Research

Surprise and destabilize: prediction error influences episodic memory reconsolidation

Alyssa H. Sinclair and Morgan D. Barense

Department of Psychology, University of Toronto, Ontario M5S 3G3, Canada

Through the process of "reconsolidation," reminders can temporarily destabilize memories and render them vulnerable to change. Recent rodent research has proposed that prediction error, or the element of surprise, is a key component of this process; yet, this hypothesis has never before been extended to complex episodic memories in humans. In our novel paradigm, we used naturalistic stimuli to demonstrate that prediction error enables adaptive updating of episodic memories. In Study 1, participants ($N = 48$) viewed 18 videos, each depicting an action–outcome event. The next day, we reactivated these memories by presenting the videos again. We found that incomplete reminders, which interrupted videos before the outcome, made memories vulnerable to subsequent interference from a new set of videos, producing false memories. In Study 2 ($N = 408$), an independent sample rated qualities of the stimuli. We found that videos that were more surprising when interrupted produced more false memories. Last, in Study 3 ($N = 24$), we tested competing predictions of reconsolidation theory and the Temporal Context Model, an alternative account of source confusion. Consistent with the mechanistic time-course of reconsolidation, our effects were crucially time-dependent. Overall, we synthesize prior animal and human research to present compelling evidence that prediction error destabilizes episodic memories and drives dynamic updating in the face of new information.

[Supplemental material is available for this article.]

Our memories are malleable. Through the process of "reconsolidation," memories can be destabilized by a reminder, modified, and then stabilized again (Nader et al. 2000). However, recent rodent research has revealed that in some cases, destabilization does not occur if the memory is reactivated in a manner identical to the original training (Díaz-Mataix et al. 2013; Exton-McGuinness et al. 2014). Reminders that present a learned conditioned stimulus (CS), but omit the anticipated unconditioned stimulus (US), most effectively initiate the reconsolidation process. One explanation is that these incomplete reminders elicit a form of "prediction error," which operates as a critical trigger for memory destabilization (Exton-McGuinness et al. 2015). Intuitively, a prediction error mechanism seems adaptive; memories would be modified only when new information necessitated updating. However, this prior animal research on prediction error relied on simple memories for stimulus-response contingencies. It is thus an open question whether prediction error similarly influences complex episodic memories in humans. The present program of research synthesizes animal and human reconsolidation research, capitalizing upon prediction error to destabilize and manipulate detailed naturalistic memories.

Although prior studies have not explicitly applied prediction error theory to human reconsolidation research, converging neural evidence supports the idea that prediction error influences human memory. The hippocampus, critical for memory reactivation and reconsolidation, is sensitive to mismatches between expectation and reality (Kumaran and Maguire 2007; Duncan et al. 2009; Chen et al. 2011). Hippocampal responses during memory reactivation are strongest when there is a similar-but-different stimulus paired with a learned cue, predicting memory updating (Long et al. 2016). Similarly, incomplete reminders of learned paired-associates make memories susceptible to subsequent interference (Forcato et al. 2009, 2010, 2016). Moreover, subtle contextual cues, rather than comprehensive reminders, can destabilize memory for a list of objects (Hupbach et al. 2007, 2009, 2013). Most recently, surprising misinformation has been found to elicit semantic prediction errors that parallel the canonical dopaminergic coding of valence and magnitude of expectancy violations (Pine et al. 2018).

Despite this accumulating evidence that reactivating a memory while violating expectations can destabilize it, to our knowledge, no existing human studies have directly bridged prediction error to the reconsolidation of richly detailed episodic memories. The bulk of prior studies that contrasted complete and incomplete reminders relied on associative learning or list memorization (Hupbach et al. 2007; Forcato et al. 2009, 2016; Dębiec et al. 2011; Sevenster et al. 2014). Faced with the challenge of generating prediction errors with naturalistic stimuli, we had the insight to use videos that violated expected action–outcome contingencies. We used multimodal, narrative stimulus videos that featured salient action–outcome events (e.g., a car crash, a baseball batter hitting the ball out of the park) (Fig. 1A). By interrupting these action–outcome events, we elicited a jarring sense of surprise, creating a naturalistic analog to the stimuli used in prior rodent research.

In Study 1, we used a three-day behavioral paradigm (Fig. 2A), described in detail under Materials and Methods. On Day 1, participants ($N = 48$) viewed 18 target videos (Fig. 2B). On Day 2, we reactivated memories for the target videos by presenting them again, either in "match" form, identical to the original viewing, or "mismatch" form, with action–outcome event interrupted

Corresponding author: allie.sinclair@mail.utoronto.ca
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.046912.117.

Figure 1. (*A*) Frames from a target video depicting a baseball batter hitting a home run. The mismatch version of this video was interrupted, generating a prediction error by violating the action–outcome contingency. (*B*) Frames from the semantically related interference video.

(Fig. 2C, top). By interrupting the videos before the expected outcome, we sought to parallel the incomplete reminders (CS without US) presented in prior rodent studies (Exton-McGuinness et al. 2015). Following reactivation of each memory, we presented an interference video with new, but semantically related, content (e.g., a different car crash, a baseball fan in the stands catching a fly-ball) (Fig. 1B). On Day 2, we also assigned participants to either the Experimental or Control group. Experimental group participants received reactivation prior to interference (initiating the reconsolidation process), whereas Control group participants received interference prior to reactivation (Fig. 2A). Finally, on Day 3, participants completed an interview-style recall test on the target videos. We hypothesized that reactivating a memory while interrupting the expected action–outcome contingency, thus generating a prediction error, would render the memory trace labile and therefore susceptible to new information from the interference video.

Furthermore, we investigated whether qualities of the stimulus videos themselves influence the memory reconsolidation process. The uniquely detail-rich, narrative stimuli used in our paradigm allowed us to explore the effects of variation in content. In Study 2, we recruited a large independent sample of participants ($N = 408$) to rate the stimulus videos for emotional valence, emotional arousal, the surprising nature of the mismatch cut, and similarity between target and interference videos. We then conducted a by-video item analysis to test whether these qualities predicted false memories in the Study 1 sample.

Last, in Study 3, we tested whether our paradigm produced genuine updating of old memory traces, or merely source confusion or association between old and new information. Reconsolidation theory posits that reactivation allows an existing memory trace to be destabilized and updated. However, the Temporal Context Model (TCM) offers an alternative account of intrusions in human memory (Howard and Kahana 2002; Sederberg et al. 2011). Whereas reconsolidation theory proposes that intrusions result from the modification of an established memory trace, TCM purports that the original memory is left intact, and intrusions are encoded in a distinct memory trace. According to TCM, reactivating an old memory reinstates the original temporal context at the same time that the new distinct memory trace is being formed. The two memory traces thus share a temporal context, which can bridge old and new information and produce source confusion. In this way, intrusions can be explained by parsimonious principles of Hebbian association, without inferring that memory traces are altered.

To investigate this alternative account for our findings, we modified the timing of our paradigm to test competing predictions of reconsolidation theory and TCM (Fig. 3). Reconsolidation theory predicts that memory effects should be time-dependent, because the process of restabilizing a memory trace through protein synthesis takes several hours (Nader et al. 2000; Debiec et al. 2002). Therefore, if we condense our paradigm to conduct the memory test on Day 2 instead of Day 3, effects should disappear because there would be insufficient time for the protein synthesis required to incorporate new information into a memory trace (Fig. 3A). In contrast, TCM makes no predictions about a delay being necessary—only that two memory traces must share a temporal context (Fig. 3B). Indeed, previous studies have sought to test these competing theories with similar manipulations, and have found that intrusions were reduced when there was no delay before test—as predicted by reconsolidation theory (Hupbach et al. 2007). However, an alternative interpretation is that the immediate test reduced intrusions because participants relied upon a recall-to-reject strategy to improve source monitoring (Sederberg et al. 2011). That is, when participants receive interference immediately before test, they may benefit from a recency effect that allows them to better distinguish between old and new information. When tested immediately after learning, participants can more easily access the new information than when tested after a 24-h delay.

To preempt this recall-to-reject claim, we included an additional manipulation: instead of presenting reactivation and interference videos in blocks, we made presentation interleaved (e.g., "Baseball" target video would be immediately followed by the corresponding "Baseball" interference video). Reconsolidation theory predicts that after reactivation, synaptic destabilization takes between 3–10 min due to protein degradation processes (Suzuki et al. 2004; Lee et al. 2008; Bustos et al. 2009). Thus, when reactivation and interference videos are interleaved, there is not enough time for a memory trace to be destabilized prior to interference (Fig. 3C). Critically, TCM makes a strong prediction in the opposite direction: if target and interference videos are presented closer in time, the association between them will be even stronger, exacerbating our effects (Fig. 3D; Howard and Kahana 2002; Sederberg et al. 2011).

Results

Study 1: prediction error influences reconsolidation, updating memories

We scored recall transcripts for four measures: intrusions, errors, correct details, and confidence. We operationally defined

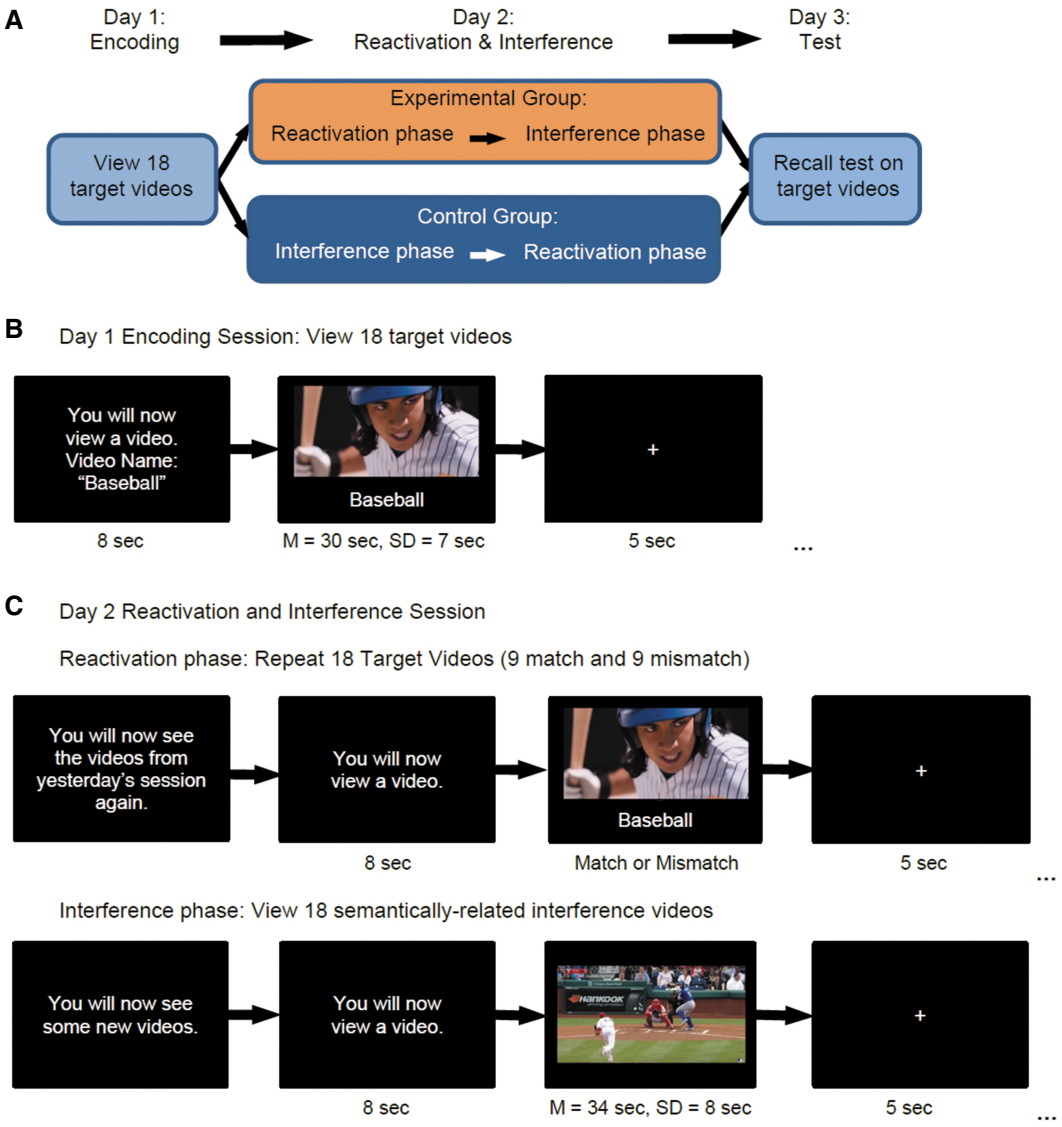

Figure 2. (*A*) Overview of the 3-d paradigm. (*B*) Example trial from the Day 1 Encoding Session, during which participants watched the 18 target videos. (*C*) On Day 2, participants were assigned to either the Experimental group (Reactivation before Interference) or Control group (Interference before Reactivation). During Reactivation, we replayed target videos, half in the same form as before (match) and half in an altered form (mismatch). During Interference, participants viewed 18 semantically related interference videos. On Day 3, participants completed an interview-style recall test on the original target videos.

"intrusions" (Fig. 4A) as visual and auditory details from an interference video mistakenly attributed to its corresponding target video (e.g., a character's gender/ethnicity/age/clothing, setting, dialogue). We defined "errors" (Fig. 4B) as incorrect details that were in neither the target nor the interference videos. Intrusions and void responses (e.g., "I don't remember") were not scored as errors. To measure the overall richness and accuracy of recall, we also coded transcripts for the total number of unique "correct details" (Fig. 4C) recalled from each video. Importantly, void responses were missed opportunities for earning correct details points. Finally, we also measured participants' "confidence" (Fig. 4D) in their memories through self-reported ratings on a Likert scale ranging from 1 ("not at all confident") to 5 ("very confident"). For each measure, we conducted 2 × 2 mixed ANOVAs to assess the differences among means for the between-subjects factor "group" (Experimental and Control) and the within-subjects factor "reactivation type" (match and mismatch). Descriptive statistics are provided in Table 1.

Our intervention produced striking between-subjects and within-subjects differences in false memories. For the intrusion rate (Fig. 3A), there was a significant interaction between reactivation type and group, $F_{(1,46)} = 9.64$, $P = 0.003$, $\eta_p^2 = 0.173$, with a large effect size. As hypothesized, for the Experimental group (reactivation before interference), mismatch-style reactivation produced significantly more intrusions than match-style, $t_{(23)} = 4.93$, $P < 0.001$, $d = 1.11$, 95% CI = [0.254, 0.623]. In contrast, in the Control group (interference before reactivation), there was no difference in the intrusion rate for match- and mismatch-reactivated videos, $t_{(23)} = 0.76$, $P = 0.453$, $d = 0.157$, 95% CI = [−0.232, 0.107]. As a follow-up, we also conducted a by-video item analysis. Videos produced more intrusions in the Experimental group when presented in mismatch form than in match form, $t_{(17)} =$

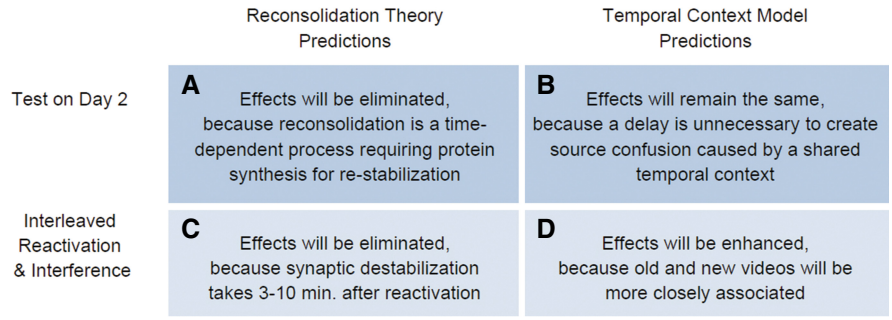

Figure 3. Competing predictions of Reconsolidation Theory and the TCM. In our modified paradigm, reactivation and interference videos are now presented in an interleaved, alternating fashion, and participants are tested immediately afterwards. (*A*,*B*) Reconsolidation Theory predicts that an immediate test will eliminate between-subjects and within-subjects differences for intrusions and errors, whereas TCM does not. (*C*,*D*) Similarly, Reconsolidation Theory predicts that interleaved reactivation and interference will eliminate effects, whereas TCM predicts that effects will be enhanced.

-3.66, $P = 0.002$, $d = 0.893$, 95% CI = [−0.675, −0.181]. Again, there was a significant interaction between group and reactivation type, $F_{(1,16)} = 7.92$, $P = 0.008$, $\eta_p^2 = 0.189$, demonstrating that this effect of mismatch-reactivation was selective for participants in the Experimental group. This finding supports the hypothesis that a prediction error can effectively destabilize a memory, rendering it temporarily vulnerable to modification via subsequent interference. Critically, this interaction demonstrates that prediction error influences the reconsolidation process, but does not affect memory when reconsolidation does not occur (i.e., in the Control group). As mismatch-reactivation did not produce more intrusions in the Control group, our findings cannot be explained by the fact that participants did not have the opportunity to review the entire mismatch video.

Overall, the intrusion rate (Fig. 4A) was higher in the Experimental group than in the Control group, $F_{(1,46)} = 22.25$, $P < 0.001$, $\eta_p^2 = 0.326$, 95% CI = [0.262, 0.653]. The large effect size provides compelling evidence in favor of episodic memory reconsolidation, as participants in the Control and Experimental groups viewed all of the same stimuli on Day 2 with only the order of the session altered. Similarly, the overall error rate (Fig. 4B) was also higher in the Experimental group than in the Control group, $F_{(1,46)} = 11.33$, $P = 0.002$, $\eta_p^2 = 0.2$, 95% CI = [0.222, 0.884]. Taken together, these dramatic between-subjects differences demonstrate that episodic memories are altered via reconsolidation. Moreover, it is worth noting that some of the details which we scored as errors may have actually been intrusions from other sources that we cannot conclusively identify. We defined our intrusions measure conservatively, counting only details that were in the specific interference video selected for a given target video. However, participants may have perceived other videos in the stimulus set to be semantically related at a broader level (e.g., all sports-related videos). In part, the higher error rate in the Experimental condition may reflect the same memory updating process which produced the intrusions. Given the richly detailed nature of our stimulus videos and the memories they evoke, we are unable to tease apart relevant intrusions from other videos and random errors.

Additionally, our prediction-error manipulation (match versus mismatch) produced overall within-subjects effects. The intrusion rate (Fig. 4A) was higher for mismatch-reactivated videos than for match-reactivated videos, $F_{(1,46)} = 17.14$, $P < 0.001$, $\eta_p^2 = 0.271$, 95% CI = [0.129, 0.372], with a large effect size. However, reactivation type did not affect the rate of other errors, $F_{(1,46)} = 0.22$, $P = 0.639$, $\eta_p^2 = 0.005$, 95% CI = [−0.184, 0.114]; correct details, $F_{(1,46)} = 1.2$, $P = 0.279$, $\eta_p^2 = 0.025$, 95% CI = [−0.263, 0.892]; or confidence ratings, $F_{(1,46)} = 1.7$, $P = 0.197$, $\eta_p^2 = 0.036$, 95% CI = [−0.141, 0.03].

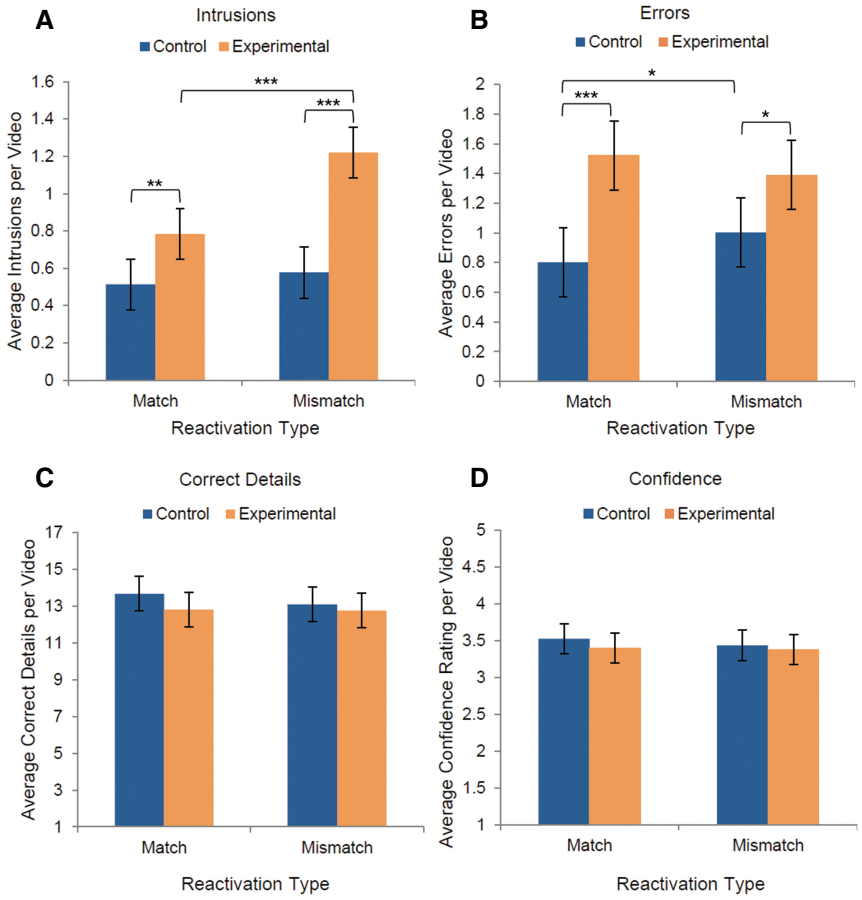

Figure 4. Mean values for the four measures, by condition and reactivation-type. Mismatch reactivation generated a prediction error, whereas Match reactivation did not. Panels depict (*A*) intrusions, false memories from the interference videos; (*B*) errors, false memories that were not classified as intrusions; (*C*) correct details recalled from the target videos; and (*D*) confidence, self-reported ratings. Error bars depict 95% confidence intervals of the mean. (*) $P < 0.05$, (**) $P < 0.01$, and (***) $P < 0.001$.

Table 1. Study 1: descriptive statistics by group and reactivation type

	Control group			Experimental group		
Measure	Mean	SD	95% CI	Mean	SD	95% CI
Intrusions, *M*	0.51	0.27	[0.382, 0.645]	0.78	0.36	[0.652, 0.915]
Intrusions, *MM*	0.58	0.32	[0.388, 0.765]	1.22	0.56	[1.033, 1.411]
Errors, *M*	0.8	0.37	[0.563, 1.043]	1.52	0.74	[1.282, 1.762]
Errors, *MM*	1	0.55	[0.732, 1.276]	1.39	0.76	[1.119, 1.663]
Correct details, *M*	13.66	1.83	[12.725, 14.59]	12.8	2.63	[11.871, 13.736]
Correct details, *MM*	13.08	2.56	[11.971, 14.19]	12.75	2.83	[11.643, 13.862]
Confidence, *M*	3.53	0.44	[3.302, 3.755]	3.4	0.64	[3.173, 3.636]
Confidence, *MM*	3.44	0.45	[3.438, 3.41]	3.38	0.53	[3.156, 3.607]
Confidence, *w/intrusions*	3.41	0.6	[3.17, 3.65]	3.33	0.58	[3.09, 3.56]
Confidence, *no intrusions*	3.53	0.46	[3.34, 3.71]	3.43	0.65	[3.17, 3.69]
Confidence, *w/errors*	3.42	0.51	[3.22, 3.63]	3.38	0.61	[3.13, 3.62]
Confidence, *no errors*	3.48	0.53	[3.27, 3.69]	3.35	0.65	[3.3, 3.4]

Note: *M*, Match form; *MM*, Mismatch form.
Confidence ratings are divided by videos with and without intrusions and errors, providing a more sensitive measure of metamemory accuracy.

For the error rate (Fig. 4B), there was a significant interaction between reactivation type and condition, $F_{(1,46)} = 5.03$, $P = 0.03$, $\eta_p^2 = 0.1$, driven by the higher error rate for mismatch-reactivated videos in the Control group, $t_{(23)} = 2.35$, $P = 0.028$, $d = 0.52$, 95% CI = [0.377, 0.24]. Participants may have made more errors because they were unable to review the entire video on Day 2. In the Experimental group, reactivation type did not influence the error rate, $t_{(23)} = 1.085$, $P = 0.289$, 95% CI = [−0.119, 0.38], suggesting that our prediction error intervention selectively updated memories with semantically related information, thus producing more intrusions but not necessarily more errors. Consistent with this, our by-video item analysis revealed that match and mismatch forms of each video did not produce different error rates in the Experimental group, $t_{(17)} = 1.39$, $P = 0.183$, $d = 0.33$, 95% CI = [−0.079, 0.385], or in the Control group, $t_{(17)} = -1.23$, $P = 0.235$, $d = 0.08$, 95% CI = [−0.38, 0.1]. (Note that because variances differed between groups for the intrusion and error measures, we also conducted nonparametric tests to verify the results, described in the Additional Analyses section of the Supplemental Material.)

Correct details did not differ between the Experimental and Control groups, $F_{(1,46)} = 0.8$, $P = 0.375$, $\eta_p^2 = 0.017$, 95% CI = [−0.738, 1.921], nor was there a significant interaction, $F_{(1,46)} = 0.842$, $P = 0.364$, $\eta_p^2 = 0.018$ (Fig. 4C). Participants from both groups successfully recalled content from the target videos, suggesting that our paradigm selectively updated memories by preserving old content while adding new information.

Average confidence ratings were above "moderately confident" (Fig. 4D), and did not differ between Experimental and Control groups, $F_{(1,46)} = 0.4$, $P = 0.531$, $\eta_p^2 = 0.009$, 95% CI = [−0.2, 0.383]. There was no interaction between group and reactivation type for confidence ratings, $F_{(1,46)} = 0.77$, $P = 0.386$, $\eta_p^2 = 0.016$. However, to further investigate whether participants experienced general source confusion or lower confidence in false memories, we conducted an additional within-subjects analysis to compare videos with reported intrusions to those without. We found that participants were no less confident in their memories when they reported intrusions in their recall, both in the Experimental group, $t_{(23)} = -1.38$, $P = 0.181$, $d = 0.281$, 95% CI = [−0.254, 0.051], and in the Control group, $t_{(23)} = -0.97$, $P = 0.344$, $d = 0.197$, 95% CI = [−0.369, 0.134]. Similarly, confidence did not differ by the presence or absence of errors in recall, either in the Experimental group, $t_{(23)} = 0.26$, $P = 0.8$, $d = 0.052$, 95% CI = [−0.194, 0.249], or the Control group, $t_{(23)} = -0.53$, $P = 0.604$, $d = 0.107$, 95% CI = [−0.275, 0.164]. Overall, it appears that metamemory judgments are not sensitive to memory distortions.

Study 2: qualities of the stimulus videos are associated with reconsolidation effects

We next conducted an item analysis to investigate associations among the memory measures from Study 1 (intrusions, errors, correct details, and confidence) and the video ratings for four attributes ("valence," "arousal," "surprise," and "similarity") that were collected from an independent sample of participants. We first assessed bivariate correlations between all measures, split by group and reactivation type (Supplemental Table S6). This initial analysis revealed that for participants in the Experimental group, surprise was positively correlated with mismatch intrusions, $r(16) = 0.489$, $P = 0.04$ (Fig. 5), and emotional arousal was negatively correlated with match intrusions, $r(16) = -0.478$, $P = 0.045$ (Fig. 6).

However, in order to account for both by-video and by-participant variability (which makes observations nonindependent), we conducted multilevel regression modeling, which can efficiently resolve these problems in a single omnibus test (Baayen et al. 2008). In a linear mixed-effects model, we predicted intrusions from the fixed factors group (Control and Experimental), reactivation type (Match and Mismatch), and "surprise ratings". In this model, the crossed random-effects were "subject" (identity of each participant) and "video" (identity of each stimulus item). We assessed significance of fixed-effects with a type III ANOVA using Kenward–Roger approximations of degrees of freedom, a method which has been shown to produce acceptably conservative Type I error rates, even with relatively smaller sample sizes (Luke 2017). Further information on model construction is provided in the Materials and Methods section (Study 2: Analyses), and descriptive statistics for parameter estimates are provided in Table 2 (left).

We found that surprise ratings were positively associated with intrusions, $F_{(1,16)} = 5.64$, $P = 0.03$, 95% CI = [0.178, 1.84] (Fig. 5). In other words, videos that were rated to be more surprising when interrupted produced the most memory updating. Moreover, there was a significant interaction between surprise and reactivation type, $F_{(1,807)} = 4.8$, $P = 0.029$, 95% CI = [0.037, 0.636], demonstrating that surprise ratings were only related to intrusions when the video was actually cut short at reactivation. Surprise was significantly associated with mismatch intrusions, $r(34) = 0.38$, $P = 0.023$, but not match intrusions, $r(34) = 0.28$, $P = 0.1$. In other words, surprise ratings reflect the response to the interruption, rather than capturing the general salience of events in a given video. There was a nonsignificant trend toward an interaction between surprise ratings and group, $F_{(1,775)} = 2.21$, $P = 0.137$, 95% CI = [−0.071, 0.522], as well as toward a three-way interaction of

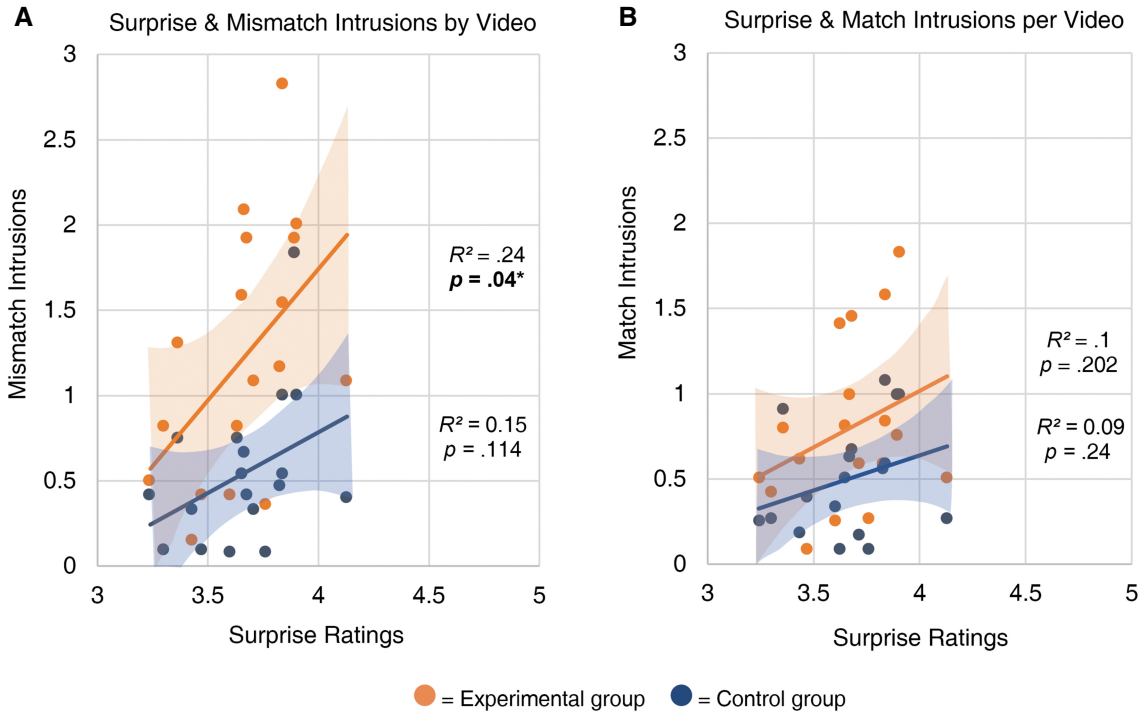

Figure 5. Surprise ratings were significantly correlated with intrusions for mismatch (*A*), but not match (*B*), forms of the videos. Videos that were more surprising when interrupted produced more intrusions. Each point represents a target video. Intrusion scores are subsetted by group. Shaded areas depict 95% confidence bands for the line of best fit.

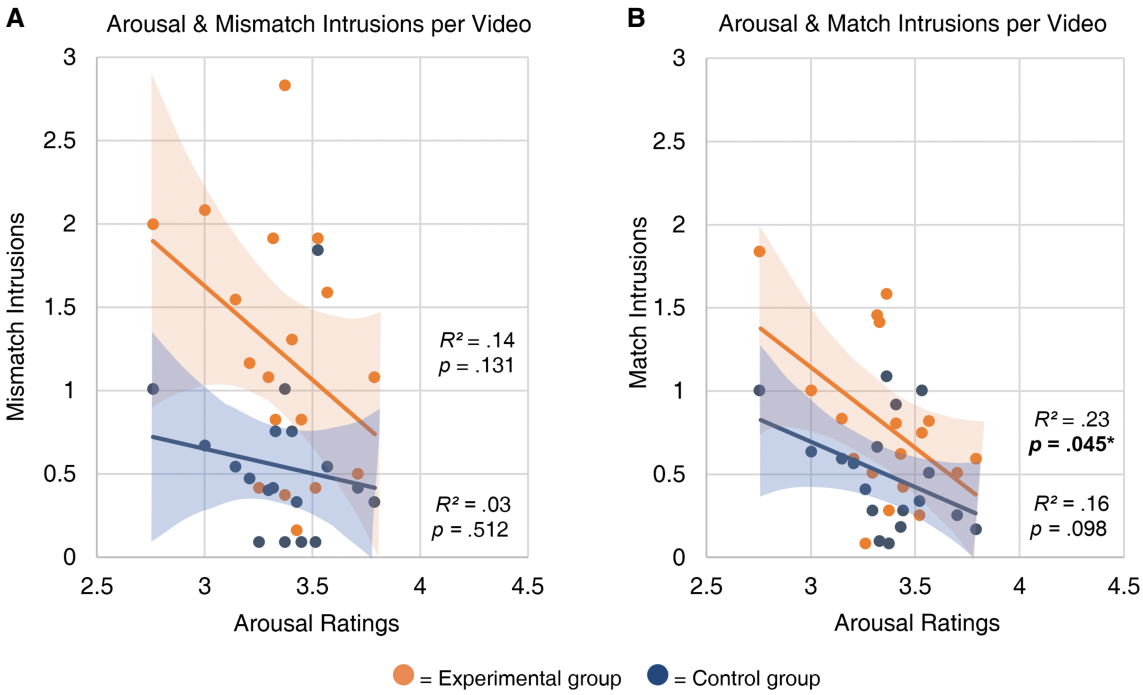

Figure 6. Emotional arousal ratings were negatively correlated with intrusions, selectively for participants in the Experimental group. This association was stronger for match-reactivation (*B*) than for mismatch-reactivation (*A*). Regardless of valence, memories with strong emotional content were more resistant to change through reconsolidation, especially when reactivated in the absence of prediction error. Each point represents a target video. Intrusion scores are subsetted by group. Shaded areas depict 95% confidence bands for the line of best fit.

Table 2. Study 2: β estimates for linear mixed effects models

	Surprise rating model			Arousal rating model		
Predictor	***b***	**SE**	**95% CI**	***b***	**SE**	**95% CI**
Intercept	0.77***	0.1	[0.575, 0.964]	0.77***	0.11	[0.563, 0.972]
Rating	1.01*	0.43	[0.178, 1.84]	−0.76	0.41	[0.563, 0.972]
Group (Exp. > Con.)	0.25***	0.05	[0.147, 0.35]	0.25***	0.05	[0.147, 0.35]
Reactivation (MM > M)	0.12***	0.03	[0.055, 0.183]	0.12***	0.03	[0.054, 0.183]
Rating*Group	0.23	0.15	[−0.071, 0.522]	−0.29*	0.14	[−0.554, −0.016]
Rating*Reactivation	0.34*	2.15	[0.037, 0.636]	−0.03	0.14	[−0.303, 0.241]
Group*Reactivation	0.1**	0.03	[0.032, 0.16]	0.1**	0.13	[0.031, 0.16]
Rating*Group*Reactivation	0.23	0.15	[−0.072, 0.527]	−0.16	0.14	[−0.432, 0.112]

Note: Rating refers to Surprise or Arousal, respectively. Estimates reflect the slope of the regression line for each fixed factor.
M , Match form; *MM*, Mismatch form.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

surprise ratings, group, and reactivation type, $F_{(1,808)} = 2.2$, $P = 0.138$, 95% CI = [−0.072, 0.527]. Overall, we found that surprise was positively associated with intrusions, specifically when videos were reactivated in mismatch form. This effect was stronger for participants in the Experimental group than for those in the Control group, though not significantly so; mismatch reactivation may also enhance memory change in the absence of reconsolidation, in accordance with past research demonstrating neural differentiation following prediction error (Kim et al. 2017). Last, reproducing the results of previously reported analyses, intrusions were significantly higher for the Experimental group than the Control group, $F_{(1,34)} = 24$, $P < 0.001$, 95% CI = [0.147, 0.35], and for mismatch than match reactivation, $F_{(1,775)} = 13.12$, $P < 0.001$, 95% CI = [0.055, 0.183]. The interaction between group and reactivation was also significant, $F_{(1,775)} = 8.59$, $P = 0.003$.

Next, we conducted the same linear mixed effects analysis for emotional arousal ratings (Table 2, right). There was a trending main effect of arousal associated with lower intrusion rates, $F_{(1,16)} = 3.36$, $P = 0.086$. There was a significant interaction between arousal and group, $F_{(1,774)} = 4.31$, $P = 0.038$, demonstrating that high emotional arousal was associated with fewer intrusions, but this effect is stronger in the Experimental group. In other words, memories with strong emotional content, regardless of valence, are resistant to change via the reconsolidation process. Numerically, this association was weaker for mismatch-reactivated videos (Fig. 6A) than for match-reactivated videos (Fig. 6B), driven by greater variability in mismatch intrusions, particularly for videos with moderately high arousal. However, the three-way interaction between arousal, group, and reactivation type was not significant, $F_{(1,808)} = 1.32$, $P = 0.25$. Overall, there was a protective effect of strong emotional arousal.

Additionally, our bivariate correlation analysis affirmed that target and interference videos were comparable in emotional and semantic content, as intended. Emotional arousal ratings for target and interference videos were positively correlated with each other, $r(16) = 0.654$, $P = 0.003$, as were valence ratings, $r(16) = 0.661$, $P = 0.003$. Moreover, target and interference videos were rated to be between "moderately similar" and "very similar" ($M = 3.47$, $SD = 0.45$), though similarity ratings were not associated with any other measures. There were no significant correlations between emotional valence ratings and memory measures in either group, suggesting that the effects of surprise and emotional arousal may apply to both positive and negative experiences. However, it should be noted that the stimulus set was not designed to cover a broad spectrum of emotional content. Overall, these intriguing exploratory findings bear implications for therapeutic applications to PTSD, and warrant future research investigating the relationship between prediction error and emotional memory.

Study 3: reconsolidation theory and the TCM describe distinct, but complementary, memory updating processes

Last, we modified the timing of our paradigm (Fig. 3) to test competing predictions of reconsolidation theory and the TCM, an alternative account of intrusions in human memory. As in Study 1, we scored recall transcripts according to the same procedure and conducted 2 × 2 mixed ANOVAs to assess the differences among means for the between-subjects factor group (Experimental and Control) and the within-subjects factor reactivation type (match and mismatch). Descriptive statistics are provided in Table 3. Strikingly, the subtle changes to the timing of our paradigm completely eliminated all between-subjects and within-subjects differences in false memories. In stark contrast to the effects observed in Study 1, were no significant differences between Control and Experimental groups for intrusions, $F_{(1,22)} = 0.27$, $P = 0.61$, $\eta_p^2 = 0.012$, 95% CI = [−0.122, 0.203] (Fig. 7A), or errors,

Table 3. Study 3: descriptive statistics by group and reactivation type

	Control group			Experimental group		
Measure	**Mean**	**SD**	**95% CI**	**Mean**	**SD**	**95% CI**
Intrusions, *M*	0.51	0.26	[0.343, 0.675]	0.51	0.25	[0.356, 0.672]
Intrusions, *MM*	0.66	0.36	[0.427, 0.888]	0.57	0.28	[0.392, 0.751]
Errors, *M*	0.95	0.4	[0.701, 1.206]	0.76	0.44	[0.476, 1.037]
Errors, *MM*	0.66	0.38	[0.419, 0.908]	0.71	0.29	[0.519, 0.892]
Correct Details, *M*	14.41	2.46	[12.843, 15.972]	12.56	2.75	[10.809, 14.309]
Correct Details, *MM*	13.53	3.18	[11.502, 15.548]	12.62	2.68	[10.916, 14.323]
Confidence, *M*	3.37	0.62	[2.98, 3.766]	3.46	0.39	[3.212, 3.702]
Confidence, *MM*	3.49	0.61	[3.1, 3.879]	3.58	0.56	[3.218, 3.949]

Note: M, Match form; *MM*, Mismatch form.

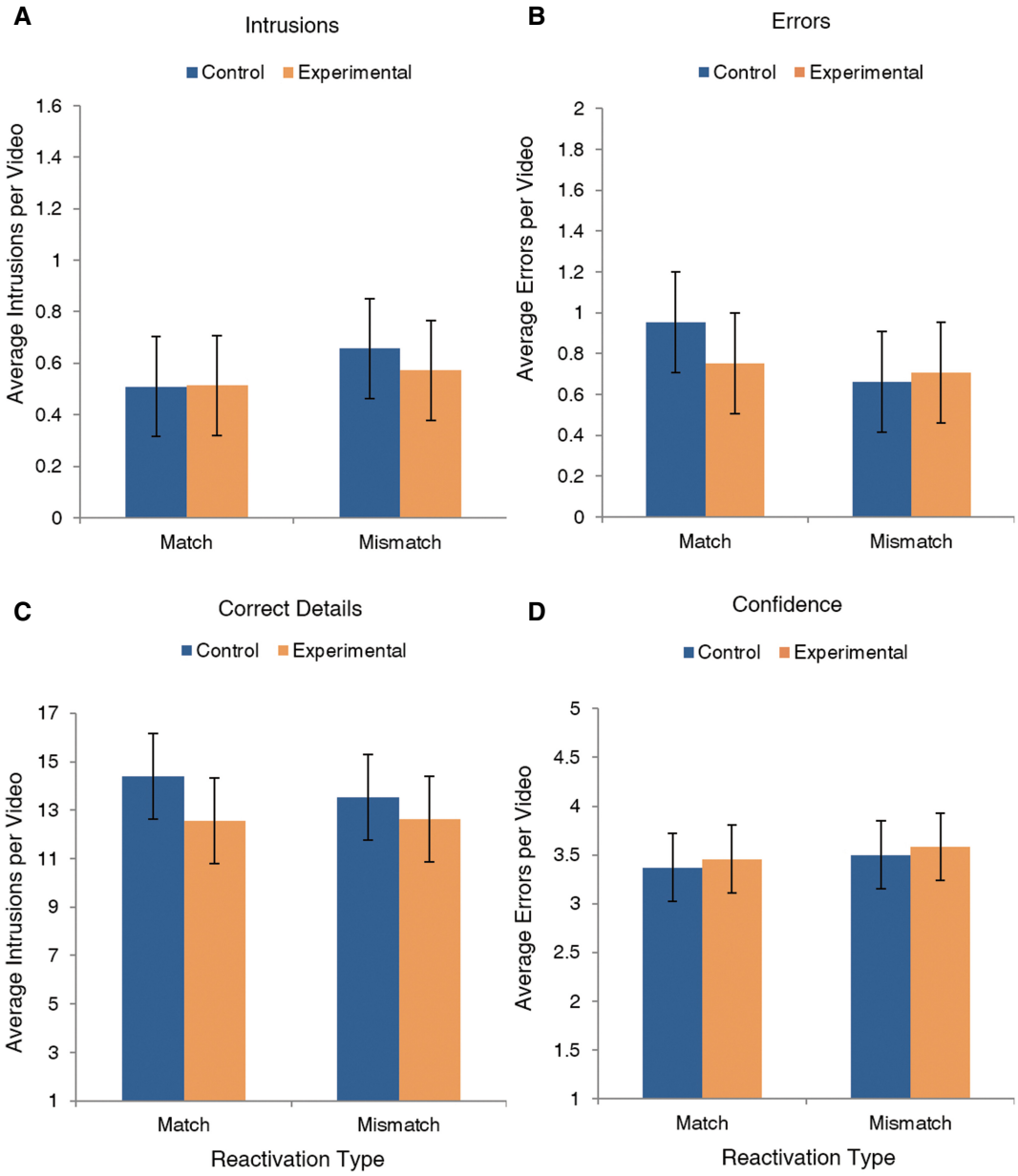

Figure 7. Mean values for intrusions (*A*), errors (*B*), correct details (*C*), and confidence (*D*), by group and reactivation type. Subtle changes to the timing of our paradigm completely obliterated all between-subjects and within-subjects differences, demonstrating that the effects observed in Study 1 are crucially time-dependent. Error bars depict 95% confidence intervals of the mean.

$F_{(1,22)} = 0.66$, $P = 0.424$, $\eta_p^2 = 0.029$, 95% CI = [−0.282, 0.123] (Fig. 7B). Moreover, the prediction error manipulation did not increase intrusions, $F_{(1,22)} = 1.31$, $P = 0.264$, $\eta_p^2 = 0.056$, 95% CI = [−0.289, 0.083], or errors, $F = 1.97$, $P = 0.174$, $\eta_p^2 = 0.082$, 95% CI = [−0.08, 0.418], and there were no significant interactions between group and reactivation type for intrusions, $F_{(1,22)} = 0.25$, $P = 0.62$, $\eta_p^2 = 0.011$, or errors, $F = 1.03$, $P = 0.321$, $\eta_p^2 = 0.045$. Furthermore, we directly compared the intrusion rates from Study 1 and Study 3, finding a significant three-way interaction among the factors reactivation type, group, and study, $F = 4.68$, $P = 0.034$, $\eta_p^2 = 0.064$. In other words, mismatch reactivation increased intrusions, but this effect was selective for the Experimental group within the Study 1 sample. The pairwise comparisons elucidating this interaction have been previously reported in the Study 1 Results. Similarly, the same three-way interaction was also significant for error rates, although this was in part driven by the higher error rate for mismatch-reactivated videos in the Study 1 Control group, $F = 4.49$, $P = 0.038$, $\eta_p^2 = 0.062$. Taken together, these null effects in the Study 3 sample affirm the predictions of reconsolidation theory, consistent with the known time-course of protein degradation and synthesis at the synapse.

Interestingly, in this study, we observed that intrusion and error rates for both groups were very similar to those observed in the Control group from Study 1. Because there were no differences between the Control and Experimental groups in Study 3, we compared the entire Study 3 sample to the Control group from Study 1. The Study 1 and Study 3 samples did not have significantly different intrusion rates, $t_{(46)} = -0.39$, $P = 0.697$, $d = 0.113$, 95% CI = [−0.141, 0.095], or error rates, $t_{(46)} = 1.34$, $P = 0.188$, $d = 0.386$, 95% CI = [−0.067, 0.332]. The TCM can account

for this nonzero baseline level of false memories, which cannot be explained by reconsolidation processes. In the Experimental group in Study 1, the higher rates of intrusions and errors above and beyond this baseline level may thus reflect a superadditive effect of two distinct processes, described by reconsolidation theory and TCM.

Consistent with Study 1, correct details (Figure 7C) did not differ by group, $F_{(1,22)} = 1.636$, $P = 0.214$, $\eta_p^2 = 0.069$, 95% CI = [–0.856, 3.61; reactivation type, $F_{(1,22)} = 1.291$, $P = 0.268$, $\eta_p^2 = 0.055$, 95% CI = –0.339, 1.16; or their interaction, $F_{(1,22)} = 1.699$, $P = 0.206$, $\eta_p^2 = 0.072$. Similarly, confidence ratings (Fig. 7D) did not differ by group, $F_{(1,22)} = 0.179$, $P = 0.676$, $\eta_p^2 = 0.008$, 95% CI = –0.498, 0.33; reactivation type $F_{(1,22)} = 1.567$, $P = 0.224$, $\eta_p^2 = 0.067$, 95% CI = –0.341, 0.084; or their interaction $F_{(1,22)} = 0.001$, $P = 0.982$, $\eta_p^2 = 0.001$. As before, participants successfully recalled content from the target videos, and reported moderately high confidence in their memories.

Discussion

Here, for the first time, we describe a fundamental mechanism that governs whether rich episodic memories will dynamically adapt to accommodate new information. The present program of research demonstrates that prediction error, or the element of surprise, renders complex memories malleable. Using our unique set of naturalistic stimulus videos, we generated prediction errors by abruptly interrupting action–outcome events. We found that prediction error destabilized memories and rendered them vulnerable to subsequent interference, modifying memories by adding relevant new content. Critically, however, prediction error only influenced memory when reactivation preceded interference, initiating the reconsolidation process. Moreover, we found that videos that were more surprising when interrupted produced more memory updating. Last, we tested competing hypotheses of reconsolidation theory and the TCM, an alternative account of source confusion. We demonstrated that our memory-updating effects were crucially time-dependent, as predicted by reconsolidation theory. However, the TCM can account for the nonzero level of false memories elicited by our control condition, which cannot be explained by reconsolidation. Bridging prior human and animal research, we implicate prediction error as a key signal which initiates the reconsolidation process and allows detailed, multimodal episodic memories to be adaptively altered.

In accordance with previous human reconsolidation research (e.g., Hupbach et al. 2007), when we reactivated memories and then presented semantically related interference, participants reported intrusions, details from the interference videos incorporated into the original memory. Critically, as predicted by reconsolidation theory (Alberini and Ledoux 2013), when we instead presented interference before reactivation, participants reported fewer intrusions and were unaffected by prediction error. We also found that the reconsolidation process produced more errors, defined as other false memories that were not from the specific interference video chosen for a given target video. However, participants may have perceived other videos in the stimulus set to be relevant at a broader level (e.g., all videos from sporting events are semantically related, not just the two baseball videos). Therefore, the higher error rate in the group that underwent reconsolidation may, in part, reflect the same adaptive updating process which produced the intrusions.

Furthermore, as first demonstrated in animal research (Díaz-Mataix et al. 2013), we found that reactivating a memory in conjunction with a prediction error produced more intrusions than reactivation without. Importantly, prediction error only influenced memory in the Experimental group, suggesting that expectancy violations selectively interact with the reconsolidation process to trigger memory updating. Despite their false memories, Experimental group participants still recalled correct details and reported moderately high memory confidence. We therefore successfully added relevant new information to destabilized memories without annihilating the existing trace, bolstering the theoretical interpretation of reconsolidation as an adaptive updating mechanism (Exton-McGuinness et al. 2015). Moreover, we found that prediction error specifically enhanced semantically related updating, potentially resolving past discrepant findings: prior reconsolidation studies have demonstrated that interference produces intrusions (Forcato et al. 2010; Hupbach et al. 2013), but does not weaken memories when participants are not permitted to choose interference items at test (Levy et al. 2017).

Last, we considered an alternative account of intrusions in human memory, the TCM. Whereas reconsolidation theory proposes that an old memory trace can be destabilized and modified, TCM proposes that old and new memories are encoded as distinct memory traces (Sederberg et al. 2011; Klingmüller et al. 2017). However, because reactivation of an old memory reinstates internal and external contextual factors, new information becomes associated with the old memory. We modified our paradigm to test competing predictions of these two theoretical frameworks, and demonstrated that subtle changes to the timing of reactivation and interference obliterated all effects of group and reactivation type. This time dependency of our effects is consistent with the synaptic mechanisms of reconsolidation. Furthermore, our findings can explain failed replications in the reconsolidation literature that arise from designs with insufficient delays between encoding, reactivation, and test (Allanson and Ecker 2017). Importantly, however, TCM can account for the nonzero level of intrusions and errors observed in the Study 1 Control group. We implicate reconsolidation as a distinct, but complementary, memory updating process.

Implications and future directions

Understanding the mechanisms that make our memories susceptible to change can inform practices for eyewitness testimony, inspire opportunities for cognitive enhancement, and motivate innovative treatments for psychiatric conditions such as posttraumatic stress disorder (PTSD). Some evidence suggests that reconsolidation-based amnestic interventions can ameliorate emotionally charged intrusive memories in PTSD (Brunet et al. 2008; Kroes et al. 2013; Schwabe et al. 2013; James et al. 2015). However, other studies have produced inconsistent and contradictory results (Golkar et al. 2012; Lázaro-Muñoz and Diaz-Mataix 2016). Here, we demonstrated that memories with strong emotional content, regardless of valence, were more resistant to change. Moving forward, designing treatments that consider the boundary conditions of reconsolidation may help to resolve past discrepant findings and overcome this protective effect of emotional arousal. Importantly, our paradigm features naturalistic stimuli that can be better generalized to the vivid intrusive memories that characterize PTSD, shedding light on the mechanisms of episodic memory change.

Although reactivation with prediction error most effectively destabilized memories, reactivation without still produced intrusions. It may be that the reactivation process is mediated by the strength of the original memory trace, such that weak memories can be destabilized without a prediction error (Díaz-Mataix et al. 2013). Additionally, a prediction error generated on a single trial may "bleed over" to temporally adjacent trials, facilitating destabilization. Finally, prediction error may selectively destabilize components of a memory within an episode. As the bulk of the details our participants reported were present throughout the

video clips (e.g., setting, appearance of characters), we were unable to assess whether memory distortions were more likely to occur in close proximity to a prediction error.

In an item analysis, we demonstrated that mismatch videos that were rated to be more surprising (by an independent sample) produced more intrusions. This linear, positive relationship between surprise and intrusions suggests that prediction error parametrically influences memory updating. However, another intriguing possibility is that the incomplete nature of the reminder is the key to memory destabilization: it may be that incomplete reminders force the subject to engage in a process of active recall. Alternatively, destabilization may be nonlinearly related to reactivation strength, such that moderately strong reactivation destabilizes memories more than weak or strong reactivation (Kim et al. 2014). Reminders that lack this incomplete or surprising element have yielded mixed success in previous research (Walker et al. 2003; Hardwicke et al. 2016), underscoring the importance of future work investigating the boundary conditions that govern reconsolidation in humans.

Conclusion

For the first time, the present work has shown that prediction error plays a critical role in the reconsolidation of complex episodic memories. With unprecedented ecological validity, we demonstrate that surprise destabilizes memories of naturalistic events. Our findings characterize prediction error as an adaptive mechanism that allows real-world memories to be dynamically updated with relevant new information.

Materials and Methods

Study 1

Participants

We recruited 53 undergraduate Students from the University of Toronto to participate in the experiment for $30 compensation. As five of these participants failed to return for all three sessions, the final sample included 48 participants. Prior to beginning the experiment, we determined the sample size ($N = 48$) necessary to satisfy three conditions: (a) achieve at least 90% power to detect a medium-sized effect ($\eta_p^2 = 0.25$) in a 2×2 design (Faul et al. 2007); (b) approximate the sample sizes used in prior studies of reconsolidation in humans (e.g., Hupbach et al., 2007); and (c) evenly allocate participants to six pseudo randomized lists, described in detail in the Procedure. We estimated the effect size based on the results from a pilot study (described under Pilot Study and Supplemental Table S3 in the Supplemental Material).

The sample consisted of 60% women and 40% men (age $M =$ 21.4 yr, $SD = 2.83$, range 18–37). Inclusion criteria were as follows: fluent in English, normal or corrected-to-normal vision and hearing, and no history of neurological, or psychiatric disorders. The sample was ethnically heterogeneous: 37.5% East Asian, 18.8% South Asian, 16.7% Caucasian, 8.3% Southeast Asian, 6.3% African–American, 4.2% Middle Eastern, 2.1% Latino, and 6.3% Other. The study was approved by the Ethics Committee at the University of Toronto. All participants provided written informed consent before beginning the experiment on Day 1.

Materials

We presented the video viewing sessions for Day 1 and Day 2 with E-Prime (Psychology Software Tools, Inc., Version 2.0) on a laptop computer. On Day 2, the participants also completed the Morningness–Eveningness Questionnaire (MEQ) online (Terman et al. 2001). The results from the MEQ are unrelated to the primary aims of the study, but can be found under Additional Analyses in the Supplemental Material.

Participants completed the Day 1 and Day 2 video viewing sessions alone in a dark, enclosed testing room with a desk and chair. We provided participants with headphones to listen to the audio tracks of the videos. On Day 3, participants completed the recall test in the same testing room as before, but with the lights on and the experimenter present. We recorded participants' responses with a video recorder and later transcribed the interviews for analysis.

Video stimuli

Thirty-six videos were used in the experiment (18 target and 18 interference); these videos are described in Supplemental Table S1 and provided online via the Open Science Framework. The target videos consisted of 18 short video clips (duration $M = 30$ sec, $SD = 6.96$ sec). No events or characters repeated across the 18 target videos, which we sourced from film and television clips found online.

During the Day 2 Reactivation Phase, each participant viewed half of the target videos in "match" form and half in "mismatch" form. (Note that each target video had a mismatch form; whether a participant viewed the match or mismatch form was fully counterbalanced across participants.) The match videos were identical to the target videos first shown on Day 1. In contrast, the mismatch videos cut off abruptly before the final seconds of the video, interrupting the salient event before its completion (e.g., a car crash cut off immediately before impact, a baseball batter cut off mid-swing) (Fig. 1A; Supplemental Table S1). During the Day 2 Interference Phase, participants also viewed 18 novel interference videos. Each interference video corresponded to one of the target videos, presenting a semantically related but distinct scene (e.g., a different car crash, a baseball fan in the stands catching a fly-ball) (Fig. 1B).

Procedure

Each participant returned to the laboratory for three sessions, each spaced 24 h apart (Fig. 2A). In order to keep contextual factors constant, the same experimenter (A.H.S.) worked with each participant in the same testing room for all three sessions (Hupbach et al. 2008).

Day 1: encoding session. Participants viewed the 18 target videos in a random order determined by the testing software (Fig. 2B). Before each video, participants viewed a screen (8 sec) which provided a name for the video to follow (e.g., "Baseball"). The name of the video remained on-screen below the video while it played. After each video, participants viewed a black screen with a fixation cross for 5 sec. We instructed participants to pay attention to the videos and their corresponding names, as there would be a memory test on Day 3.

Day 2: reactivation and interference session. Participants were sequentially allocated to either the Experimental group or Control group (each described below). Participants in the Experimental group completed the Reactivation phase prior to the Interference phase; participants in the Control group completed the Interference phase prior to the Reactivation phase (Fig. 2C). At the beginning of the session, the experimenter asked the participant approximately how many hours he/she had slept the night before. We collected this information because sleep is a critical factor for memory consolidation and reconsolidation.

Experimental group. Before beginning the viewing session, we informed participants that they would see the videos from Day 1 again. We sequentially assigned each participant to one of six pseudo random ordered lists of match and mismatch reactivation videos, counterbalancing such that an equal number of participants from the Experimental and Control groups received each list. We also counterbalanced the lists such that each target video was reactivated in match form for half of the participants and mismatch form for the other half. To ensure that the mismatch videos still generated a sense of surprise throughout the session, we prepared each list so that there were never more than two consecutive

mismatch videos. Although the first experience of a mismatch video is likely always the most surprising, our video clips featured action–outcome events that drew on prior expectations about the sequence of events in the real world (e.g., one has a strong expectation that a baseball batter will finish swinging the bat). Therefore, interruptions likely still feel surprising even if the participant is aware that some of the videos will be cut short.

Before each video, participants viewed a screen (8 sec) which stated, "You will now view a video." We did not inform participants whether they would be seeing a match or mismatch video on each trial. As in the Day 1 session, we displayed the name of each target video at the bottom of the screen as the video played (Fig. 2C, top). After each video, participants viewed a black screen with a fixation cross (5 sec).

Following this reactivation procedure, the participants completed the MEQ during a 7-min interlude. We included this delay because evidence from animal studies suggests that at the synaptic level, the destabilization process takes 3–10 min (Monfils et al. 2009). After completing the MEQ, participants viewed the 18 novel interference videos (Fig. 2C, bottom), in the same pseudorandom order as their respective target videos had been presented in the Reactivation phase. We omitted names for the interference videos in order to account for prior critiques of human reconsolidation research, which have argued that incorrect associations between new content and an old context (i.e., the target video names) can explain intrusions (Klingmüller et al. 2017). The trial structure was the same as during the reactivation phase (8-sec information screen, video, 5-sec fixation cross). At the end of the session, participants reported whether they had seen any of the stimulus videos prior to the experiment.

Control group. The procedure for the Control group was identical to that of the Experimental group, except that we reversed the order of the session. Participants in the Control group viewed the interference videos before the match and mismatch reactivation target videos. According to reconsolidation theory, a memory must be reactivated before it can be destabilized and made vulnerable to change. Thus, interference presented before reactivation should have less effect on the original memory.

Day 3: testing session. The experimenter again asked the participant approximately how many hours he/she had slept the night before. The participant then completed a memory test on the 18 target videos, in a random order determined by the testing software. We emphasized that the participant was to answer based on his/her memory of the original target videos, not the interference videos viewed on Day 2. Participants were permitted to provide void responses (e.g., "I don't remember,") and were explicitly instructed not to guess or confabulate. The experimenter verbally cued the participant with the name of each target video. When the participant concluded his/her initial free-recall, the experimenter prompted him/her with a series of predetermined questions addressing any aspects of the video not already reported (e.g., "Can you describe what the driver of the car looked like?"). Supplemental Table S1 provides a full list of the interview questions. Following recall of each video, the participant verbally reported their overall memory confidence on a 5-point Likert scale. In total, the interview session lasted ~1-h. At the end of the session, participants were debriefed.

Analysis method

The experimenter (A.H.S) scoring the recall transcripts was fully blinded to both reactivation type and condition. In some cases, we did not use participants' reports of all 18 target videos. In brief, there were six instances in which we excluded videos from analysis because the participant had seen them prior to the experiment (Supplemental Table S2, "Knew"). In eight cases, participants entirely failed to recall a video when cued with its name (Supplemental Table S2, "Forgot"). Finally, there were six instances in which participants exhibited a total source-monitoring failure, describing only the interference video (Supplemental Table S2, "SM"). As our research question concerned updating of the original memory trace, we were specifically interested in evidence that information from the target and interference videos had been integrated, but not entirely confused. It was never necessary to exclude more than three of 18 total videos for any given participant.

Visual inspection of boxplots prior to analysis revealed that for the intrusion and error rates, there were several outliers in both the Control and Experimental groups. In order to attenuate their influence, we winsorized high and low outliers to the 95th and 5th percentiles, respectively. However, the results reported do not qualitatively change depending on whether outliers were unaltered, winsorized, or omitted from analysis. Statistical analyses were conducted with IBM SPSS Statistics (Version 24).

Inter–rater reliability. A second trained scorer, also blinded to group and reactivation type, independently coded half of the transcripts for intrusions and errors. The Pearson correlation between the two sets of intrusion scores was 0.82 for match-reactivated videos and 0.95 for mismatch-reactivated videos, demonstrating excellent inter-rater reliability. Similarly, the Pearson correlation between the two sets of error scores was 0.92 for match-reactivated videos and 0.9 for mismatch-reactivated videos.

Study 2

Participants

Participants were recruited via Amazon Mechanical Turk to participate in a 3-min survey for $0.50 compensation. Upon accepting the advertised task, participants were redirected to a Qualtrics survey. Completion of the survey yielded a confirmation code to be entered for credit on the Mechanical Turk website. In total, 443 participants completed the study. However, we excluded 35 participants for failing sanity checks (e.g., exhibiting poor attention), reported in the Supplemental Material (Study 2 Excluded Participants). The final sample consisted of 408 adults (Age $M = 34$ yr, $SD = 10.36$, range = 18–63) (62.3% male, 37.3% female, 0.5% other.

Procedure

Each participant viewed a trio of videos from the stimulus set: the match version of a target video, the mismatch version, and the corresponding interference video. Participants answered six questions about the videos, each on a five-point Likert scale. Due to the exclusions, the number of participants providing ratings for each triad ranged from 19 to 27. Descriptive statistics for the video ratings (Supplemental Table S4) and the item analysis (Supplemental Table S5) are provided in the Supplemental Material.

Before beginning the survey, participants provided informed consent with a digital signature. Participants first viewed the match version of a target video, embedded within the survey. They then rated on a scale of 1 to 5 the emotional valence ("very negative" to "very positive") and emotional arousal ("very weak" to "very strong") of their responses to the video. Participants then viewed the mismatch version of the same target video and rated how "surprising/unsettling/unexpected" it felt when the clip was interrupted ("very expected" to "very unexpected"). Prior to watching the mismatch video, participants were not informed that it would be cut short. Last, participants viewed the interference video and again provided emotional valence and arousal ratings, as well as a rating of how similar the target and interference videos were ("very different" to "very similar").

Analyses

We investigated the association between video ratings and memory measures by using linear mixed effects regression, with restricted maximum likelihood estimation. In RStudio (Version 1.1.442), we constructed the model with the *lme4* (Bates et al. 2014) and *lmerTest* packages (Kuznetsova and Christensen 2017), and obtained *P*-values with the car package (Fox and Weisberg 2011). These models included crossed-random effects to account for the by-video and by-participant variability: we included random intercepts (shifting the regression line) for each video and each subject.

Furthermore, because our "a priori" prediction was that the reactivation type manipulation would influence participants in the Control and Experimental groups differently, we included random slopes for the factor group (allowing each group to have its own regression line slope). Predictors were coded as follows—group: Control = −1, Experimental = 1, reactivation type: Match = −1, Mismatch = 1, rating: mean-centered surprise or arousal ratings. We calculated 95% confidence intervals on summary statistics by using a profile procedure (Bates et al. 2014). Parameter estimates for fixed-effects at the population-level are reported in Table 2.

Study 3

Participants

We recruited 24 young adults from the University of Toronto community to participate in the experiment for $25 compensation. Prior to beginning data collection, we performed a power analysis using the effect size for the within-between interaction term (intrusions measure, $\eta_p^2 = 0.173$) from Study 1 and determined that a sample size of 22 participants would yield 99% power (Faul et al. 2007). We increased the sample to 24, to evenly allocate participants to the six pseudo-random lists.

The sample consisted of 83% women and 17% men (age M = 21.8 yr, SD = 2.5, range 18–27). Inclusion criteria were as follows: fluent in English, normal or corrected-to-normal vision and hearing, and no history of neurological or psychiatric disorders. The sample was ethnically heterogeneous: 41.7% East Asian, 25% Southeast Asian, 12.5% Caucasian, 8.3% African–Canadian, 8.3% Middle Eastern, and 4.2% Hispanic. The study was approved by the Ethics Committee at the University of Toronto. All participants provided written informed consent before beginning the experiment on Day 1.

Procedure

The procedure was consistent with Study 1, with the exception of two modifications to the timing of the paradigm. First, the memory test was conducted on Day 2, immediately after the video viewing phase. Participants did not return for a Day 3 session. Second, during the video viewing phase on Day 2, we presented reactivation and interference videos in an interleaved fashion. Previously, in Study 1, we blocked presentation to group all old and all new videos together. In Study 2, old and new videos alternated, such that a reactivated target video and its corresponding interference video were presented as a pair. We scored and analyzed memory tests in the same manner as in Study 1.

As in Study 1, there were cases in which we did not use participants' reports of all 18 target videos, reported in full in Supplemental Table S7. In brief, there were three instances in which we excluded videos from analysis because the participant had seen them prior to the experiment (Supplemental Table S7, "Knew"). In four cases, participants entirely failed to recall a video when cued with its name (Supplemental Table S7, "Forgot"). Finally, there was one instance in which a participant exhibited a total source-monitoring failure, describing only the interference video (Supplemental Table S7, "SM"). It was never necessary to exclude more than two of 18 total videos for any given participant.

Data access

The full set of stimulus videos and the data set analyzed during the current study have been deposited in a public repository through the Open Science Framework, with the identifier DOI:10.17605/OSF.IO/GRNJW (Sinclair and Barense 2017).

Acknowledgments

A.H.S. and M.D.B. developed the study concept and study design. A.H.S. performed stimulus selection, data collection, interview transcription and scoring, and data analysis. A.H.S. drafted the manuscript, with input from M.D.B. All authors approved the final version of the manuscript for submission. This research was funded by an NSERC Discovery Grant and Accelerator Award to M.D.B., and a James S. McDonnell Scholar Award to M.D.B. M.D.B. also receives support from the Canada Research Chairs program.

References

Alberini CM, Ledoux JE. 2013. Memory reconsolidation. *Curr Biol* **23:** R746–R750.
Allanson F, Ecker UKH. 2017. No evidence for a role of reconsolidation in updating of paired associates. *J Cogn Psychol* **29:** 912–919.
Baayen RH, Davidson DJ, Bates DM. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* **59:** 390–412.
Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. Retrieved from http://arxiv.org/abs/1406.5823
Brunet A, Orr SP, Tremblay J, Robertson K, Nader K, Pitman RK. 2008. Effect of post-retrieval propranolol on psychophysiologic responding during subsequent script-driven traumatic imagery in post-traumatic stress disorder. *J Psychiatr Res* **42:** 503–506.
Bustos SG, Maldonado H, Molina VA. 2009. Disruptive effect of midazolam on fear memory reconsolidation: decisive influence of reactivation time span and memory age. *Neuropsychopharmacology* **34:** 446–457.
Chen J, Olsen RK, Preston AR, Glover GH, Wagner AD. 2011. Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and CA1 mismatch detection. *Learn Mem* **18:** 523–528.
Debiec J, LeDoux JE, Nader K. 2002. Cellular and systems reconsolidation in the hippocampus. *Neuron* **36:** 527–538.
Dębiec J, Bush DEA, LeDoux JE. 2011. Noradrenergic enhancement of reconsolidation in the amygdala impairs extinction of conditioned fear in rats-a possible mechanism for the persistence of traumatic memories in PTSD. *Depress Anxiety* **28:** 186–193.
Díaz-Mataix L, Ruiz Martinez RC, Schafe GE, LeDoux JE, Doyère V. 2013. Detection of a temporal error triggers reconsolidation of amygdala-dependent memories. *Curr Biol* **23:** 467–472.
Duncan K, Curtis C, Davachi L. 2009. Distinct memory signatures in the hippocampus: intentional states distinguish match and mismatch enhancement signals. *J Neurosci* **29:** 131–139.
Exton-McGuinness MTJ, Patton RC, Sacco LB, Lee JLC. 2014. Reconsolidation of a well-learned instrumental memory. *Learn Mem* **21:** 468–477.
Exton-McGuinness MTJ, Lee JLC, Reichelt AC. 2015. Updating memories—the role of prediction errors in memory reconsolidation. *Behav Brain Res* **278:** 375–384.
Faul F, Erdfelder E, Lang AG, Buchner A. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* **39:** 175–191.
Forcato C, Argibay P, Pedreira M, Maldonado H. 2009. Human reconsolidation does not always occur when a memory is retrieved: the relevance of the reminder structure. *Neurobiol Learn Mem* **91:** 50–57.
Forcato C, Rodríguez MLC, Pedreira ME, Maldonado H. 2010. Reconsolidation in humans opens up declarative memory to the entrance of new information. *Neurobiol Learn Mem* **93:** 77–84.
Forcato C, Bavassi L, De Pino G, Fernández RS, Villarreal MF, Pedreira ME. 2016. Differential left hippocampal activation during retrieval with different types of reminders: an fMRI study of the reconsolidation process. *PLoS One* **11:** e0151381.
Fox J, Weisberg S. 2011. *An {R} companion to applied regression, 2nd ed*. Sage, Thousand Oaks, CA. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.
Golkar A, Bellander M, Olsson A, Ohman A. 2012. Are fear memories erasable?-reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. *Front Behav Neurosci* **6:** 80.
Hardwicke TE, Taqi M, Shanks DR. 2016. Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proc Natl Acad Sci* **113:** 5206–5211.
Howard MW, Kahana MJ. 2002. A distributed representation of temporal context. *J Math Psychol* **46:** 269–299.
Hupbach A, Gomez R, Hardt O, Nadel L. 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learn Mem* **14:** 47–53.
Hupbach A, Hardt O, Gomez R, Nadel L. 2008. The dynamics of memory: context-dependent updating. *Learn Mem* **15:** 574–579.
Hupbach A, Gomez R, Nadel L. 2009. Episodic memory reconsolidation: updating or source confusion? *Memory* **17:** 502–510.
Hupbach A, Gomez R, Nadel L. 2013. Episodic memory reconsolidation: an update. *Mem Reconsol*. doi:10.1016/B978-0-12-386892-3.00011-1
James EL, Bonsall MB, Hoppitt L, Tunbridge EM, Geddes JR, Milton AL, Holmes EA. 2015. Computer game play reduces intrusive memories of experimental trauma via reconsolidation-update mechanisms. *Psychol Sci* **26:** 1201–1215.

Kim G, Lewis-Peacock JA, Norman KA, Turk-Browne NB. 2014. Pruning of memories by context-based prediction error. *Proc Natl Acad Sci* **111:** 8997–9002.
Kim G, Norman KA, Turk-Browne NB. 2017. Neural differentiation of incorrectly predicted memories. *J Neurosci* **37:** 2022–2031.
Klingmüller A, Caplan JB, Sommer T. 2017. Intrusions in episodic memory: reconsolidation or interference? *Learn Mem* **24:** 216–224.
Kroes MCW, Tendolkar I, van Wingen GA, van Waarde JA, Strange BA, Fernández G. 2013. An electroconvulsive therapy procedure impairs reconsolidation of episodic memories in humans. *Nat Neurosci* **17:** 204–206.
Kumaran D, Maguire EA. 2007. Match-mismatch processes underlie human hippocampal responses to associative novelty. *J Neurosci* **27:** 8517–8524.
Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest package: Tests in linear mixed effects models. *J Stat Software* **82:** 1–26.
Lázaro-Muñoz G, Diaz-Mataix L. 2016. Manipulating human memory through reconsolidation: stones left unturned. *AJOB Neurosci* **7:** 244–247.
Lee SH, Choi JH, Lee N, Lee HR, Kim JI, Yu NK, Choi SL, Lee SH, Kim H, Kaang BK. 2008. Synaptic protein degradation underlies destabilization of retrieved fear memory. *Science* **319:** 1253–1256.
Levy DA, Mika R, Radzyminski C, Ben-Zvi S, Tibon R. 2017. Behavioral reconsolidation interference with episodic memory is elusive. *PsyArXiv Preprints*. doi: 10.17605/OSF.IO/W273C
Long NM, Lee H, Kuhl BA. 2016. Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *J Neurosci* **36:** 12677–12687.
Luke SG. 2017. Evaluating significance in linear mixed-effects models in R. *Behav Res Methods* **49:** 1494–1502.
Monfils MH, Cowansage KK, Klann E, Ledoux JE. 2009. Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. *Science* **324:** 951–955.
Nader K, Schafe GE, Le Doux JE. 2000. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature* **406:** 722–726.
Pine A, Sadeh N, Ben-Yakov A, Dudai Y, Mendelsohn A. 2018. Knowledge acquisition is governed by striatal prediction errors. *Nat Commun* **9:** 1673.
Schwabe L, Nader K, Pruessner JC. 2013. β-Adrenergic blockade during reactivation reduces the subjective feeling of remembering associated with emotional episodic memories. *Biol Psychol* **92:** 227–232.
Sederberg PB, Gershman SJ, Polyn SM, Norman KA. 2011. Human memory reconsolidation can be explained using the temporal context model. *Psychon Bull Rev* **18:** 455–468.
Sevenster D, Beckers T, Kindt M. 2014. Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learn Mem* **21:** 580–584.
Sinclair AS, Barense MD. 2017. Surprise and destabilize: prediction error influences episodic memory reconsolidation dataset and stimuli. doi: 10.17605/OSF.IO/GRNJW
Suzuki A, Josselyn SA, Frankland PW, Masushige S, Silva AJ, Kida S. 2004. Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *J Neurosci* **24:** 4787–4795.
Terman M, Rifkin J, Jacobs J, White T. 2001. *Automated morningness-eveningness questionnaire*. Center for Environmental Therapeutics, New York.
Walker MP, Brakefield T, Hobson JA, Stickgold R. 2003. Dissociable stages of human memory consolidation and reconsolidation. *Nature* **425:** 616–620.

Received October 23, 2017; accepted in revised form June 15, 2018.

Surprise and destabilize: prediction error influences episodic memory reconsolidation

Alyssa H. Sinclair and Morgan D. Barense

Learn. Mem. 2018, **25:**
Access the most recent version at doi:10.1101/lm.046912.117