valent items for CMR2 (black bars), CMR3 (gray bars), and eCMR (white bars). Error bars represent the standard error of the mean.

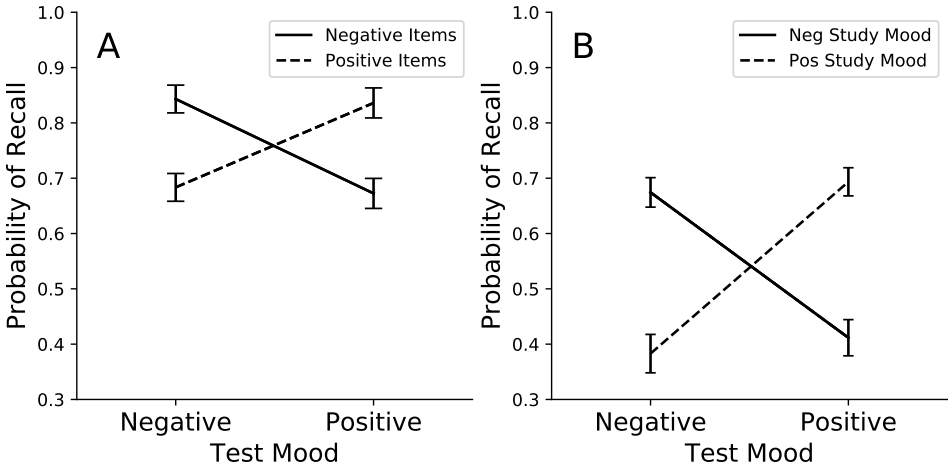


*Figure 5. Simulating Mood-congruent and Emotion-state Dependent Recall.* Using parameters obtained from the behavioral fits in Simulation 1, we used CMR3 to predict recall outputs based on two factors: (1) the match between an item's valence and the emotional state during retrieval (mood-congruent recall), 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.

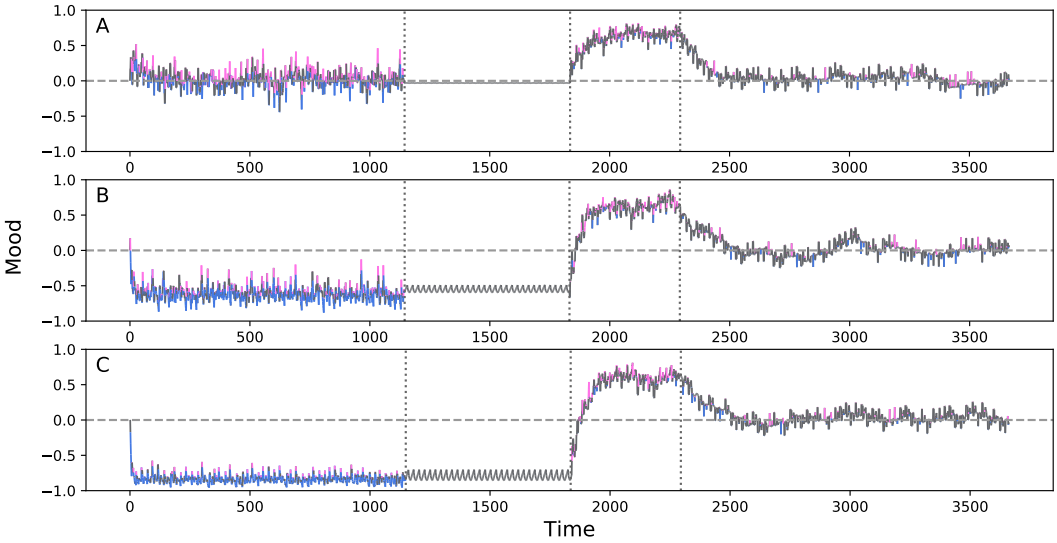


*Figure 6. Simulating negative-event frequency effects.* Each panel illustrates the evolution of model-simulated mood across four periods, marked by vertical lines: (1) A developmental period, in which a virtual subject experienced a mixture of both emotional and neutral events, (2) A neutral period, in which a virtual subject experienced solely unique, neutral events, (3) A simulated course of positive event scheduling, a component 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).

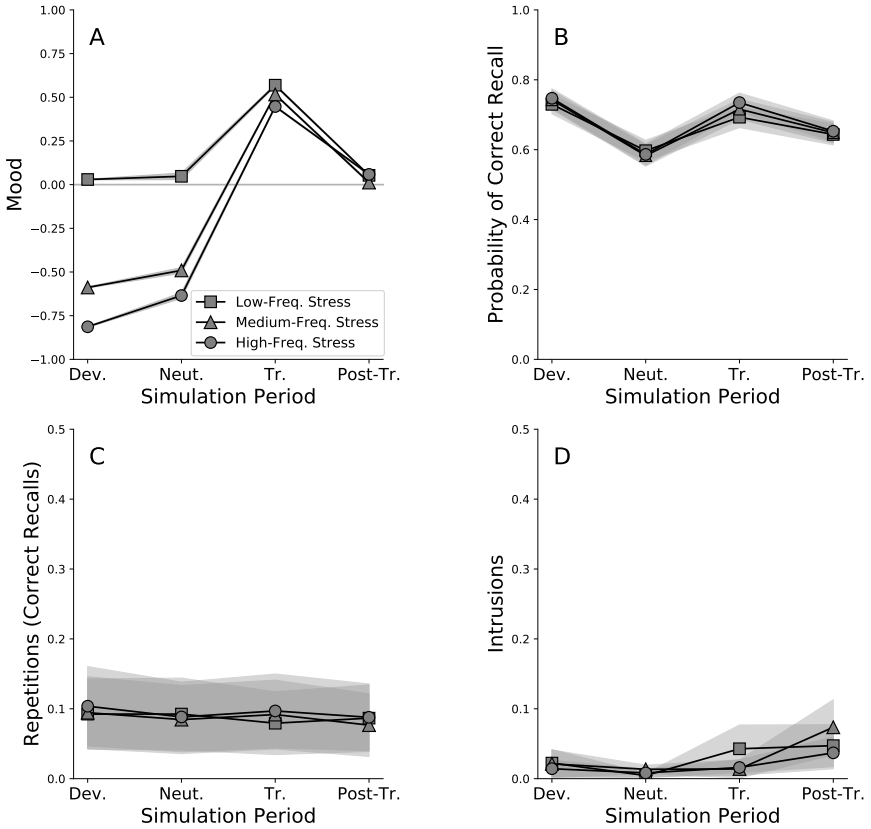


*Figure 7. Aggregated results for negative-event frequency effects.* This figure presents the model-predicted patterns of mood and recall for Simulation 4, aggregated across the results for all virtual subjects (i.e., parameter sets). Results are averaged within each simulation stage: Developmental (Dev), Neutral (Neut), Treatment (Tr), and Post-Treatment (Post-Tr). Subjects experienced one of three conditions during the developmental period: Low-Frequency negative events (square markers), 30% negative, 30% positive events; Moderate-Frequency negative events (triangle markers), 60% negative, 20% positive events; and High-frequency negative events (circle markers), 70% negative, 10% positive events. After the developmental period, all virtual subjects experienced a neutral period (100% neutral events), followed by a “treatment period" simulating a positive event scheduling intervention, with a 3:1 ratio of positive to negative events, 30% positive and 10% negative. Finally, we evaluated a post-treatment period of 80% neutral, 10% negative, and 10% positive events. **A.** Mean levels of virtual subjects’ mood. **B.** Probability of correct recall. **C.** Mean number of memory-repetitions per list. **D.** Mean number of memory-intrusions per list. Error regions represent +/-the standard error of the mean, calculated across all parameter sets’ predictions.

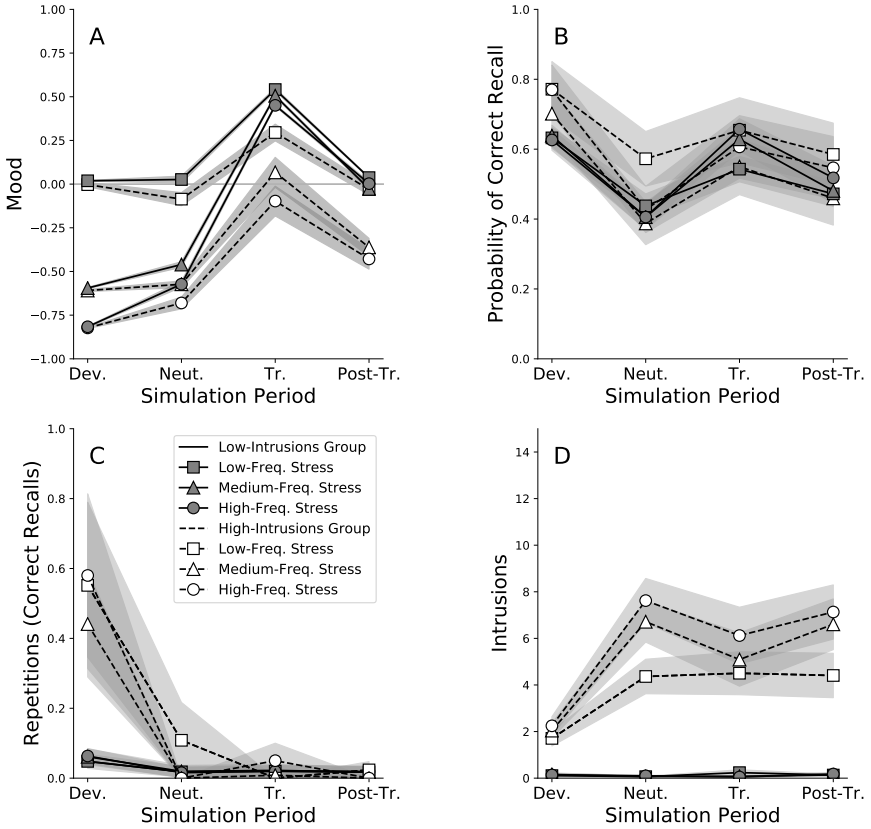


*Figure 8. Aggregated results for negative-event repetition effects.* This figure presents the model-predicted patterns of mood and recall for Simulation 5, aggregated across the results for all virtual subjects. Results are averaged within each simulation stage: Developmental (Dev), Neutral (Neut), Treatment (Tr), and Post-Treatment (Post-Tr). Subjects experienced one of three conditions during the developmental period: Low-Frequency negative events (square markers), 30% negative, 30% positive events; Moderate-Frequency negative events (triangle markers), 60% negative, 20% positive events; and High-frequency negative events (circle markers), 70% negative, 10% positive events. One negative event repeated every 10 lists. After the developmental period, all virtual subjects experienced a neutral period (100% neutral events), followed by a "treatment period" simulating a positive event scheduling intervention, with a 3:1 ratio of positive to negative events, 30% positive and 10% negative. Finally, we evaluated a post-treatment period of 80% neutral, 10% negative, and 10% positive events. **A.** Mean levels of virtual subjects' mood. **B.** Probability of correct recall. **C.** Mean number of memory-repetitions per list. **D.** Mean number of memory-intrusions per list. Solid lines show virtual subjects with low memory-intrusion counts (Low-Intrusions Group). Dashed lines show virtual subjects with high memory-intrusion counts (High-Intrusions Group). Error regions represent +/- the standard error of the mean, calculated across all parameter sets' predictions.

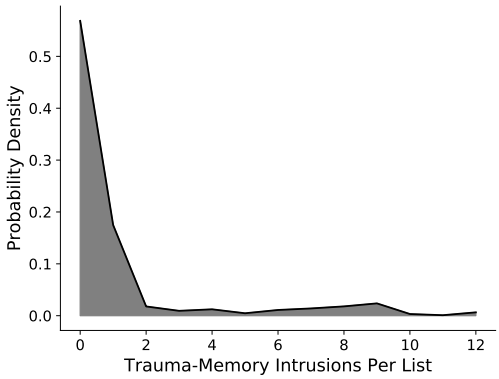


*Figure 9. Distribution of trauma-memory intrusions per list for virtual subjects.* Whereas 85 of 97 virtual subjects exhibited a mean of fewer than one trauma-memory intrusion per list during the period of neutral events following their trauma, 12 virtual subjects experienced trauma-memory intrusions at higher frequencies.

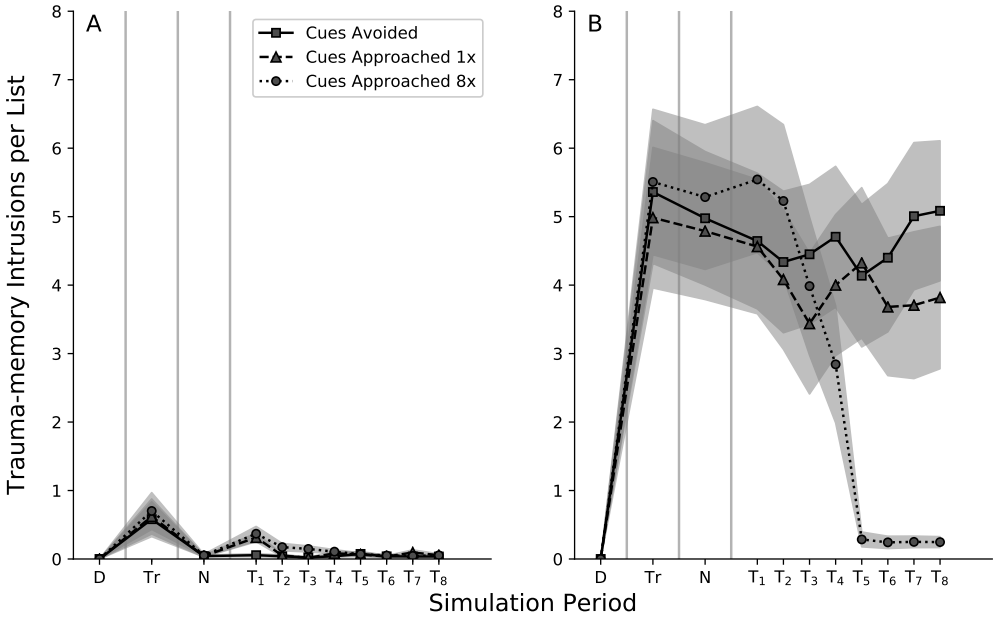


*Figure 10. Effect of temporal context on trauma-memory intrusions.* This figure presents the model-predicted patterns of trauma-memory intrusions resulting from encoding a traumatic event, aggregated across all virtual subjects. Panel A shows results for virtual subjects with minimal or zero trauma-memory intrusions (Low-Intrusions Group). Panel B shows results for virtual subjects with high frequencies of trauma-memory intrusions (High-Intrusions Group). 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. $T_{1-8}$ = each of the eight time-points at which a virtual subject has an opportunity to either approach or avoid trauma cues.

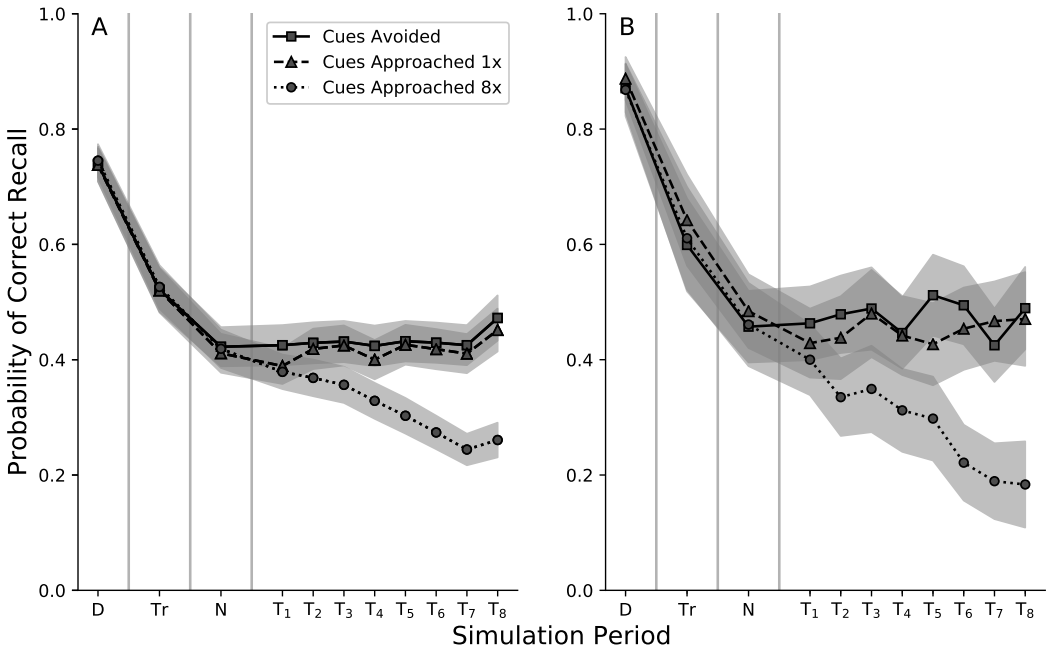


*Figure 11. Effect of temporal context on correct recall.* This figure presents the model-predicted patterns of correct recall of recent events after encoding a traumatic event, aggregated across the results for all virtual subjects (i.e., parameter sets). Panel A shows results for the "Low-Intrusions 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. $T_{1-8}$ = each of the eight time-points at which a virtual subject has an opportunity to either approach or avoid trauma cues.

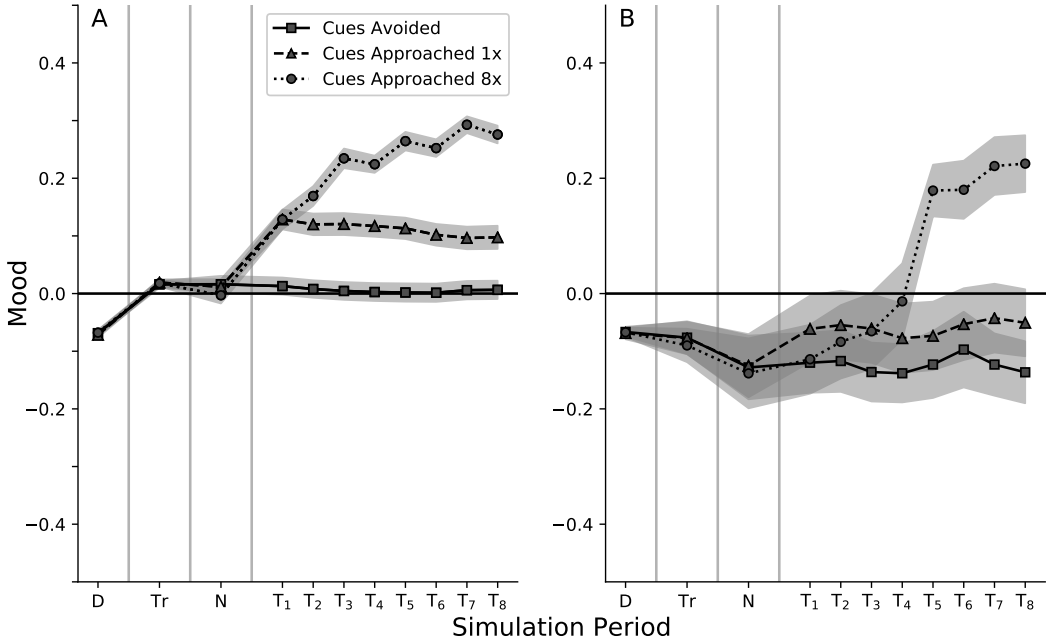


*Figure 12. Effects of temporal context on mood.* Model-predicted mood following a traumatic event, aggregated across virtual subjects. **A.** Low-Intrusions Group, i.e., virtual subjects with low rates of trauma-memory intrusions after their traumatic event. **B.** High-Intrusions Group, i.e., virtual subjects with high rates of trauma-memory intrusions after their traumatic event. Square markers indicate a baseline condition, in which the neutral period consists solely of unique, neutral events. Triangle markers indicate a single exposure to a set of three trauma cues, or the nonnegative stimuli that preceded the traumatic event during encoding the neutral period. Circle markers indicate 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. $T_{1-8}$ = each of the eight time-points at which a virtual subject has an opportunity to either approach or avoid trauma cues.

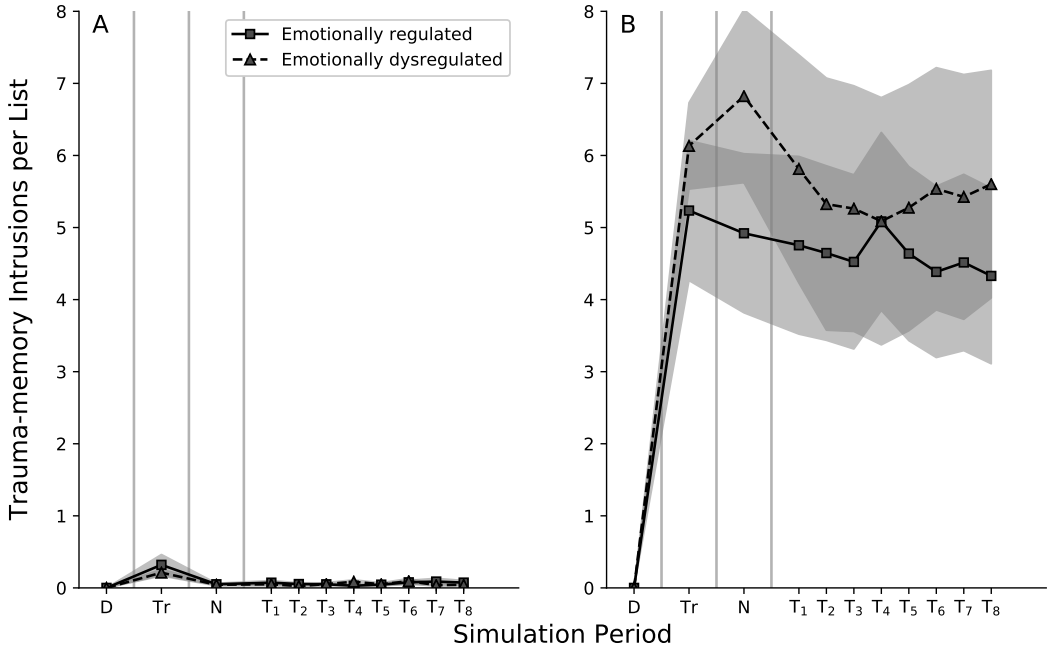


*Figure 13*. Effects of emotion dysregulation on trauma-memory intrusions. Panel A: Low-Intrusions Group. Panel B: High-intrusions group. Without intervention (as in this simulation), virtual subjects with high emotion dysregulation ($\beta_{emot} = 0.8$, triangle markers) experience heightened rates of trauma-memory intrusions relative to emotionally regulated virtual subjects ($\beta_{emot} = 0.3$, square markers). D = Developmental period preceding the traumatic event. Tr = Developmental period including and immediately following the traumatic event. N = Neutral events-only period, which continues throughout timepoints $T_1 - T_8$. Error regions represent +/- standard error of the mean across virtual subjects.

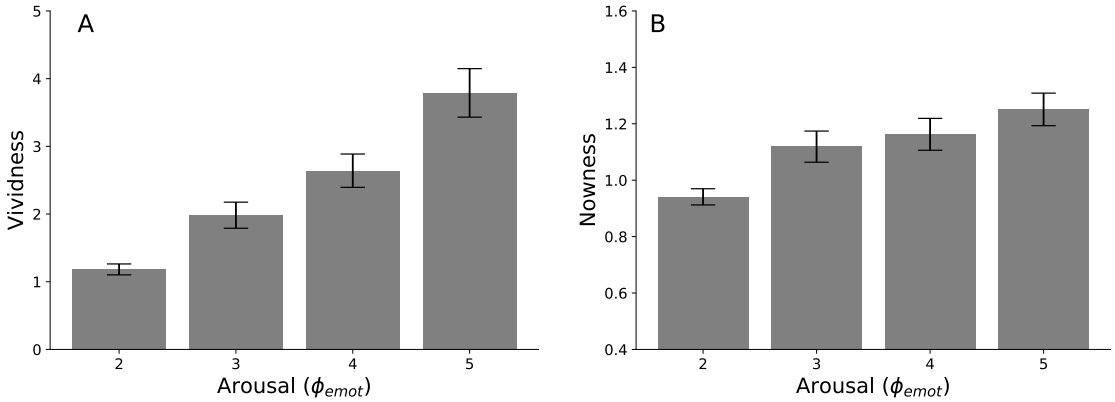


*Figure 14. Effects of negative arousal on memory vividness and "nowness."* To simulate arousal we varied the learning rate for the associations between context features and item features during encoding ($\phi_{emot}$). We then define vividness as the activation of a retrieved memory's features, and we define nowness as the similarity between current context and retrieved context. **A.** Vividness increases with the simulated arousal during the initial encounter with a subsequently remembered experience. **B.** Nowness also increases with the simulated arousal of an experience. Error regions represent +/- 1 *S.E.M.* across the values for each of 97 virtual subjects.

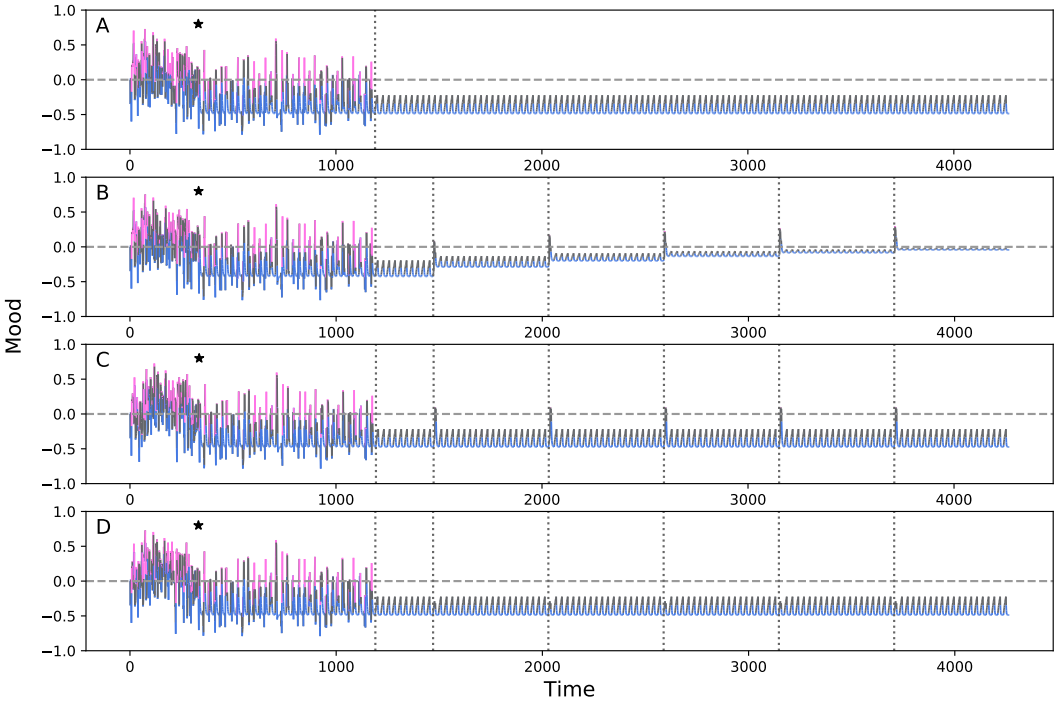


*Figure 15*. Simulating the effects of trauma and imaginal exposure therapy on intrusive memories. Each panel illustrates the evolution of model-simulated mood across two periods: (1) A developmental period, in which a virtual subject experienced an even mixture of negative and positive life events (30% 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).

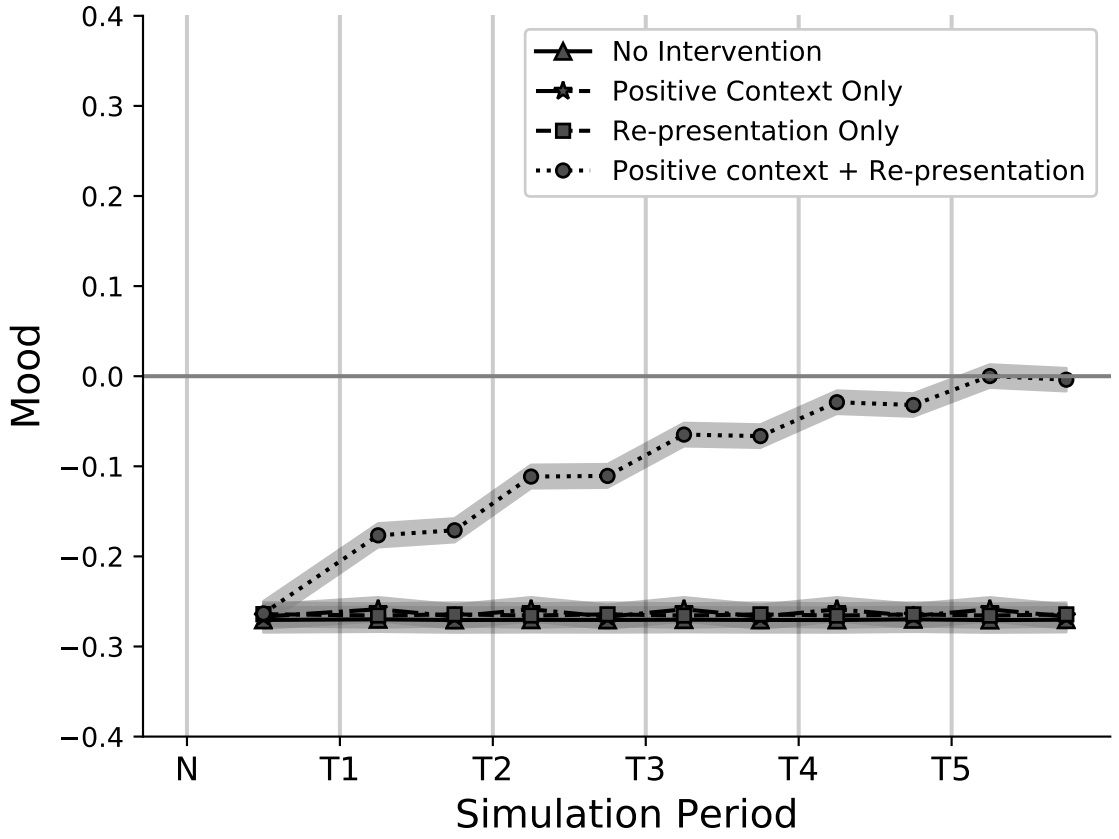


*Figure 16*. Effects of trauma exposure and imaginal exposure therapy on mood. This figure presents the model-predicted patterns of mood for each of four conditions, aggregated across the results for all virtual subjects (i.e., parameter sets). Each virtual subject experienced a developmental period, 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.

# Appendix

## Parameter Estimation and Model Comparison Methods

We used a particle swarm optimization method to fit each model (CMR2, eCMR and CMR3) to each individuals' behavioral data. Like a genetic algorithm or downhill simplex technique, a particle swarm searches a high-dimensional parameter space to find the optimal value of some objective function – in our case, the objective function is minimizing the error between predicted and observed data. Each particle in a swarm refers to a set of parameters (we used 200 particles) for a given model. We evolved the particles over 30 iterations, searching for the parameter set that provided the best fit to the behavioral data, separately for each model (CMR2, eCMR, and CMR3) and for each subject. In a given "run" of the model with one "particle," we simulate an entire 24-session experiment, using the actual list items that a given subject experienced to determine the sequence of items presented to the model. The model then generates a set of simulated recall sequences, which we analyze to estimate the same behavioral patterns shown in Figure 2.

Having generated simulated behavioral data obtained for a given particle for one of our three models, we can now calculate a goodness-of-fit statistic quantifying the deviation between model and data. For our fit index, we used the $\chi^2$ 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 $\chi^2$ error for each particle is determined, the optimal parameter set is identified as the one having the lowest $\chi^2$ error in this iteration. The particle-swarm algoirthm then adjusts the other particles to produce results similar to the current optimal value (hence the term "swarm"). Over the course of iterations, the particles converge on an optimal set of parameters, thereby minimizing the $\chi^2$ error.

The 56 points contributing to the $\chi^2$ error for each model fit comprised the first 15 and last three positions of the serial position curve (Fig 2a); the first 15 positions of the probability of first recall curve (Fig 2b); the lag-CRP curve at lags of -3 through +3; the

nine conditional probabilities of transitioning among negative, positive, and neutral item types (negative-to-negative, negative-to-positive, negative-to-neutral, and so forth); all six points on the semantic CRP curve; the frequency of PLI's per list; and the frequency of ELI's per list.

Particle swarm optimization yielded best-fitting parameters for each of the 97 subjects in Experiment 1. At the end of the particle swarm process, the mean $\chi^2$ 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 $\chi^2$ error values indicate better model fits. Degrees of freedom for the $\chi^2$ 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. Table A1 presents the means and standard errors of the sets of parameters obtained for CMR2, eCMR, and CMR3, taken across individual-subject fits.

After obtaining the best-fitting parameters for each model (Table A1), we used these parameters to simulate a full dataset of 24 sessions of delayed free recall for each model. To assess each model's ability to fit the aggregate data, we calculated three measures of fit: (1) the same $\chi^2$ 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 et al., 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 $\chi^2$ 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 $\chi^2$ 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 et al., 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:

$$BIC = kln(n) + nln(\frac{RSS}{n}) \tag{12}$$

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 $\chi^2$ 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 2). 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.

Table A1

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

| | CMR2 | | eCMR | | CMR3 | |
|---|---|---|---|---|---|---|
| | Mean | SEM | Mean | SEM | Mean | SEM |
| $\beta_{enc}$ | .298 | .008 | .295 | .008 | .286 | .007 |
| $\beta_{rec}$ | .713 | .013 | .717 | .013 | .701 | .012 |
| $\beta_{distract}$ | .458 | .018 | .481 | .020 | .453 | .020 |
| $\beta_{post}$ | .672 | .021 | .675 | .025 | .733 | .020 |
| $\beta_{emot}$ | – | – | .369 | .018 | .322 | .016 |
| $\gamma_{FC}$ | .799 | .010 | .800 | .010 | .813 | .009 |
| $\gamma_{CF}$ | .872 | .009 | .845 | .010 | .869 | .009 |
| $\gamma_{emot}$ | – | – | .394 | .025 | .417 | .026 |
| $\phi_S$ | 1.413 | .074 | 1.009 | .064 | .957 | .072 |
| $\phi_D$ | .611 | .042 | .747 | .037 | .728 | .037 |
| $\kappa$ | .257 | .012 | .279 | .013 | .293 | .012 |
| $\eta$ | .184 | .010 | .181 | .010 | .189 | .010 |
| $\lambda$ | .118 | .010 | .115 | .011 | .125 | .010 |
| $s$ | 1.044 | .041 | 1.049 | .039 | 1.117 | .049 |
| $\omega$ | 22.821 | .383 | 22.702 | .406 | 22.605 | .382 |
| $\alpha$ | .934 | .006 | .958 | .0006 | .954 | .006 |
| $c_{thresh}$ | .339 | .018 | .991 | .038 | .992 | .038 |
| $\phi_{emot}$ | 1.0 | | 1.0 | | 1.0 | |

Note. $\beta_{enc}$ = drift rate for temporal context during encoding. $\beta_{rec}$ = drift rate for temporal context during recall. $\beta_{distract}$ = drift rate for temporal context during a distractor item or shift between mental tasks. $\beta_{post}$ = drift rate for temporal context at the end of a recall session. $\beta_{emot}$ = drift rate for emotional context. $\gamma_{FC}$ = learning rate for item-to-temporal context associations. $\gamma_{CF}$ = learning rate for temporal context-to-item associations. $\gamma_{emot}$ = learning rate for emotional context-to-item associations. $\phi_S$ scales the primacy effect and $\phi_D$ determines its rate of decay over the presentation of new items. $\kappa$ = rate of evidence decay in the leaky accumulator due to the passage of time. $\eta$ = width of the normal distribution of noise values for each item's recall evidence in the leaky accumulator. $\lambda$ = degree of lateral inhibition between items in the leaky accumulator. $s$ = scales the strength of inter-item semantic associations. $\omega$ scales the raised threshold for item repetition, and $\alpha$ 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. $\phi_{emot}$ = a scalar that magnifies the strength of context-to-item associations during the encoding of a high-arousal negative item.