**Emotional learning retroactively enhances item memory at the cost of source memory misattribution**

**Introduction**

Emotional experiences gain privileged access to the neurobehavioral mechanisms of long-term memory (McGaugh 2015; LaBar & Cabeza 2006; Yonelinas & Ritchey 2015; Mather, Clewett et al. Brain Something; A Kensinger review--these are people who might review this). Interestingly, emotional enhancements of memory are not limited to temporal window of the emotional event. Memory for seemingly mundane details encoded close in time to an emotional experience can also be enhanced. On one hand, memory enhancement via temporal proximity can be adaptive, allowing us to use information in our environment to avoid aversive situations in the future. Alternatively, this same mechanism can result in previously neutral cues becoming unwanted reminders for traumatic emotional memories. But what aspects of memory are modulated via temporal proximity to an emotional event? Episodic memories, for instance, are composed of stimulus information (e.g., item memory) embedded with contextual details (e.g., source context memory) (Johnson, Hashtroudi, Lindsay 1993; Other good episodic memory REFS). While emotion enhances item memory (Sharot & Phelps 2004; LaBar & Cabeza 2006; Mather 2007), the effect of emotion on contextual details is less consistent (Yonelinas & Ritchey 2015). Here, we investigated how temporal proximity to a strong emotional event (fear conditioning) influences both item and contextual memory for related information. Our goal was to advance our understanding of how strong learning events transform weakly learned information into durable long-term episodic memories.

Enhancement in memory via a temporal association between mundane and salient events is consistent with neurobiological models of long-term memory. For instance, the behavioral tagging hypothesis (derived from the synaptic tagging hypothesis; Frey & Morris 1997) proposes that weakly learned information is strengthened in memory if it is encoded within a critical time window of a more salient event, and if the two events share overlapping neural ensembles (Wang, Redondo, Morris 2010; Moncada & Viola 2007; Ballarini et al. 2009; Takeuch, Duszkiewicz,i et al., 2016). As emotion is a powerful learning event, neutral information encoded within temporal proximity might be strengthened in memory via a mechanism of behavioral tagging. But which features of rich memory episodes are influenced by this putative behavioral tagging mechanism?

There is prior evidence that novel or emotional events retroactively enhance item memory in humans. For example, Pavlovian conditioning using exemplars of an object category as conditioned stimuli (CS) has been shown to produce selective increases in long-term (24-hour) recognition memory for conceptually related stimuli encoded before, during, and after fear conditioning (Dunsmoor et al., 2015). A similar result was obtained using an appetitive outcome (Patil, Murty, et al. 2016). This finding accords with a behavioral tagging hypothesis in a number of ways, including the temporal symmetry of enhancement for neutral information weakly encoded before or after the stronger event, and necessity of a period of consolidation to observe retroactive memory enhancements (Ballarini et al. 2009).

However, it is unclear whether and how this mechanism may affect memory for the contextual details associated with the weak event. Emotion effects memory accuracy for information directly associated with an emotional stimulus, such as the spatial and temporal context in which the emotional stimulus was encoded (Talmi et al., Psych Review 2019; Rimmele et al., 2012 Emotion; Schmidt, Patnaik, Kensinger Cogn Emot 2012; Kensinger & Schacter 2006 Memory & Cognition). Therefore, one possibility is that emotional learning improves both item and source memory for information in temporal proximity to an emotional event. In the case of a Pavlovian fear conditioning design using category exemplars as CSs, we might expect selectively enhanced recognition for threat relevant stimuli to be accompanied by selective accuracy in sourcing those items to their correct temporal context. Alternatively, emotional learning might have selective effects on item memory, but have no effect (Wang and Fu 2010, Neurobio of Learning & Memory; Sharot & Yonelinas 2008) or even impair memory for contextual information. Indeed, it is plausible that linking weak and strong events by temporal proximity might improve item memory at the cost of accurate source context memory. Memory retrieval for weakly encoded items may in fact depend on a source context misattribution to the emotional context. In other words, it may simply be easier to remember an item if it is associated with an emotional context, regardless of when it was actually encoded (Takashima, van der Ven, et al., Neuroimage 2016). In this way, we may expect an association between selective retroactive and proactive item memory enhancements and a misallocation of source memory to the more salient context.

In addition to our exploration of how emotional learning impacts memory for episodic events, we also tested a parallel hypothesis that emotional learning alters abstract properties of stimuli. Previous work demonstrates that emotional learning is sensitive to abstract stimulus properties such as how well an item represents its broader category, i.e., typicality (Dunsmoor, Kragel, et al., Cerebral Cortex; Dunsmoor & Murphy 2014; Dunsmoor & Murphy 2015 TICS; Struyf, Hermans, Vervliet, 2018, BRAT; Lei, Wang, Dou, Qiu, Li, 2019, Int J Psychophys). Typicality itself may be an important feature in predicting memorability of items encoded in the penumbra of an emotional event (Rust & Mehrpour TICS 2020; Lissman, Ducel, Grace 2010). Whether emotional learning has the power to transform an abstract stimulus property such as subjective typicality, however, is unknown.

Specifically, we hypothesized that Pavlovian conditioning would enhance the subjective typicality of basic level category exemplars, as a function of what emotional association is learned about the category during fear conditioning. This hypothesis was derived in part from past behavioral research showing that typical stimuli serve as stronger sources of fear generalization than less typical stimuli (Dunsmoor & Murphy Psych Science 2014; Strufy, Hermans, Vervliet, 2018). Moreover, typical items preferentially engage the hippocampus as well as hippocampal-amygdala functional connectivity during Pavlovian fear conditioning (Dunsmoor et al., Cerebral Cortex, 2014), which is a putative substrate for retroactive and proactive enhancement in memory via a behavioral tagging mechanism.

To test these hypotheses, subjects underwent a two-day Pavlovian fear conditioning task that included trial-unique (i.e., non-repeating) pictures of animals and tools as CSs, based on the protocol from Dunsmoor et al. (2015). Items were encoded before, during, and after Pavlovian fear conditioning. Relevant to our hypotheses, we collected measures of both item recognition memory and temporal context memory. We predicted that emotional learning would have divergent effects on retroactive and proactive memory accuracy for items and contextual details. Specifically, we predicted that items encoded before and after fear conditioning would be selectively remembered, at the cost of incorrectly sourcing the encoding of those items to the temporal context of fear conditioning. We also collected subjective typicality judgements for each item, and predicted that emotional learning would alter subjective typicality such that exemplars conceptually related to threat would be judged as more representative than control exemplars. Such findings might indicate that encoding in the wake of an emotional experience binds the memory of a weakly encoded item to the more salient temporal context, and additionally has the power to transform abstract stimulus properties such as how well an item represents its category.

**Discussion**

While emotional events are often prioritized in memory, it is far less clear how emotional events affect memory for other information encoded close in time. Here, we provide support for a behavioral tagging mechanism in human episodic memory, whereby emotional learning selectively and retroactively enhanced item memory conceptually related to the emotional event, replicating prior findings (Dunsmoor; Patil). Building on prior findings, we found that related information encoded before or after fear conditioning is consistently misattributed to the more salient context. Put in terms of a tag-and-capture mechanism: source memory errors were biased to the temporal context of the strong event that provided memory stabilization for the weak event. We also found new evidence that emotional learning selectively enhances subjective typicality for related exemplars, and that typical items were better remembered overall. Collectively, these findings advance knowledge on how emotional learning affects mnemonic and abstract stimulus representations for information encoded before and after an emotional experience.

While effects of emotional arousal on item memory are fairly consistent (LaBAR & CABEZA 2006; McGUAGH 2015; MURTY, RITCHEY, ADCOCK, LABAR), the effect of emotion on memory for contextual details is less straightforward. Contextual details sometimes receive a boost in memory (REF), but other studies find no effect or worse memory for contextual details associated with an emotional item (SHAROT & PHELPS, OTHER REFS FROM ABOVE AGAIN). The divergent effects of emotion on item and source memory may be related to different temporal dynamics in forgetting between item-emotion binding, supported by the amygdala, and item-context binding, supported by the hippocampus (YONELINAS & RITCHEY). Specifically, amygdala-dependent emotional item memory is resistant to forgetting, whereas hippocampus-dependent item-context memory decays at a faster rate. Extending this model to memory for related items encoded *around the time* of emotional learning has interesting implications for interpreting episodic memory through the lens of a putative behavioral tagging mechanism. Specifically, it suggests that emotional events act predominately on the amygdala to support retroactive and proactive item memory enhancements, which would accord with the use of the Pavlovian conditioning design employed here. In this way, amygdala-dependent fear conditioning may upregulate hippocampal processing to support selective recognition of weakly encoded items. However, this memory benefit would not necessarily be accompanied by strengthening hippocampus-dependent item-context memory for those same weakly encoded items.

One important question regarding source memory results is the relative contribution of consolidation versus retrieval processes. One possibility is that emotional learning boosts memory for weakly encoded items in part by linking these memories to the temporal context that provides memory stabilization (i.e., fear conditioning). This explanation is in keeping with a tag-and-capture model, which is primarily a mechanistic account for the consolidation of newly formed memories (Redondo & Morris NRN). This account may afford some explanatory power for the strong association between selective item memory enhancements (CS+ > CS-), which our prior work shows is consolidation dependent (Dunsmoor et al., 2015; Dunsmoor et al., 2018 Nature Human Behaviour; Patil et al.), and the bias to misattribute CS+ items to the fear conditioning context. Alternatively, source memory errors may be a factor of retrieval processes, per se. In this case, CS+ item memory may be accompanied by retrieval of the more salient temporal context, which is the moment when some CS+’s were paired with shocks. Thus participants may attribute more CS+ items to the emotional context because those items are accompanied by a feeling of remembering (SHarot, Delgado, Phelps), regardless of when the item was actually encoded. This explanation is consistent with the general properties of source monitoring that involve retrospectively attributing the origin of a memory to a particular source (JOHNSON et al. 1993), which could change given the context of retrieval (Tulving Thomson 1973; Smith & Vela 2001; Howard & Kahana 2002). It is also consistent with a retrieval focused account of emotional memory enhancements (TALMI), wherein emotional item memory is bound to changing representations of mental contexts. In terms of context retrieval, remembering a CS+ item may reactivate the mental context associated with the fear conditioning phase, thereby promoting retrieval for related items encoded in temporal proximity. Interestingly, there is evidence that distinct phases of a fear conditioning experiment can be understood as separable mental contexts that uniquely organize memory (Hennings; Dunsmoor et al. nature human behaviour) in a manner consistent with how shifts in mental context segment experience for distinct episodes (Dubrow; Clewett; etc.).

We also found that items from the CS+ category were rated as more typical than CS- items, regardless of when the item was encoded. Importantly, there was no inherent difference in typicality as a function of superordinate category (animal, tool), suggesting that fear conditioning itself modulated typicality judgements. These results are new evidence that conditioning alters an abstract stimulus property like how well an item represents its category. This enhancement in representativeness may enable CS+ items to be better remembered over time, as we also found that typical items were more memorable overall. There is surprisingly little research on the link between stimulus typicality and memorability, but there is evidence that semantic similarity promotes free recall (Howard & Kahana 2002b, journal of memory and language—be sure to get the right citation, as there are 2 howard & kahana papers from 2002). Thus, one mechanistic possibility is that fear conditioning augmented semantic organization by raising the overall typicality of distinct category members from the CS+ category, which in turn facilitated memory for related items encoded close in time. This would be consistent with prior neuroimaging evidence that fear conditioning enhances representational similarity of semantic categories in object-selective occipitotemporal cortex and the amygdala, and that more typical items engage the hippocampus and hippocampal-amygdala connectivity (Dunsmoor et al. Cerebral cortex). More research on the relationship between typicality and memorability is warranted to help bridge the current findings to the rest of the episodic memory literature.

Recent work has made progress on translating neurobiological models of long-term memory to explain memory strengthening for weak memories encoded in temporal proximity to a more salient event. According to the synaptic tagging hypothesis (Frey & Morris 1997; Redondo & Morris NRN, 2011), and its behavioral counterpart (Moncada & Viola 2007, J Neuro), weak activation that is only sufficient to produce a transient short-term memory can be transformed into a durable long-term memory if accompanied by a more salient event that relies on the same neural system within a critical time window. The weak event sets a putative learning tag that is stabilized by release of plasticity related proteins induced by the strong event, broadening the window of time by which salient events can modulate long-term memory. But memories are multifaceted, and how a putative behavioral tagging mechanism impact separate mnemonic aspects of the same experience (item and source), or other abstract stimulus properties (typicality), has been understudied. Matching the task demands of episodic memory and Pavlovian conditioning allowed us here to show that an emotional event has divergent effects on item and source memory and enhances subjective typicality of related events encoded close in time. Mechanisms that link seemingly inconsequential information to a future or past emotional event can serve an adaptive function. For instance, we do not always know the significance of numerous details we encounter throughout the day. By integrating memories encoded close in time to a meaningful event, we can remember information that may be relevant to seeking out or avoiding similar outcomes in the future. Enhancing the representativenessof the experience may serve a similar mnemonic function. But linking memories by temporal proximity may come at the cost of misattributing neutral memories to an emotional context. Such an organization proves maladaptive when innocuous cues experienced around the time of highly negative events trigger retrieval of unwanted emotional memories. Thus, as a consequence of organizing memories based on temporal proximity to a salient event, a host of potentially irrelevant neutral information acquires the capacity to reactivate emotional memories in the future.