

# Forgiveness Updates Interpersonal Memories to Be Less Negative

Songzhi Wu<sup>1</sup>, Timothy W. Broom<sup>2</sup>, Sasha Brietzke<sup>2</sup>, Jonathan Phillips<sup>1</sup>, Kevin N. Ochsner<sup>2</sup>,  
Lila Davachi<sup>2</sup>, and Meghan L. Meyer<sup>2</sup>

<sup>1</sup>Department of Psychological and Brain Sciences, Dartmouth College

<sup>2</sup>Department of Psychology, Columbia University

Forgiveness is crucial for restoring social bonds, yet how it shapes impressions of poor treatment remains unclear. Building on memory updating research, we propose forgiveness can change the memory of a negative experience by incorporating information considered during the forgiveness process. On Day 1 of neuroimaging, participants ( $N = 23$ , data collected 2022–2023) observed which stimuli two other participants (or “targets”) chose for them to view, believing the targets selected from neutral and negative images (encoding phase). Most chosen images were highly negative, indicating the target had treated the participant poorly. Participants then learned each target’s reasoning, with one being apologetic and the other nonchalant. While still undergoing neuroimaging, participants rated the negative images again while instructed to either “forgive the target” or simply “view the selections again” for the apologetic and nonchalant targets, respectively (experimental manipulation phase). On Day 2 of neuroimaging, participants rerated the images (reconsideration phase). Forgiveness reduced the negativity ratings of the images, an effect that persisted into Day 2. Two brain regions demonstrated that information considered while forgiving was incorporated into the memory of a forgiven act: the dorsomedial prefrontal cortex, associated with mentalizing, and the posterior hippocampus, linked to episodic memory. These findings suggest at least one way forgiveness works is by understanding the transgressor, updating related details, and consolidating them into memory. Instead of “forgive and forget,” forgiveness may involve a “forgive and update” process, revising memories to aid reconciliation.

*Keywords:* forgiveness, social cognition, mentalizing, memory, affect

*Supplemental materials:* <https://doi.org/10.1037/emo0001611.supp>

Forgiveness is key to restoring social connection after a transgression (e.g., Bono et al., 2008; vanOyen Witvliet & Luna, 2018). Sayings like “forgive and forget” suggest forgiveness may work by making the upsetting aspects of an interpersonal experience less memorable. Consistent with this idea, psychological research shows forgiveness can reduce memory for mistreatment-relevant content (Lichtenfeld et al., 2015; Noreen et al., 2014). Yet, despite the importance of forgiveness in sustaining social bonds, how forgiveness shapes memory for negative interpersonal experiences remains largely unclear.

Forgiveness is defined as the voluntary transformation of feelings, attitudes, and behavior such that a person is no longer upset with someone who wronged them (American Psychological Association,

2018). Here, we propose that forgiveness can update the representation of a negative experience by incorporating information considered while forgiving into the memory of the act being forgiven. That is, information considered during forgiveness may get included into memories of forgiven transgressions. This would mean that when we reconsider a previously forgiven event, the representation brought to mind should appear similar to the representation elicited during the moment of forgiveness. Specifically, the more effective the forgiveness (i.e., the less upsetting the memory when reconsidered later), the more the memory should look similar to the moment of forgiveness. Put simply, rather than “forgive and forget” the event, we may “forgive and update” our representation of the event.

This article was published Online First January 8, 2026.

Mauricio R. Delgado served as action editor.

Meghan L. Meyer  <https://orcid.org/0000-0002-9834-5925>

Data and analysis code are available on the Open Science Framework at <https://osf.io/bujms> (Wu & Meyer, 2025). This research was supported by the National Institute of Mental Health and Neurosciences, National Institutes of Health (Grant MH125406 awarded to Meghan L. Meyer).

This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0; <https://creativecommons.org/licenses/by-nc-nd/4.0>). This license permits copying and redistributing the work in any medium or format for non-commercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Songzhi Wu played a lead role in data curation and formal analysis, a supporting role in writing–review and editing, and an equal role in conceptualization and writing–original draft. Timothy W. Broom played a supporting role in formal analysis. Sasha Brietzke played a supporting role in formal analysis. Jonathan Phillips played a supporting role in conceptualization and writing–review and editing. Kevin N. Ochsner played a supporting role in writing–review and editing. Lila Davachi played a supporting role in writing–review and editing. Meghan L. Meyer played a lead role in writing–review and editing and an equal role in conceptualization.

Correspondence concerning this article should be addressed to Meghan L. Meyer, Department of Psychology, Columbia University, 1190 Amsterdam Avenue, Room 406, Schermerhorn Hall, New York, NY 10027, United States. Email: [mlm2378@columbia.edu](mailto:mlm2378@columbia.edu)

This proposal is supported by a range of learning and memory research showing that the initial encoding and subsequent consolidation of memories is a flexible process (Cowan et al., 2021; FeldmanHall et al., 2021; Tomparry & Davachi, 2024). The consolidation process, in which newly encoded information gets committed to memory, takes time, and new information can update an initial representation prior to long-term storage (Sekeres et al., 2017; Shohamy & Wagner, 2008). For example, when participants consecutively encode related information by learning a piece of information and then a new piece of related information, the new piece of information gets incorporated into the memory of the initial event (Sekeres et al., 2017; Tomparry & Davachi, 2024). Emotional arousal also updates memory consolidation, and this effect emerges slowly, after a 24-hr delay (Park, 2005). This area of work further suggests the impact of forgiveness on memory may emerge up to a day later, given that interpersonal interactions that warrant forgiveness are by their nature emotionally arousing. Collectively, multiple findings from the learning and memory literature suggest that if a transgressor is forgiven after wrongdoing, the information from the moment of forgiveness may get incorporated into the memory of the wrongdoing. If this were the case, when the wrongdoing is reconsidered later, particularly after some delay, the memory of the wrongdoing should include representations considered during forgiveness.

In the memory literature, this is often demonstrated with evidence that when recalling a memory, the information that was incorporated into it during encoding is reinstated (i.e., “reappears” when the memory is recalled; for a review, see Danker & Anderson, 2010). Greater reinstatement of encoding neural patterns during recall also scales with participants’ subjective ratings of stimuli made during recall. For example, more strongly reinstating a neural pattern from encoding during memory retrieval corresponds with subjectively perceiving the stimulus as more familiar (Johnson et al., 2009). This is consistent with our prediction that neural patterns for memories subjectively perceived as the least upsetting after forgiveness should look the most like the neural pattern reflecting information considered during forgiveness. However, to our knowledge, the memory updating literature has not investigated whether the information considered during forgiveness can become incorporated into the memory of a transgression. The primary goal of the present study is to test whether effective forgiveness is characterized by reinstating neural representations of forgiveness when the transgression is later reconsidered.

## What Information Considered During Forgiveness Becomes Incorporated Into a Memory?

### *Possibility 1: Forgiveness Incorporates Changed Affect Into the Memory*

If forgiveness works, at least in part, by incorporating information considered during forgiveness into memories of poor treatment, what information might this be? Forgiveness is complex and likely draws on multiple kinds of information, such as how one feels about the wrongdoing and how to regulate those feelings while also trying to consider the wrongdoer’s intentions. It is therefore worth considering which pieces of information involved in forgiveness may demonstrate memory updating. In the next sections, we consider what information may become incorporated into forgiven memories and how that may be reflected by neural patterns in distinct brain regions.

By definition, forgiveness assumes a reduction in the negative affect associated with poor treatment by a wrongdoer. This new, less negative affective response may get incorporated into the memory trace of the wrongdoing. If this were the case, then we would expect to see that forgiven experiences, when reconsidered later, show neural pattern similarity to the moment of forgiveness in the amygdala. The amygdala signals the presence of goal-relevant and potentially threatening stimuli and is particularly responsive to negative stimuli, including those used in traditional memory malleability research (Agren et al., 2012; Hamann et al., 1999; Kensinger & Corkin, 2003; Zald, 2003). Indeed, prior affective neuroscience research in animals and humans implicates the amygdala in the consolidation (and reconsolidation) of negative memories (Lee et al., 2017; McGaugh, 2002; Vazdarjanova & McGaugh, 1999). If the reduced negative affect becomes incorporated into memory in the amygdala, then in this scenario, the updated feelings are the information incorporated into the memory for the event and recalled when the event is reconsidered later.

While to date this possibility has not been examined, forgiveness scholars suggest that forgiveness creates lasting reductions in negative affect. One philosophical perspective of forgiveness is the “emotional fading model,” which suggests that forgiven experiences are characterized by less upsetting memories of the initial transgression (Amaya, 2019). Recent empirical work supports this model, demonstrating that relative to a number of memory dimensions measured, forgiven memories comprise significantly less negative valence and arousal than unforgiven memories (Fernández-Miranda et al., 2025). Social psychological research also shows that the reduced negative affect that emerges with forgiveness increases over time (McCullough et al., 2003), further pointing to the possibility that this attenuation of negative affect may occur through a delayed, memory updating process. If the primary mental representation that gets incorporated into a forgiven memory is attenuated negative affect, then we should see evidence of this effect in the amygdala.

### *Possibility 2: Forgiveness Incorporates Inhibitory Responses Into the Memory*

The second possibility is related to the first possibility described above. That is, if forgiveness dampens negative affect, this dampening may draw on response inhibition mechanisms that help regulate affect. Forgiveness involves thinking and acting in opposition to one’s initial reaction to a negative event, and regulating initial reactions is broadly associated with the ventrolateral prefrontal cortex (VLPFC; Ridderinkhof et al., 2004). For instance, this region supports response inhibition during Stroop tasks (e.g., reading the word “green” despite seeing it in a purple font color) and stop-signal tasks (e.g., withholding a practiced response to a stimulus). Perhaps most relevant to forgiveness, cognitive reappraisal tasks, which instruct participants to regulate affect in response to emotionally evocative stimuli by reframing them, reliably engage the VLPFC, particularly the left VLPFC (Ochsner et al., 2012). The left VLPFC is also implicated in the cognitive control of memories (Badre & Wagner, 2007). This further points to its potential role in updating negative memories through forgiveness. To the extent forgiveness requires a person to reinterpret a negative interpersonal event in part to regulate their emotions, neural representations in the left VLPFC may be updated to include information from the reinterpretation. In this scenario, when a person recalls a forgiven event, they may still

understand the transgressor and remember details of the experience as they did prior to forgiveness, but they bring to mind the inhibitory response they experienced during forgiveness.

### ***Possibility 3: Forgiveness Incorporates Mentalizing Content Into the Memory***

A third way that forgiveness may update memories of negative experiences is through the neural processes that support mentalizing. Mentalizing, the process of inferring other people's thoughts, feelings, and personality traits, offers a potential route for integrating information from the act of forgiveness into the memory of the event. Several existing models of forgiveness emphasize the importance of mentalizing, specifically considering the wrongdoer's perspective (e.g., Billingsley & Losin, 2017; Enright & North, 1998; Fourie et al., 2020; McCullough et al., 1997). Studies suggest that taking the transgressor's perspective reduces the victim's personal distress and that empathy motivates increased caring for the offender as well as relationship-constructive behavior, thus facilitating interpersonal forgiveness (McCullough et al., 1997; Takaku, 2001).

While multiple brain regions are implicated in mentalizing, the dorsomedial prefrontal cortex (DMPFC) may be particularly key to updating negative interpersonal memories with information considered during forgiveness. This is because the DMPFC is both robustly associated with mentalizing (Denny et al., 2012; Lieberman et al., 2019; Tan et al., 2022), as well as the consolidation of social memories (Brietzke et al., 2024; Jimenez & Meyer, 2024), including memories of interpersonal exchanges (Collier & Meyer, 2020). Interestingly, DMPFC is also implicated in several aspects of emotion regulation, particularly when it involves considering the abstract meaning behind experiences and employing strategies in cognitive reappraisal (Banks et al., 2007; Kensinger & Ford, 2021; Ochsner & Gross, 2005; Silvers & Guassi Moreira, 2019; Wager et al., 2008). In this scenario, the act of forgiveness involves reconsidering the meaning, or intentions, behind a transgressor's actions. This reconsideration of the transgressor's intentions is incorporated into the memory for the event and brought to mind upon later reconsideration.

### ***Possibility 4: Forgiveness Incorporates New Episodic Details Into the Memory***

Finally, it is also possible that forgiving updates neural representations in the hippocampus, a region reliably associated with episodic memory. Memory updating research frequently finds evidence that the hippocampus is sensitive to memory alterations (e.g., Sekeres et al., 2018), including when learning new information about how well people treat us (FeldmanHall et al., 2021). Pertinent to the present study, Speer et al. (2021) showed that finding positive meaning in a negative memory updates its representation such that 1 day later, the multivariate neural pattern in the hippocampus is altered—appearing dissimilar from how it responded during initial recall. Our hypothesis is similar, though still distinct from this prior work. That is, we specifically predict that information considered during forgiveness should get incorporated into memories of forgiven transgressions. Hence, rather than predicting neural dissimilarity, we predict that when we recall a memory for a forgiven event, it should appear similar to the moment of forgiveness. This would be consistent with the idea that information considered during forgiveness is incorporated into the memory. Given the role of the

hippocampus in representing episodic details of an experience, in this scenario, the act of forgiveness may change which details of an event are emphasized. These details would be emphasized in the memory for the event and recalled when the event is later reconsidered.

## **The Present Study**

We ran a multisession functional magnetic resonance imaging (fMRI) study to test whether and how forgiveness updates the memory of a negative experience by incorporating the information considered while forgiving into the representation of the transgression. On the first day of the study, participants were informed that two other participants (i.e., “targets”) selected images for them to view in the scanner. During the encoding phase, participants rated their affect felt in response to each image paired with the relevant target. Afterward, they were given the targets' reasons for negative image selection. One of the targets gave an apologetic response, while the other one was unapologetic and nonchalant.

Participants then went through the experimental manipulation phase where they viewed and rated the negative images again and were instructed to either “forgive” or simply “view” the respective targets, taking into consideration reasons given earlier. On the second day of fMRI scanning, participants reconsidered the initially shown stimuli from the targets and rerated all of the images. We hypothesized that forgiveness would update memories of the negative images, with the multivariate neural patterns on Day 2 reconsideration appearing more similar to the forgiveness (vs. viewing) patterns—specifically for the images that were rated the least negatively on Day 2 (i.e., when forgiveness was effective). The multivariate pattern similarity approach used here is ideal to test this hypothesis because it (a) quantifies the fine-grained neural patterns reflecting specific mental representations and (b) compares pairs of neural patterns, here within subjects, to determine their degree of similarity (Johnson et al., 2009; Kriegeskorte et al., 2008). In other words, the approach is designed to assess the similarity between sets of mental representations in the brain. We tested our prediction in brain regions previously associated with affect (amygdala), response inhibition (IVLPFC), mentalizing (DMPFC), and/or episodic memory updating (hippocampus). We also explored whether any other brain regions outside of those we predicted demonstrated forgiveness-related effects.

## **Method**

### **Transparency and Openness**

We report how we determined our sample size, all data exclusions, and all manipulations, and the study follows Journal Article Reporting Standards (Appelbaum et al., 2018). Data and analysis code are available at <https://osf.io/bujms>. We are unable to share our stimuli because the International Affective Picture System (IAPS) images are maintained by a separate database and permission must be granted for access. Data analysis software packages used are cited in the relevant sections below. The study design and its analysis were not preregistered.

### **Participants**

We recruited 43 participants from Dartmouth College and the greater community between November 2022 and May 2023. Sample size was determined based on available funding. Six participants

failed to complete the experiment (i.e., did not complete Day 2 of scanning), leaving a total sample size of 37 (12 male, 21 female, four gender nonconforming;  $M_{\text{age}} = 23.11$  years,  $SD_{\text{age}} = 7.19$  years; 27.03% Asian, 8.11% Black/African, 45.95% Caucasian, 24.32% Hispanic/Latinx, 2.70% Native American, 2.70% Pacific Islander). Of these participants, 14 were suspicious that the “other participants” that chose their stimuli were real people (as opposed to artificially constructed for the purpose of the study; see the Day 2 fMRI Scan section under the Procedure section). Given that forgiveness necessitates a real person to forgive, all results are reported with the 23 participants who believed the experiment (nine male, 12 female, two gender nonconforming;  $M_{\text{age}} = 22.83$  years,  $SD_{\text{age}} = 6.16$  years; 26.09% Asian, 13.04% Black/African, 43.48% Caucasian, 26.09% Hispanic/Latinx, 4.35% Native American, 4.35% Pacific Islander), and follow-up analyses reported from the full sample (including suspicious participants) are reported in the [Supplemental Materials](#). Note that a sample size of 23, while not large, is similar to those used in other recent brain imaging studies investigating multivariate neural pattern similarity within subjects (Monsa et al., 2024; Yu et al., 2025). The full study protocol was approved by the Dartmouth College Institutional Review Board. All participants provided informed consent. Participants were awarded course credit or paid \$20 per hour for study completion.

## Procedure

### Cover Story and Procedure

To enhance paradigm believability, a few days prior to their scan, participants completed an “image selection task” in which they decided which images another supposed participant would see during their scan session and provided an explanation for their image selection choices. Throughout the experiment, including the said task, participants were shown images from the IAPS, which is maintained and distributed by researchers at the National Institute of Mental Health Center for the Study of Emotion and Attention at the University of Florida and includes images that were rated on a 9-point rating scale along two primary dimensions: valence (1 = *low pleasure*, 9 = *high pleasure*) and arousal (1 = *low arousal*, 9 = *high arousal*). More specifically, in this part, the images were 30 pairs of images, one negative and one neutral in each pair. The participants were told that

We are partnering with other universities in conducting this study. Now, you will select images that a participant from another university (e.g., Johns Hopkins) will view while they are in the fMRI scanner. Similarly, participants at other universities will be selecting which images you will eventually see in your fMRI sessions. After this screen, you will see a series of images and we would like you to select one from each pair for students at another university to view while they are in their fMRI scanner. When you are ready, click the arrow to proceed.

Subsequently, participants chose one out of two IAPS images across 30 trials and were asked to describe the strategy they used upon reflecting on their choices. Finally, we asked participants to submit an electronic copy of a headshot of theirs. They were told that their photograph would be shown with the images they selected for the future participant. Again, the primary purpose of this procedure was to enhance the believability of the experiment when participants took part in the subsequent fMRI sessions. It is noteworthy that the mean proportion of negative images selected by participants was 36.23%, with little variation (variance = 4.83%) and positive skew (i.e., a distribution biased

toward a low proportion of negative image selection). This suggests a norm to give other participants mostly neutral images (relative to negative images) and that participants were presumably not expecting the other supposed participants to give them a disproportionate amount of negative images. Indeed, when we examined the participants’ strategy, two themes emerged: Participants tended to note trying to avoid selecting aversive images for the other participant while also trying to select interesting images for the other participant.

### Day 1 fMRI Scan

On the day of the first scheduled fMRI scan, participants went through a safety check, provided informed consent, and watched an instruction video about the tasks to be completed in the scanner.

During fMRI, participants first completed the encoding phase, where they saw on each screen (a) a headshot of “another participant” (hereafter referred to as “target”) together with her or his name, (b) an IAPS image, and (c) a slider of 1–4 at the bottom (see [Figure 1](#)). With the buttonbox in their hands, participants indicated their rating of the image on each screen on a 1 (*negative*) to 4 (*neutral*) scale. Each participant saw a total of 90 IAPS images associated with two targets (45 images for each target): 2/3 of the images were negative in valence (2–4 in IAPS mean valence), and 1/3 were neutral in valence (4–5 in IAPS mean valence), as predefined by prior ratings (Lang et al., 2008). According to paired-sample *t* tests, there was no statistically significant difference in valence or arousal between the negative images—valence:  $t(29) = 0.59$ ,  $p = .559$ ; arousal:  $t(29) = 1.37$ ,  $p = .183$ —and the neutral images—valence:  $t(14) = 0.71$ ,  $p = .489$ ; arousal:  $t(14) = 0.03$ ,  $p = .978$ —assigned to the two targets. The reason why participants were shown mostly negative images is to help them form an overall negative impression about the targets initially so that we could investigate the effect of attempting to subsequently forgive them. Note that all analyses reported below pertain to the negative IAPS images. The neutral images are unrelated to our hypotheses and were included to enhance believability, as it seems unlikely that someone would only choose negative images.

Each image stayed on the screen for 6 s, with jittered fixation created by easy-optimize-x in between each image (Spunt, 2016; mean jitter = 3,  $SD$  jitter = 0.87). Targets were gender-matched to each participant. The entire encoding phase took place across six blocks over two runs that each lasted around 8 min. Block order was randomized across participants.

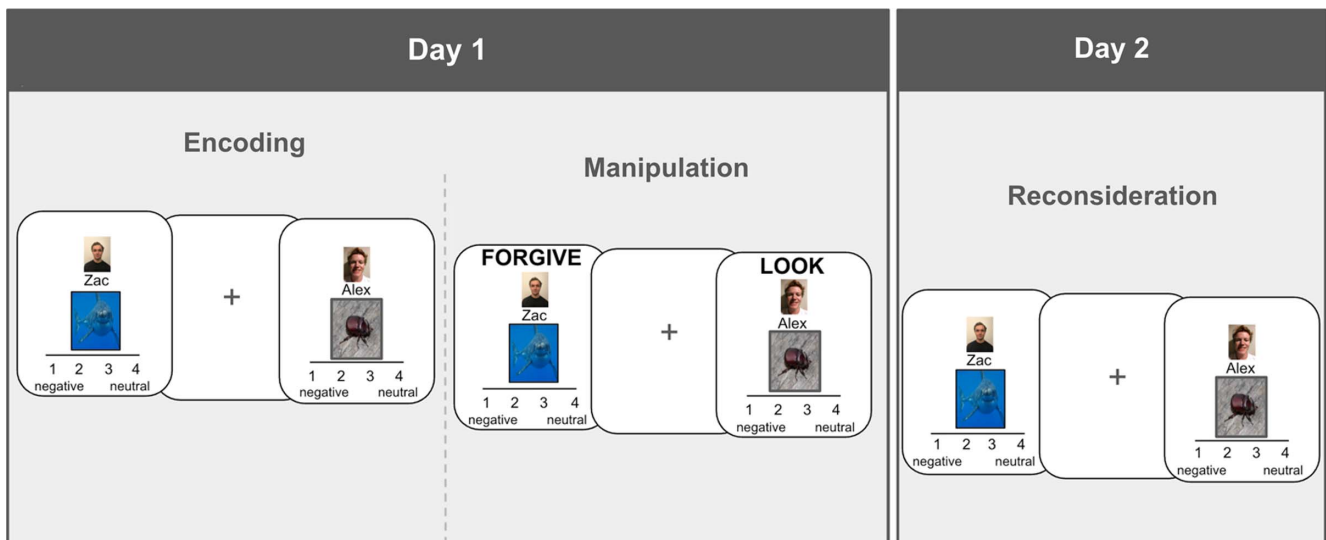
Once encoding was completed, participants would see on the screen the targets’ reasons for their image selections. It was then revealed to participants that one target said she or he actually had a rough day and would like to apologize to whoever had to view those negative images and the other responded that she or he randomly chose the images. More specifically, one target said:

To be honest, I was having a really hard day. ... My boss really harshly blamed me for some mistake that actually had nothing to do with me! I guess I took it out on whomever has to view those horrifying images. I really want to apologize to them. I’m very sorry and hope you can forgive me for doing that!

The other target’s explanation was:

Hmmm, I’m not sure exactly why I picked some images over other images. I could tell that some of the images were really very scary and creepy and others were actually pretty chill. But I’m not sure why I chose one or the other. Now that I think about it, I guess I just sort of went with my gut.

**Figure 1**  
Overview of the Experimental Design



*Note.* During the Day 1 encoding phase, participants made affective ratings for images paired with two separate targets in the functional magnetic resonance imaging scanner, after which they were given the targets' reasons for negative image selection. Then, during the Day 1 experimental manipulation phase, participants viewed and rated the negative images again and were instructed to either "forgive" or "view" the respective targets, considering reasons given earlier. During the Day 2 reconsideration phase, participants saw and rated all the images again in another functional magnetic resonance imaging scan. We received official consent from the individuals depicted in the photos to allow their images to be shown in this publication. These are the faces male participants saw while undergoing functional magnetic resonance imaging. Participants observed the International Affective Picture System images; however, the negative images shown here are approved for redistribution under creative commons licensing and can be found at [https://commons.wikimedia.org/wiki/File:Great\\_white\\_shark\\_south\\_africa.jpg](https://commons.wikimedia.org/wiki/File:Great_white_shark_south_africa.jpg) (Creative Commons Attribution 2.0 Generic license) and [https://commons.wikimedia.org/wiki/File:Oryctes\\_nasicornis\\_male\\_2012\\_G1.jpg](https://commons.wikimedia.org/wiki/File:Oryctes_nasicornis_male_2012_G1.jpg) (public domain). See the online article for the color version of this figure.

We used a nonchalant target, as opposed to an intentionally unapologetic target, as our control condition target for two reasons. First, this aligns with reappraisal paradigms that use a "look" condition without any additional context provided to participants with respect to why they are being asked to look at the image and rate its valence again. Second, we decided against using an unambiguously and/or intentionally unforgiving target because that would be another experimental, rather than control, condition. That is, if one target intentionally and unambiguously apologized and another target intentionally and unambiguously did not apologize, it would be hard to know if decreased negative affect and/or neural changes associated with the forgiveness condition were due to forgiveness versus increased negative affect/resistance to memory change in response to an actively unapologetic response. By analogy, our logic was similar to studies that choose not to directly compare a negative affect condition to a positive affect condition and instead start with a comparison of negative affect versus neutral affect condition. Moreover, prior work shows that transgressors who take responsibility for their wrongdoing in their apology are more easily forgiven than transgressors who do not (McCullough et al., 2003); this is thought to occur because when the apologizer takes responsibility for their initial intentions, it prompts the victim to actively engage with the forgiveness process, and when transgressors admit their original intention, their apology appears more sincere (McCullough et al., 1997). This prior work would suggest that the nonchalant target is more difficult to forgive than the target who apologizes, in part because they do not provide a clear intention behind their mostly negative image selections.

We informed participants that the negative images would be shown to them again, but this time, they would simply look at the images associated with the target who did not apologize (hereafter referred to as the "looked at target") and try to forgive the target who apologized for image choices (hereafter referred to as the "forgiven target"). The "look" condition is based on extensive prior research on cognitive reappraisal, which also uses a "look" condition as a control comparison to emotion regulation (Ochsner et al., 2004, 2012).

At this point, participants entered the "experimental manipulation" phase. Participants saw on each screen a target's face, a target's name, an IAPS image chosen by the target as already seen during encoding, and a slider to make their affective rating. They also saw an instruction cue of "view" or "forgive" next to the image as a reminder for them to merely look at or try to forgive the target (see Figure 1). The manipulation phase included only the negative images shown during encoding (because there was no reason to forgive a neutral image selection) and took place across four blocks over two runs that each lasted around 5 min.

Once outside of the scanner, participants filled out a survey where they were asked to (a) rate how they felt about each target as a person on a scale of 1–10 (1 = *extremely negative*, 10 = *extremely positive*) and (b), in 15 trials per target (30 in total), select one out of two images for the targets to view when we had them come back for a new version of the study. The purpose of the second image selection task was to probe participants' motivation for "revenge" toward the two targets (i.e., a greater ratio of negative-to-neutral image selections reflects giving negative treatment to the targets in response to receiving negative treatment from them). To compute an index of revenge, we calculated the average negative rating of the 15 selected

trials such that lower overall average ratings indicated picking more upsetting images for the target. These additional tasks were carried out because prior work suggests that forgiveness corresponds with more positive feelings toward a transgressor and less motivation for revenge (McCullough et al., 2014; Tabak et al., 2012). These additional outcomes helped us assess the effectiveness of our experimental task in capturing forgiveness. That is, if the affective ratings during the forgiveness manipulation relate to these other outcomes implicated in forgiveness, it would suggest the paradigm effectively captures forgiveness.

### Day 2 fMRI Scan

On the following day, participants came in for their second fMRI scan. As before, they were checked for MRI safety and provided informed consent. In the scanner, participants went through the reconsideration phase where they repeated their task during encoding; namely, they rated the negativity of each image the targets gave them on Day 1.

We wanted to ensure participants believed the targets were real people (as opposed to artificially constructed by the experimenters); otherwise, they would be unmotivated and/or unable to authentically forgive them. To address this, outside of the scanner on Day 2, we probed the extent to which participants believed the experimental paradigm and their thoughts about the study's goals. Specifically, they were asked the following: (a) Is there anything about this study that seems off to you? (b) What do you think this study was about? (c) On a scale of 1–6, to what extent did you believe that our study paradigm was veracious (1 = *completely did not believe*, 6 = *completely believed*)? Participants who mentioned suspicion about the veracity of the targets in their free response and/or reported a believability score less than 3 (on a scale of 1–6;  $n = 14$ ) were excluded from analyses reported here. This exclusion criterion was determined prior to examining the data. For the interested reader, the [Supplemental Materials](#) show results from the full sample. It is noteworthy that all statistically significant results reported below remain statistically significant when the full sample is assessed except for the right DMPFC (rDMPFC) results, which become marginally significant in the Neurosynth “mentalizing” region of interest (ROI) and nonsignificant but in the same direction in the Shen rDMPFC ROI (see the [Supplemental Materials](#)). At the end of the study, participants were debriefed and compensated for their time.

### fMRI Data Acquisition

fMRI data were collected on a Siemens Prisma 3T scanner across 2 days. On the first day, four functional runs—two encoding runs and two manipulation runs—were acquired using an echo planar imaging gradient-echo sequence ( $2.5 \times 2.5 \times 2.5$  mm voxels, repetition time = 1,000 ms, excitation time = 30 ms, 2.5 mm slice thickness, field of view = 24 cm, matrix =  $96 \times 96$ , flip angle = 59, simultaneous multislice = 4). A T2-weighted structural image was acquired coplanar with the functional images ( $0.9 \times 0.9 \times 0.9$  mm voxels, repetition time = 2,300 ms, excitation time = 2.32 ms, 0.9 mm slice thickness, field of view = 24 cm, matrix =  $256 \times 256$ , flip angle = 8). Sequence optimization was derived using optimize design software (<https://www.bobspunt.com/easy-optimize-x/>). Each encoding run was 8 min in duration; each manipulation run was 5.4 min in duration. On the second day during reconsideration,

two functional reconsideration runs were conducted with the same setting as those on Day 1. Each reconsideration run was 8 min in duration.

### fMRI Preprocessing

Results included in this article come from preprocessing performed using *fMRIPrep* 23.1.3 (Esteban et al., 2018, 2019; RRID: SCR\_016216), which is based on *Nipype* 1.8.6 (Esteban et al., 2025; Gorgolewski et al., 2011; RRID:SCR\_002502). As recommended by Esteban et al. (2019), for transparency and reproducibility, we provide fMRIPrep's boilerplate text.

For each of the blood oxygen level dependent (BOLD) runs per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using *mcfliirt* (FMRIB Software Library; Jenkinson et al., 2002). The estimated field map was then aligned with rigid registration to the target echo planar imaging reference run. The field coefficients were mapped onto the reference echo planar imaging using the transform. BOLD runs were slice-time corrected to 0.452 s (0.5 of slice acquisition range = 0–0.905 s) using 3dTshift from Analysis of Functional NeuroImages (Cox & Hyde, 1997; RRID:SCR\_005927). The BOLD reference was then coregistered to the T1w reference using *mri\_coreg* (FreeSurfer) followed by *flirt* (FMRIB Software Library, Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009) cost function. Coregistration was configured with six degrees of freedom. Several confounding time series were calculated based on the preprocessed BOLD: framewise displacement (FD), derivative of root-mean-square variance over voxels, and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions; Power et al., 2014) and Jenkinson (relative root-mean-square displacement between affines; Jenkinson et al., 2002). FD and derivative of root-mean-square variance over voxels are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014).

The three global signals are extracted within the cerebral spinal fluid (CSF), the white matter (WM), and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor; Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time series (using a discrete cosine filter with 128 s cutoff) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM, and combined CSF + WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. (2007) in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of grey matter is subtracted from the aCompCor masks. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of grey matter. Finally, these masks are resampled into BOLD space and binarized by

thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the  $k$  components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50% of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head-motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized derivative of root-mean-square variance over voxels were annotated as motion outliers. Additional nuisance time series are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by Patriat et al. (2017).

The BOLD time series were resampled into standard space, generating a preprocessed BOLD run in MNI152Nlin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and coregistrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms, configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Nongridded (surface) resamplings were performed using mri\_vol2surf (FreeSurfer).

## ROIs

We used the meta-analytic platform Neurosynth (Yarkoni et al., 2011) to derive functional ROIs in the amygdala, left VLPFC, and DMPFC related to two of the constructs relevant to our predictions. Because emotional changes to affect-inducing stimuli through response inhibition, including cognitive reappraisal, are typically studied under the construct "emotion regulation" (Buhle et al., 2014), we used the search term "emotion regulation" to generate functional ROIs for the amygdala and IVLPFC. Indeed, this search term identified clusters in the bilateral amygdala (which we broke down into separate left and right ROIs) as well as the left VLPFC. The search term "mentalizing" yielded bilateral clusters in DMPFC, which were also broken into a left DMPFC and right DMPFC ROI.

To investigate the possibility that the hippocampus may be involved in memory updating through forgiveness, we used the six hippocampus ROIs from the 268 Shen parcellation (Parcels 93–94 and 229–232; Shen et al., 2013). Note that we did not use Neurosynth-derived hippocampus ROIs because relevant search terms such as "memory" or "episodic memory" yield the entire hippocampus, though different subportions of the hippocampus are thought to support different aspects of memory (Bein et al., 2020; Thorp et al., 2022; Treves & Rolls, 1994). For this reason, the Neurosynth-derived ROI was deemed too broad, and instead, the Shen parcellation allowed for us to assess functionally relevant subdivisions of the hippocampus. We chose the Shen parcellation for this because our follow-up, exploratory whole-brain analyses also used the Shen parcellation. In this way, we consistently used the same functional divisions of the brain throughout analyses.

## General Linear Models

General linear models (GLMs) were constructed using nltools (<https://nltools.org>). GLMs incorporated nuisance regressors (the six standard motion parameters and their derivatives, the signal extracted from WM regions, the signal extracted from cerebrospinal fluid regions, and first- and second-order polynomial trends) and were convolved with a canonical hemodynamic response function to compute parameter estimates ( $\beta$ ) at each voxel. GLMs were run separately for each phase (i.e., encoding, manipulation, and reconsideration phases). For univariate analyses, data were spatially smoothed using an 8-mm full-width-at-half-maximum Gaussian kernel. For these analyses, conditions were defined at the "category" level; for example, all negative images viewed from the forgiven target were defined as a single condition in the GLM. For multivariate analyses, data were spatially smoothed using a 4-mm full-width-at-half-maximum Gaussian kernel following prior multivariate analysis work (e.g., Broom & Wagner, 2023; Chavez & Wagner, 2020; Guthrie et al., 2022). For these analyses, each individual trial was defined as an independent condition, allowing for estimation of the multivariate neural representation of each specific image viewed.

## Univariate Comparisons Between Experimental Phases

We assessed the extent to which each of our ROIs was engaged during each experimental phase by examining their univariate activation in response to negative images for the forgiven versus looked at target. Each condition was modeled relative to the implicit baseline. Parameter estimates for each condition were extracted for each subject. We then ran 2 (target: forgiven vs. looked at)  $\times$  3 (experimental phase: encoding, manipulation, reconsideration) repeated-measures analyses of variance (ANOVAs) on the negative images to determine which ROIs' univariate activity varied between phases as a function of the target. For these targeted ROI analyses, statistical significance was determined by an false discovery rate (FDR)-corrected  $p$  value of  $p < .05$ , which took into account the number of ROIs assessed. We complemented these targeted ROI analyses with parallel exploratory, whole-brain repeated-measures ANOVAs across all parcels from the 268 Shen parcellation testing for any parcels (FDR-corrected  $p < .05$ ) that may also vary neural responding as a function of the phase and target.

## Neural Pattern Similarity Between Experimental Phases

We used multivariate neural pattern similarity to test our primary hypothesis that information from forgiveness becomes incorporated into forgiven experiences to make them less negative. For each negative trial and for each participant, we computed the similarity (i.e., Pearson's correlation coefficient) between multivariate neural responses for (a) similarity between encoding and manipulation, (b) similarity between manipulation and reconsideration, and (c) similarity between encoding and reconsideration. It is noteworthy that we statistically compared these correlation values, rather than Fisher  $z$  transformed values. This is because both visual inspection and a Kolmogorov–Smirnov test used to determine whether data significantly deviated from normality showed that our  $r$  values were more normally distributed ( $D = 0.003$ ,  $p = .408$ ) than the Fisher  $z$  transformed values ( $D = 0.023$ ,  $p < .001$ ). We also computed, for each negative trial and for each participant, the change in the affective

rating. Specifically, we calculated the numerical difference between [encoding – manipulation], [reconsideration – manipulation], and [reconsideration – encoding]. Higher affective difference values indicate less negativity between Timepoint 1 and Timepoint 2. For example, a value of zero between manipulation and reconsideration would indicate a trial that stayed as negative during reconsideration as during manipulation; a positive value would indicate the image became less negative during reconsideration relative to during manipulation; a negative value would indicate the image became more negative during reconsideration versus manipulation.

For each ROI, we performed an item-level analysis assessing changes in neural pattern similarity as a function of change in negative affect rating for the forgiven versus looked at target. Although we computed pattern similarity between each phase in the experiment, the primary focus was on changes in neural pattern similarity between the manipulation and reconsideration phases. This is based on prior memory updating and forgiveness research, which both suggest that the new information from the moment of forgiveness should become incorporated into the memory representation after some delay (Cowan et al., 2021; McCullough et al., 2003; Sekeres et al., 2017). This allowed us to assess whether memories that are less negative as a function of forgiveness upon Day 2 reconsideration preferentially show increased neural pattern similarity to the forgiveness brain state measured on Day 1. In other words, memories that are less upsetting on Day 2 have the greatest trace of the information considered during forgiveness incorporated into them. Statistical significance of these ROI analyses was determined with an FDR-corrected  $p$  value of  $p < .05$ , accounting for the number of ROIs assessed.

We complemented our targeted ROI analyses with exploratory analyses across all parcels from the 268 Shen parcellation testing for any parcels (FDR-corrected  $p < .05$ ) that were significant in the abovementioned comparisons. Pairwise neural similarity was computed by correlating the patterns of neural activity for every trial pair (e.g., Trial 1 during manipulation and Trial 1 during reconsideration). To account for the nested nature of the data (i.e., the pictures evaluated were nested within participants), linear mixed-effects modeling was implemented in the R statistical language (R Core Team, 2022) using the lmerTest package (Kuznetsova et al., 2017) with a random intercept for each participant and a random intercept for each IAPS image. The same package was used to calculate Satterthwaite approximated degrees of freedom and corresponding  $p$  values for all linear mixed-effects models reported. Confidence intervals for each fixed-effect parameter were estimated using the confint function of the lme4 package (1,000 simulations; Bates et al., 2015).

We next ran follow-up, nonparametric permutation tests to ensure the robustness of our observed neural pattern similarity results, especially since nonparametric permutation tests are well-suited for small sample sizes. Here, we used nonparametric permutation testing to assess whether our observed neural pattern similarity results are meaningful, above and beyond what would be expected from noise and without reliance on parametric testing assumptions. We conducted a permutation test ( $n = 1,000$  permutations) to assess the interaction between affective rating differences and target condition (forgive vs. look) on neural pattern similarity in each of our ROIs showing the parametric results (DMPFC ROIs and posterior hippocampus ROI). Specifically, we fit a linear mixed-effects model for each ROI predicting neural pattern similarity using the interaction term, while including a random intercept for distinct participants and

a random intercept for distinct image trials to account for repeated measures. The fixed-effect coefficient for the interaction term was extracted from each observed model. Then, to generate a null distribution of these interaction effects, we randomly permuted target condition labels within participants, preserving the within-subject structure, and refit the mixed-effects model for each permutation. With the null distribution, we calculated two-tailed  $p$  values by comparing the observed interaction with the null distribution.

## Results

### Behavioral Results

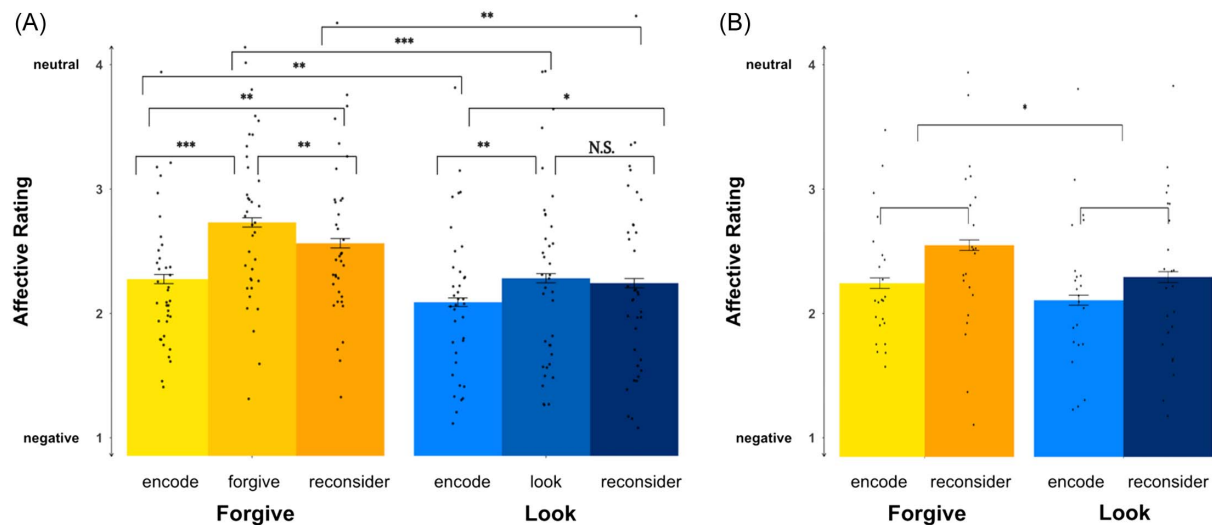
#### *Changes in Negative Affect Across Experimental Phases and Conditions*

The affective ratings made during fMRI scans indicate that the forgiveness condition was more effective than the look condition in attenuating negative affect. A 2 (target: forgiven vs. looked at)  $\times$  3 (experimental phase: encoding, manipulation, reconsideration) repeated-measures ANOVA showed that the interaction for the negative image ratings was significant,  $F(2, 44) = 4.99$ ,  $p = .011$ ,  $\eta_g^2 = 0.01$  (Figure 2A). Within the forgiven target condition, we found that forgiving during Day 1 manipulation reduced negativity ratings relative to encoding,  $t(22) = 5.11$ ,  $p < .001$ . Moreover, the ratings of negative images on Day 2 reconsideration were still less negative than during encoding on Day 1,  $t(22) = 3.37$ ,  $p = .003$ . That said, reconsideration ratings were also more negative than forgiveness manipulation ratings,  $t(22) = -2.84$ ,  $p = .010$ , although there was variability, within subjects, in affect ratings in response to negative images on Day 2 across the negative images (average variance in negative image affect ratings within subjects on Day 2 = 0.81; average standard deviation within subjects in negative image affect ratings on Day 2 = 0.88). This pattern suggests two things about the affect felt upon reconsideration of the negative images on Day 2: (a) The negativity of the memory remains attenuated on Day 2 relative to initial encoding on Day 1, and (b) forgiveness created lasting and even enhanced reductions to negative affect for some but not all trials.

Ratings for the looked at target were also less negative during the look manipulation versus encoding on Day 1,  $t(22) = 3.09$ ,  $p = .005$ , and less negative on Day 2 reconsideration versus Day 1 encoding,  $t(22) = 2.29$ ,  $p = .032$ , but showed no difference in negativity on Day 2 reconsideration versus Day 1 manipulation,  $t(22) = -1.16$ ,  $p = .259$ .

We next ran follow-up analyses to assess whether the significant results observed for the forgiveness conditions persisted above and beyond effects observed for the look conditions (as would be implied by the significant 2  $\times$  3 interaction reported above). It is noteworthy that on Day 1, negative trials paired with the forgiven target were rated as less negative than those paired with the looked at target,  $t(22) = 3.40$ ,  $p = .003$ . We therefore tested whether differences in affect ratings for the forgiven versus looked at target emerge during the other phases of the experiment when controlling for each participant's baseline difference in ratings during encoding. The negative trials shown with the forgiven target were rated as less negative than the trials shown with the looked at target during manipulation, while controlling for baseline differences during encoding ( $\beta = -0.23$ ),  $t(2653.09) = -4.02$ ,  $p < .001$ , 95% CI  $[-0.35, -0.12]$ . The reduction in negative affect from encoding to reconsideration was also greater for the forgiven (vs. looked at) target, while controlling for baseline differences in affect ratings at

**Figure 2**  
*Behavioral Results*



*Note.* (A) Data shown in the plots come from the negative image trials; participants rated their affect in response to each negative image during each phase of the experiment. Because the images shown during encoding and reconsideration ranged from negative to neutral, participants made their ratings on a 1 (*negative*) to 4 (*neutral*) scale. (B) For negative trials, the reduction in negative affect from encoding to reconsideration was greater for the forgiven versus the looked at target, controlling for baseline differences in affect ratings at encoding. Note that the data shown in Panel B are also shown in Panel A; they are shown again in Panel B to demonstrate the difference in affective rating change between the two target conditions. N.S. = nonsignificant. See the online article for the color version of this figure.  
\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

encoding ( $\beta = -0.12$ ,  $t(2667.01) = -2.05$ ,  $p = .041$ , 95% CI  $[-0.23, -0.01]$  (Figure 2B).

Finally, we explored (a) whether the negative rating at encoding interacts with the target (forgive vs. look) to predict the affect rating change during manipulation (vs. encoding) and (b) whether the negative rating at encoding interacts with the target (forgive vs. look) to predict the affect rating change during reconsideration (vs. encoding). Both interactions are significant—manipulation versus encoding model:  $\beta = 0.10$ ,  $t(1239.67) = 2.54$ ,  $p = .011$ , 95% CI  $[0.02, 0.17]$ ; reconsideration versus encoding model:  $\beta = 0.12$ ,  $t(1336.34) = 3.22$ ,  $p = .001$ , 95% CI  $[0.05, 0.20]$ —and are driven by stronger relationships in the forgive versus look condition—manipulation versus encoding: forgive,  $\beta = -0.56$ ,  $t(1280) = -18.93$ ,  $p < .001$ , 95% CI  $[-0.62, -0.50]$ ; look,  $\beta = -0.44$ ,  $t(1359) = -14.71$ ,  $p < .001$ , 95% CI  $[-0.50, -0.38]$ ; reconsideration versus encoding: forgive,  $\beta = -0.48$ ,  $p < .001$ , 95% CI  $[-0.54, -0.42]$ ; look,  $\beta = -0.38$ ,  $p < .001$ , 95% CI  $[-0.44, -0.32]$ . This suggests that forgiveness was most effective for the trials that were the most subjectively upsetting at baseline, perhaps because they had the most “room” on the 1–4 scale to become reduced or perhaps because forgiveness is more easily induced for more versus less poor treatment.

More generally, the affective rating results suggest that while there was some habituation to the negative images during the control condition (i.e., “look”) manipulation and reconsideration phases, forgiveness attenuated the negative experience above and beyond what would be expected from habituation, as well as above and beyond baseline differences in affect during encoding. In other words, in terms of subjective experience, forgiveness updates negative interpersonal memories to be less upsetting.

### *Linking Participants’ Image Selections at Baseline, Revenge, and Target Liking With Changes in Negative Affect Across Experimental Phases and Conditions*

We next explored whether the additional behavioral measures we collected—participants’ baseline image selection, feelings of revenge, and liking of the forgive and look target—related to participants’ changes in negative affect across experimental phases and conditions. The extent to which a participant selected negative images “at baseline” for another target before undergoing the experiment corresponded with their revenge score for the looked at target ( $r = .413$ ,  $p = .050$ ) and was in the same overall direction (but not significant) for the forgiven target ( $r = .355$ ,  $p = .096$ ). It is notable that, for both targets, the relationship is positive (i.e., the “meaner” a participant in their selections for another, random participant prior to knowing what kinds of images someone else was going to choose for them corresponds with later being more revengeful toward the transgressors), but we refrain from interpreting these results because they are not statistically significant.

Turning to relationships between affect rating change across conditions, we found evidence that baseline meanness and liking, but less so revenge, meaningfully related to forgiveness effectiveness. We tested for separate interaction models assessing whether (a) baseline meanness and target, (b) liking and target, and/or (c) revenge and target predict reductions in negative affect between manipulation versus encoding and, separately, between reconsideration and encoding.

Baseline “meanness” interacted with the target (forgive vs. look) to predict the reduction in affect ratings on the task between the manipulation versus encoding phase ( $\beta = 0.38$ ),  $t(1274.85) = 3.25$ ,  $p = .001$ , 95% CI  $[0.10, 0.60]$ , as well as the reconsideration versus

encoding phase ( $\beta = 0.46$ ),  $t(1288.16) = 3.249$ ,  $p = .001$ , 95% CI [0.24, 0.70]. For reductions in negative affect between manipulation and encoding, baseline meanness corresponded with a trend toward less effective forgiveness—manipulation versus encoding forgiven target:  $\beta = -0.45$ ,  $t(23.8) = -1.95$ ,  $p = .063$ , 95% CI [-0.94, 0.03]—and this effect is attenuated for the looked at target ( $\beta = -0.08$ ),  $t(23.9) = -0.33$ ,  $p = .743$ , 95% CI [-0.56, 0.40]. For reductions in negative affect between reconsideration and encoding, baseline meanness corresponded with significantly less effective forgiveness ( $\beta = -0.59$ ),  $t(24.7) = -2.79$ ,  $p = .010$ , 95% CI [-1.03, -0.15], and this effect is attenuated for the looked at target ( $\beta = -0.14$ ),  $t(24.7) = -0.64$ ,  $p = .528$ , 95% CI [-0.57, 0.30]. The baseline image selection task, while designed to increase paradigm believability, may have also tapped into individual differences in prosociality related to forgiveness, particularly the forgiveness that emerges over a delay (i.e., reconsideration phase), which is consistent with both the memory updating and forgiveness literatures (Cowan et al., 2021; McCullough et al., 2003; Sekeres et al., 2017).

Revengefulness did not interact with the target to predict reductions in negative affect rating scores between manipulation versus encoding ( $\beta = 0.17$ ),  $t(1283.56) = 1.22$ ,  $p = .224$ , 95% CI [-0.14, 0.48], nor between reconsideration versus encoding ( $\beta = -0.05$ ),  $t(1288.70) = -0.32$ ,  $p = .751$ , 95% CI [-0.33, 0.24]. That said, participants also rated how much they liked each target at the end of Day 1 participation, and these liking scores did meaningfully interact with the target to predict reductions in negative affect. While greater reductions in negative affect between the manipulation and encoding phase were associated with greater liking of both targets ( $\beta = 0.10$ ),  $t(487.50) = 5.76$ ,  $p < .001$ , 95% CI [0.07, 0.14]; forgive target ( $\beta = 0.10$ ),  $t(497) = 5.69$ ,  $p < .001$ , 95% CI [0.07, 0.14]; and look target ( $\beta = .06$ ),  $t(497) = 3.28$ ,  $p = .001$ , 95% CI [0.026, 0.10], the relationship is greater for the forgiven target as reflected by the significant interaction ( $\beta = -0.041$ ),  $t(1267.46) = -1.98$ ,  $p = .048$ , 95% CI [-0.08, 0.00]. Liking and target also significantly interacted to predict reductions in negative affect between reconsideration and encoding ( $\beta = -.06$ ),  $t(1299.41) = -2.81$ ,  $p = .005$ , 95% CI [-0.10, -0.02], such that participants who liked the forgiven target the most by the end of Day 1 also showed the greatest reductions in negative affect between reconsideration and encoding ( $\beta = -0.06$ ),  $t(630) = -3.07$ ,  $p = .002$ , 95% CI [0.02, 0.10], and this relationship is attenuated for the looked at target ( $\beta = 0.0002$ ),  $t(632) = -0.01$ ,  $p = .992$ , 95% CI [-0.04, 0.04]. Overall, these exploratory analyses suggest that the forgiveness manipulation was effective, both because it meaningfully related to baseline individual differences in prosociality and how much participants liked the target they forgave.

## Neural Results

### *Univariate Neural Activity for Forgiven Versus Looked at Targets*

Our primary hypotheses pertain to whether forgiveness content becomes incorporated into forgiven experiences to make them less upsetting. However, for the sake of completion, we additionally assessed how strongly each ROI was engaged (i.e., univariate activation) for each condition during each experimental phase. None of the targeted ROIs demonstrated a significant 2 (target: forgiven vs. looked at)  $\times$  3 (experimental phase: encoding, manipulation, reconsideration) repeated-measures ANOVA interactions ( $F_s < 3.31$ , FDR-corrected

$ps > .429$ ). The exploratory, whole-brain analyses across all parcels from the 268 Shen parcellation also yielded null results ( $F_s < 4.08$ , FDR-corrected  $ps > .976$ ). Thus, the overall degree to which participants engaged neural regions did not vary as a function of the target and experimental phase.

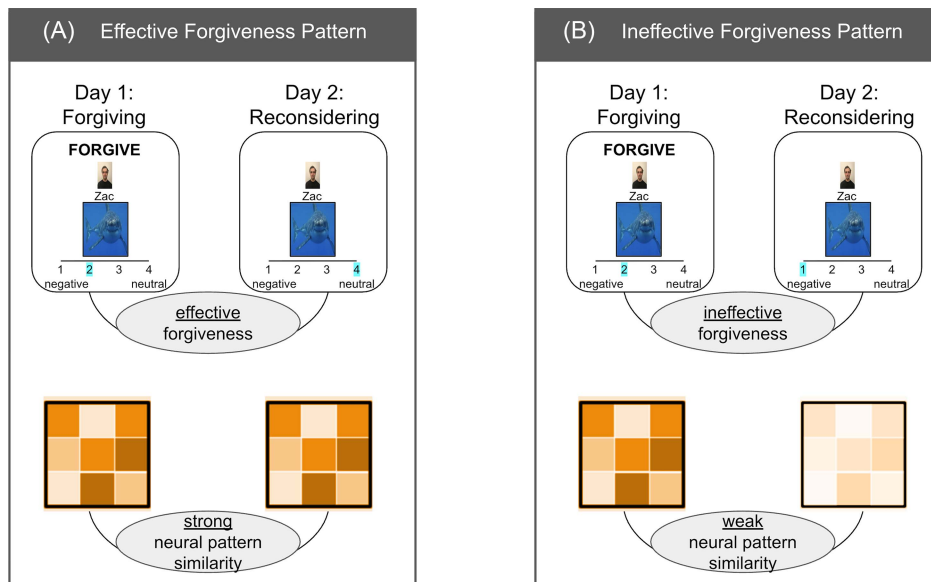
### *Changes in Neural Pattern Similarity for Forgiven Versus Looked at Targets as a Function of Affect Rating Change*

Next, we tested our primary hypothesis: Information from the moment of forgiveness becomes incorporated into forgiven experiences to make them less negative upon reconsideration. We performed an item-level analysis on the negative image trials that assessed the extent to which multivariate neural patterns increased (or decreased) in similarity between two phases of the experiment as a function of how much the affective rating of that trial changed to be less negative. Figure 3 shows a conceptual schematic of what the approach assesses, indicating that neural pattern similarity between the forgiveness manipulation on Day 1 and reconsideration on Day 2 should be high when forgiveness was effective. We tested for significant interactions to see if changes in pattern similarity varied as a function of target (forgiven vs. looked at target) and the extent to which an item's affective ratings changed.

Two of our targeted ROIs demonstrated our prediction: The rDMPFC ROI derived in Neurosynth from the search term “mentalizing” and the left posterior hippocampus from the Shen parcellation (Parcel Number 229) showed a significant change in neural pattern similarity between manipulation and reconsideration for the forgiven versus looked at target as a function of affective rating change—Neurosynth rDMPFC:  $\beta = -0.062$ ,  $t(1352.81) = -2.78$ , FDR-corrected  $p = .030$ , 95% CI [-0.105, -0.016] (Figure 4B); Shen Parcel Number 229 left posterior hippocampus:  $\beta = -0.052$ ,  $t(1351.99) = -2.83$ , FDR-corrected  $p = .030$ , 95% CI [-0.088, -0.016] (Figure 4E). We also checked if our rDMPFC Neurosynth result replicated when using the rDMPFC region from the Shen parcellation. Indeed, the Shen rDMPFC parcel also showed a significant interaction: Shen Parcel 10, rDMPFC,  $\beta = -0.052$ ,  $t(1349.82) = -2.47$ ,  $p = .014$ , 95% CI [-0.094, -0.011].

To help interpret these significant interactions, we performed follow-up simple slope analysis to reveal where in the continuum of affective change scores the pattern similarity was significantly different between conditions. This approach compares the slopes for each target at each level of affective change score, which are all reported in Figure 5. The overall pattern across the rDMPFC and posterior hippocampus ROIs was that the slopes were significantly different between the forgiven and looked at target when the image became (a) more negative between manipulation and reconsideration and (b) less negative between manipulation and reconsideration. This suggests two things: (a) Forgiven trials that were “effective” (i.e., the ratings became less negative on Day 2 reconsideration than the moment of forgiveness) correspond with greater pattern similarity between manipulation and reconsideration for the forgiven (vs. looked at) target as well as (b) forgiven trials that were “ineffective” (i.e., the ratings became more negative on Day 2 reconsideration than the moment of forgiveness) correspond with less pattern similarity between manipulation and reconsideration for the forgiven (vs. looked at) target, indicating the memory representation was actually dissimilar from the representation of forgiveness.

**Figure 3**  
*Conceptual Schematic of the Neural Pattern Similarity Analysis Plan*



*Note.* Panel A displays a hypothetical example of a forgiven target trial during the forgiveness manipulation phase on Day 1 and the reconsideration phase for the same trial on Day 2. In this example, the participant's rating changed from a value of 2 while forgiving to a value of 4 during reconsideration, indicating that forgiveness updated the memory to be less upsetting by Day 2 (i.e., forgiveness was "effective"). In this scenario, we predict that the neural pattern when reconsidering this memory on Day 2 should look similar to the neural pattern while forgiving on Day 1, indicating that information from forgiveness became incorporated into the memory representation. Panel B displays a hypothetical example for a trial for which forgiveness was ineffective—because the affect rating on Day 2 was actually more negative than during forgiveness, indicating that forgiveness did not help make the memory of the event less upsetting. Here, we do not expect information from forgiveness to have become incorporated into the memory trace. The image shown is approved for redistribution under creative commons licensing and can be found at [https://commons.wikimedia.org/wiki/File:Great\\_white\\_shark\\_south\\_africa.jpg](https://commons.wikimedia.org/wiki/File:Great_white_shark_south_africa.jpg) (Creative Commons Attribution 2.0 Generic license). See the online article for the color version of this figure.

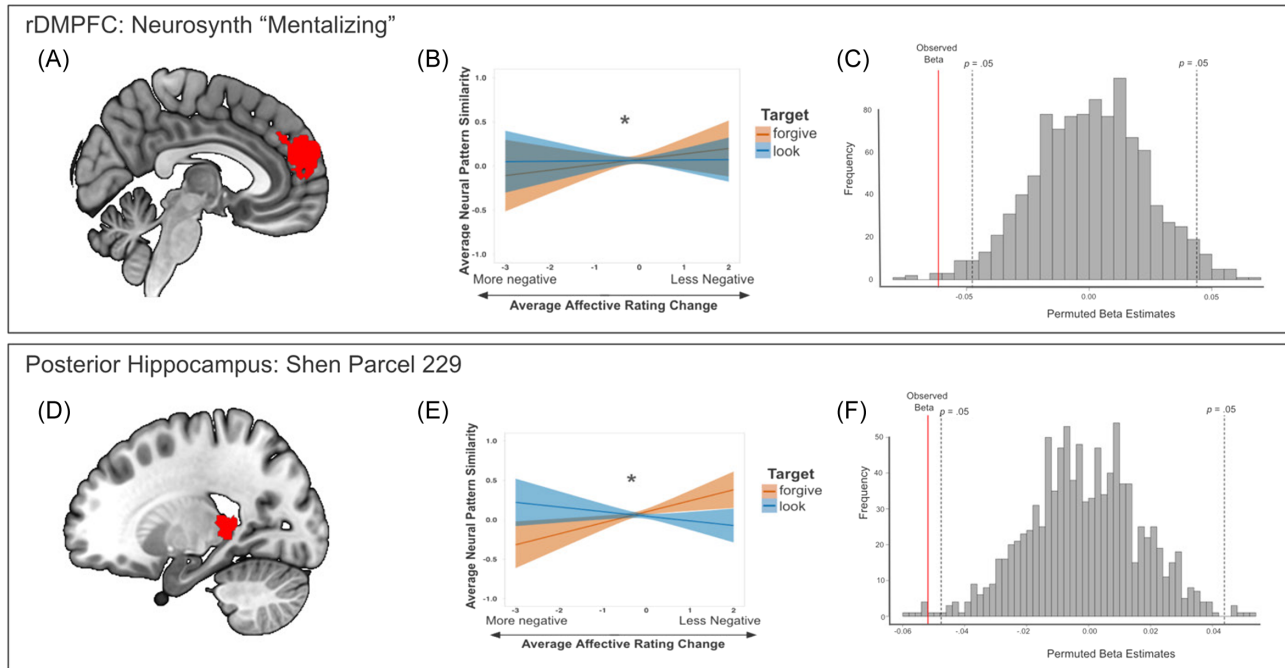
Finally, to check the robustness of our neural pattern similarity results, we performed nonparametric permutation testing to assess whether our observed neural pattern similarity results are meaningful, above and beyond what would be expected from noise. We conducted a permutation test ( $n = 1,000$  permutations) to assess the interaction between affective rating differences and target condition (forgive vs. look) on neural pattern similarity in each of our ROIs showing the parametric results reported above. This added further support for our neural findings. The significant change in neural pattern similarity between manipulation and reconsideration for the forgiven versus looked at target as a function of affective rating change was significantly greater than what would be expected from noise in both of the DMPFC ROIs (Neurosynth rDMPFC,  $\beta = -0.062$ ,  $p = .006$ , Figure 4C; rDMPFC Shen Parcel 229,  $\beta = -0.052$ ,  $p = .016$ ). This was also true for the left posterior hippocampus Shen Parcel 229 ( $\beta = -0.052$ ,  $p = .002$ ; Figure 4F).

Notably, across all of our targeted ROIs, neural pattern similarity did not change as a function of affective rating for the forgiven versus looked at target when pattern similarity between encoding and manipulation was assessed ( $\beta s < 0.03$ ,  $t s < 1.94$ , FDR-corrected  $p s > .573$ ), which is consistent with prior work indicating incorporation should emerge after a delay (Cowan et al., 2021; McCullough et al.,

2003; Sekeres et al., 2017). The same analysis when encoding and reconsideration were assessed was also null ( $\beta s < 0.04$ ,  $t s < 2.72$ , FDR-corrected  $p s > .073$ ), likely because trials compared between these two timepoints are identical, and thus, similarity will overshadow subtle incorporation of forgiveness content. Finally, no additional regions emerged as statistically significant in the parallel, exploratory, whole-brain searches across all 268 Shen parcels ( $\beta s < 0.02$ ,  $t s < 1.32$ , FDR-corrected  $p s > .526$ ).

### *Changes in Neural Pattern Similarity for Forgiven Versus Looked at Targets, Regardless of Affect Ratings*

As described above, our primary prediction about forgiveness considers participant's affect ratings. This is because forgiveness is defined as the voluntary transformation of feelings, attitudes, and behavior such that a person is no longer upset with someone who wronged them (American Psychological Association, 2018). Thus, forgiveness should pertain specifically to the instances in which participants' affect ratings changed to be less negative. Nonetheless, it is still worth examining whether neural pattern similarity may have changed between trials as a function of having *tried* to forgive the target (vs. look at the target), regardless of whether the attempt to forgive was effective.

**Figure 4***Neural Pattern Similarity Results***Upsetting memories that are less negative on Day 2 reconsideration show greater neural pattern similarity with the forgiveness pattern from Day 1 in DMPFC and hippocampus**

*Note.* The figure shows the multivariate neural pattern similarity between trials from the reconsideration and manipulation phases as a function of how much the affective rating of that trial changed to be less negative. (A) The DMPFC Neurosynth region of interest is derived from the search term “mentalizing.” (D) The posterior hippocampus region of interest is derived from the Shen parcellation. (B, E) Multivariate neural patterns were more similar for reconsideration and forgiveness trials (vs. reconsideration and look trials) for trials in which the affective rating was less negative on Day 2 reconsideration than during Day 1 manipulation in the DMPFC and posterior hippocampus. Note that while these analyses are run for each item and each subject, the individual items are averaged in the displayed graphs for ease of visualization. (C, F) The DMPFC and hippocampus results were confirmed with a nonparametric permutation test, showing that the observed effect exceeds what would be expected from noise. DMPFC = dorsomedial prefrontal cortex; rDMPFC = right DMPFC. See the online article for the color version of this figure.

\*  $p < .05$ .

To examine this possibility, for each trial, we examined neural pattern similarity between each set of experimental phases and compared how item similarity varied for the forgiven (vs. looked at) targets, leaving out a given trial’s affect rating. None of the targeted ROIs showed significant changes in neural pattern similarity for the forgiven (vs. looked at) targets between encoding and manipulation ( $\beta < 0.04$ ,  $t < 2.76$ , FDR-corrected  $ps > .064$ ) or encoding and reconsideration ( $\beta < 0.05$ ,  $t < 2.94$ , FDR-corrected  $ps > .052$ ). However, we did observe results in the amygdala and hippocampus when examining neural similarity between manipulation and reconsideration. Amygdala neural patterns became less similar between the manipulation and reconsideration phase for the forgiven (vs. looked at) target—this was observed both in the left amygdala Neurosynth ROI associated with the search term “emotion regulation” ( $\beta = -0.03$ ),  $t(1356.00) = -2.66$ , FDR-corrected  $p = .043$ , 95% CI  $[-0.06, -0.01]$ , as well as with a follow-up replication using the Shen left amygdala ROI, Parcel Number 228 ( $\beta = -0.03$ ),  $t(1356.00) = -2.46$ ,  $p = .014$ , 95% CI  $[-0.06, -0.01]$ . Additionally, one hippocampus ROI showed greater neural pattern similarity between manipulation and reconsideration—left hippocampus Shen

Parcel 230 ( $\beta = 0.05$ ),  $t(1356.00) = 3.31$ , FDR-corrected  $p = .005$ , 95% CI  $[0.02, 0.08]$ . Although these amygdala and hippocampus results were not predicted, in the Discussion section, we speculate on their potential causes. The same analyses performed across all 268 Shen parcels did not reveal any additional significant results ( $\beta < 0.04$ ,  $ts < 1.72$ , FDR-corrected  $ps > .261$ ).

## Discussion

While it is well-known that forgiveness can restore social connections (Bono et al., 2008; vanOyen Witvliet & Luna, 2018), how forgiveness changes our perception of the harm done is not fully understood. We developed a paradigm to assess how forgiveness shapes memories of poor treatment. Participants were led to believe that another set of supposed participants selected which stimuli they saw while undergoing fMRI. On the first day of fMRI scanning, participants were instructed to view and make affective ratings for images that had been selected for them to view by two other supposed participants (i.e., “targets”) whose faces appeared on the same screen. Afterward, participants were given reasons for the two

**Figure 5**  
Simple Slopes Results

Change Value	rDMPFC Neurosynth "Mentalizing"			rDMPFC Shen Parcel 10			Posterior Hippocampus Shen Parcel 229		
	$\beta$	$t$	$p$	$\beta$	$t$	$p$	$\beta$	$t$	$p$
-3	.17	2.62	.009**	.14	2.24	.025*	.12	2.21	.027*
-2	.11	2.47	.014*	.09	2.07	.039*	.07	1.85	.064
-1	.05	1.90	.058	.04	1.47	.142	.02	.80	.426
0	.01	0.64	.522	.02	0.91	.369	.03	2.30	.025*
1	.07	2.39	.017*	.07	2.32	.021*	.09	3.40	<.001***
2	.13	2.68	.008**	.12	2.49	.013*	.14	3.31	<.001***

*Note.* The figure shows the results from simple slope analyses to determine where in the affective rating change continuum neural pattern similarity was different for the forgiven versus looked at target. Overall, across the rDMPFC and posterior hippocampus regions of interest, the simple slope analysis indicates a crossover interaction such that neural pattern similarity is meaningfully different between the forgiven versus looked at target when affect ratings become more and less subjectively negative. rDMPFC = right dorsomedial prefrontal cortex.

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

targets' negative image selections, with one apologetic and the other one nonchalant. Negative images were shown once again while participants were asked to make their best attempt to "forgive" or simply "look" while they saw and rated the images. On the second day of fMRI scanning, participants saw and rated all of the images once more. Results showed that effective forgiveness updates neural representations in the DMPFC and posterior hippocampus such that information from the moment of forgiveness becomes incorporated into forgiven experiences to make them less negative later on. Specifically, trial-by-trial analyses revealed that images that were rated as less negative on Day 2 (vs. Day 1) showed correspondingly greater neural pattern similarity in the DMPFC between the reconsideration and forgiveness trials (vs. the reconsideration and look trials). The results suggest that information considered during the momentary act of forgiveness may get incorporated into memories of a transgression to help make them less negative moving forward.

The results provide a novel mechanism by which forgiveness can change perceptions in a recipient of maltreatment. While there are many definitions of forgiveness, scholars agree that it should reflect some change in the individual such that they are less upset by a transgression afterward (McCullough, 2001). Yet, the various ways in which this change may occur remain to be fully determined. Meanwhile, cognitive psychology and neuroscience research indicates that new information can update an initial representation of an event prior to long-term storage (Sekeres et al., 2017; Shohamy & Wagner, 2008). To our knowledge, no prior memory research has tested these principles in interpersonal phenomena where representations of social experiences are changed, such as in the case of forgiveness. Our findings update both literatures, providing a novel way to conceptualize how forgiveness may work while also extending prior memory updating research into the interpersonal realm, specifically here in the context of forgiveness. Memory updating appears to be one novel pathway in which forgiveness can occur.

Memory updating through forgiveness did not occur throughout the brain, but instead emerged in two regions: the DMPFC and posterior hippocampus. The DMPFC demonstrated this result both in the region identified in the meta-analytic platform, Neurosynth (Yarkoni et al., 2011), with the search term "mentalizing," as well as in a follow-up analysis with the DMPFC parcel identified by the Shen parcellation (Shen et al., 2013). The DMPFC is robustly implicated in mentalizing, with convergent evidence from human and nonhuman primates and across neuroscience methods, including fMRI and electrocorticography (Báez-Mendoza et al., 2021; Denny et al., 2012; Tan et al., 2022). Recent work also implicates the DMPFC in interpersonal memory consolidation (Brietzke et al., 2024; Collier & Meyer, 2020; Jimenez & Meyer, 2024). For example, when participants rest after watching videos of people and places, the DMPFC preferentially reinstates the social videos during rest, which predicts subsequent memory for them (Jimenez & Meyer, 2024). The findings reported here extend this growing literature, as, to our knowledge, they are the first to show that representations of interpersonal experiences can be dynamically updated in the DMPFC.

The forgiveness updating result in the posterior portion of the left hippocampus may indicate that fine-grain details about the transgressor are becoming incorporated into memories via forgiveness, given what is known about the posterior hippocampus. Extensive research suggests there are functional gradients along the hippocampus, with different populations of neurons likely supporting different aspects of episodic memory (Bein et al., 2020; Thorp et al., 2022; Treves & Rolls, 1994). One perspective argues that the level of details in memories aligns with anterior-to-posterior hippocampal functional divisions, with the anterior hippocampus more associated with "gist"-like representations and the posterior hippocampus more associated with fine-grained details of specific content (Poppenk et al., 2013; Sekeres et al., 2018). In rodents, research examining the Cornu Ammonis (CA) areas of the hippocampus demonstrates that CA2 neurons—which fall within the human posterior hippocampus

(Schafer & Schiller, 2018)—causally support recognition of littermates (Hitti & Siegelbaum, 2014; Meira et al., 2018; Smith et al., 2016). Recent work links these two functions, showing that CA2 neurons represent littermates in high-dimensional detail. Thus, it is possible that specific details about a transgressor that are considered during forgiveness become incorporated into memories for the event. Future research is needed to identify the precise aspects of the representational change in memories that occurs in the hippocampus, as well as the DMPFC, through forgiveness.

The neural pattern similarity results in the DMPFC and posterior hippocampus demonstrated “crossover” interactions, meaning it is not only the case that there was greater neural pattern similarity for the forgiven (vs. looked at) target when forgiveness was effective (i.e., the affect rating was less negative on Day 2 reconsideration than during the moment of forgiveness itself). We also observed *less* neural pattern similarity for the forgiven (vs. looked at) target when forgiveness was ineffective (i.e., the affect rating was more negative on Day 2 reconsideration than during the moment of forgiveness itself). This suggests that in some cases, attempting to forgive someone may backfire by making the experience even more upsetting, and when this happens, the memory trace in the DMPFC and posterior hippocampus is actually dissimilar from the attempt to forgive.

Forgiveness is defined as the voluntary decrease in negative feelings, attitudes, and behavior felt toward a wrongdoer (American Psychological Association, 2018). This is why our primary multivariate analyses focused on changes in neural representations as a function of affect rating changes. That said, additional, exploratory analyses revealed that simply trying to forgive (vs. look at) a target, collapsing across affect ratings, changed neural patterns in the left amygdala. The left amygdala ROI derived from the search term “emotion regulation” in Neurosynth demonstrated that neural patterns in this region became less similar between manipulation and reconsideration for the forgiven (vs. looked at) target. This result was replicated with the left amygdala region from the Shen parcellation (Parcel 228; Shen et al., 2013).

The amygdala findings complement prior work on emotion regulation on the one hand and memory consolidation on the other. In terms of emotion regulation, prior work finds amygdala responses are changed in the moment of cognitive reappraisal (Buhle et al., 2014). Our finding is similar; however, here the change in the amygdala response emerges only after time has passed. Relatedly, a prior emotion-regulation study examined whether there are lasting impacts on the amygdala response to negative images and found that negative images that were reappraised multiple times (vs. a single time) also showed altered amygdala responses 1 week later (Denny et al., 2012). Collectively, these results suggest (a) that various forms of emotion regulation, ranging from “classic” cognitive reappraisal strategies to interpersonal forgiveness, may have lasting impacts on amygdala responses and (b) that the form of regulation (e.g., interpersonal forgiveness vs. other reappraisal strategies) and frequency (e.g., number of times the stimulus is reconsidered) may determine when and for how long amygdala responses are changed. In terms of memory consolidation, memory updating can take time, and as the emotional content of a memory changes, so do multivariate patterns in subcortical regions (Speer et al., 2021). Forgiveness may trigger consolidation-dependent changes in the amygdala response that are only detectable when time has passed. This possibility dovetails longitudinal forgiveness research demonstrating that the affective consequences of forgiveness increase over time, for example, over at least 4 weeks (McCullough et al., 2003).

We also observed evidence that neural pattern similarity in portions of the hippocampus increased between manipulation and reconsideration in the forgive (vs. look) condition. This result was in a portion of the hippocampus (Parcel 230) distinct from the posterior hippocampus parcel demonstrating the effective forgiveness results (Parcel 229). It is possible that in the memory consolidation process, different subcomponents of the posterior hippocampus are sensitive to incorporating episodic details considered during (a) simply the attempt to forgive and (b) more specifically effective forgiveness. Interestingly, one philosophical account of forgiveness is the “episodic fading model,” which suggests that episodic details of a transgression reduce over time (Blustein, 2014; Boleyn-Fitzgerald, 2002; Noreen et al., 2014). Yet, recent empirical work failed to find support for the episodic fading account (Fernández-Miranda et al., 2025). This prior work, paired with our findings, suggests that rather than episodic details fading, forgiveness may change which episodic details are emphasized, an effect modulated by forgiveness effectiveness.

Our hypothesis pertained to representational changes that occur in the process of forgiveness, which is why our primary analyses focused on multivariate neural pattern similarity. Nonetheless, it is noteworthy that we did not observe changes in univariate levels of activation across our conditions. Yet, prior emotion-regulation work using similar paradigms (outside of forgiveness) does find differences in univariate activity as a function of trying to reappraise (vs. look at) negative stimuli. The instructions we provided participants to try to forgive a target are more subtle than those used in emotion-regulation research, which may have made it harder to detect differences in univariate levels of activity. In fMRI reappraisal research, just prior to a participant’s brain scan, they are given explicit instructions on how to reappraise negative images and complete reappraisal training exercises (Buhle et al., 2014; Denny et al., 2012). In contrast, to mirror how spontaneous forgiveness may occur in everyday life, we did not tell our participants before their brain scan that they would later be asked to try to forgive a target while in the scanner. Instead, after initial encoding, participants observed the rationales of each target; through the intercom that the experimenter used to communicate to the participant in the scanner, participants were then told to try to forgive the target who apologized, without explicit instructions on how to do so. This more subtle instruction means that participants tried to forgive in whatever way felt natural to them, and may explain why we did not observe “macro” differences between conditions in the average amount of activity across voxels in a brain region, despite observing forgiveness corresponded with “micro” changes in fine-grained, multivariate neural patterns between conditions.

That said, the left VLPFC, a region reliably associated with response inhibition, including cognitive reappraisal (Ochsner & Gross, 2005), also did not show “micro” changes in the fine-grained, multivariate neural patterns as a function of effective forgiveness. Reappraisal broadly refers to the process of changing the meaning of the stimulus, and prior work identifying the IVLPFC in this process tends to examine reappraisal strategies in which participants imagine changing features of the stimulus, such as imagining a threatening stimulus is far away from themselves (i.e., “distancing”; Ochsner et al., 2004). In contrast, forgiveness may require reappraising the wrongdoer’s intentions that triggered the events on the ground, rather than changing the facts on the ground. Indeed, prior work suggests portions of DMPFC can regulate emotions by considering the abstract meaning behind experiences (Kensinger &

Ford, 2021; Ochsner & Gross, 2005), which is conceptually similar to considering the intentions driving another person's behavior. In this way, the DMPFC may support the incidental reappraisal of negative interpersonal experiences through perspective taking. Consistent with this possibility, a prior study found that when participants encode negative IAPS images from a "tough" or "fearful" person's perspective, neural patterns in the medial prefrontal cortex, but not the IVPFC, decrease and increase emotional responses according to the imagined target's perspective (Gilead et al., 2016). There is growing research interest on the topic of "social emotion regulation," in which people regulate one another's emotional reactions, including their subjective affect and physiological responding (He et al., 2025; Inagaki et al., 2023; Reeck et al., 2016; Sahi et al., 2021, 2023, 2025; Shu et al., 2021; Zaki & Williams, 2013). It is possible that dorsal and medial portions of the prefrontal cortex, rather than lateral prefrontal regions, play a key role in social emotion regulation. Representational change in DMPFC through forgiveness may be one instance of a broader class of phenomena in which other people's minds regulate our own.

In our paradigm, the wrongdoing and forgiveness occur close in time, which means our observed neural similarity between Day 2 reconsideration and Day 1 forgiveness likely reflects an integration process during initial learning and consolidation of the memory. However, forgiveness often occurs not directly after a transgression, but instead after more time has passed. For example, a transgressor may apologize to a victim weeks, months, or even years later. An interesting future direction will be to examine whether delayed forgiveness can update old memories of wrongdoing. Even after a memory is formed, later reconsideration can make it labile and susceptible to incorporating new information (Alberini & Ledoux, 2013; Lee et al., 2017; Loftus, 2005; Schacter, 1999; Schiller et al., 2010), a process termed "reconsolidation." Relevant to forgiveness, a great deal of reconsolidation research focuses on reductions in the negative affect associated with a memory (Agren et al., 2012; Kessler et al., 2017; Kredlow et al., 2016; Lane et al., 2015). Since forgiving occurs in response to aversive events and makes them less negative, it is possible that a similar updating process occurs endogenously during forgiveness. If this were the case, future work would need to demonstrate that long-term memory updating via forgiveness occurs specifically within the "reconsolidation window" (Monfils & Holmes, 2018; Schiller et al., 2010). For example, in humans, there is a heightened skin conductance response to conditioned fear stimuli (Schiller et al., 2010). If a day after committing these fear associations to memory there is a "manipulation phase" in which the fear association is retrieved and within 6 hr extinguished, the skin conductance response is greatly reduced (relative to "encoding"). And critically, the next day when participants see the stimuli again, their skin conductance response is still reduced, appearing similar to their "manipulation phase" skin conductance response. However, if the extinction occurs later than this 6-hr window, it does not effectively update the original fear memory. Future research could develop similar paradigms to test whether the negative affect-reducing qualities of forgiveness not only occur if an apology occurs close enough in time to influence the consolidation of the original experience, but also whether forgiveness adheres to the principles of reconsolidation.

The goal of this study was to search for neural evidence of memory updating through forgiveness. This is because there is a robust cognitive neuroscience literature to inform the hypotheses

and multivariate neural pattern similarity is well-suited to test for fine-grain changes in mental representations. That said, recent progress in natural language processing methods offers an exciting opportunity to test whether the effects we observed at the neural level of analysis replicate at the subjective level of analysis, with respect to how people describe their memories of transgressions and forgiveness. Natural language processing approaches quantify semantic information in a very similar fashion to multivariate neural pattern similarity analyses: In both approaches, multivariate data are transformed into a vector, with each cell of the vector reflecting "representational content"—either neural activity or a sentence's semantics. The next steps in natural language processing proceed similarly to the approach used here, for example, by comparing the similarity between a semantic vector representing a participant's description of why they forgave someone and their memory of the transgression the next day. Thus, future natural language processing studies are well-positioned to replicate and extend our neural results.

In the present study, there were two targets—one apologetic and one nonchalant. We chose a nonchalant target as the control condition target instead of a target that was actively unapologetic, so that our forgive condition was compared to a relatively "neutral" condition. This mirrors prior emotion-regulation paradigms and helps ensure neural results preferential to forgiveness, rather than, for example, increased negative affect and resistance to memory change in response to an actively unapologetic target. An interesting future direction will be to add an unambiguously unapologetic target to this type of paradigm. It is possible that in this scenario, affective ratings will become more negative because participants will resent that lack of remorse, and this could actually also correspond with enhanced memory updating as reflected by increased neural pattern similarity to the attempt to forgive—though in this case, the neural pattern similarity pertains to mental representations of resentment. The memory malleability literature suggests that memories can be updated to be either less negative or more negative—what matters is that new affectively relevant information is considered in conjunction with the other details of the encoded event (Schiller et al., 2010; Speer et al., 2021). A transgressor's genuine apologies may update memories of the transgression to be less negative, while actively unapologetic behaviors from a transgressor may update memories to be more negative. Both possibilities are relevant to the victim meeting their social connection goals, as a clearly unremorseful transgressor may be someone with whom we do not want to reconcile and instead are best off avoiding in the future.

A final note about our approach—we developed a novel paradigm to investigate how forgiveness influences memories for negative treatment from a stranger. However, forgiveness often unfolds between people in existing relationships, rather than strangers (McCaughey et al., 2022). Additionally, in everyday life, full forgiveness does not always happen right away and instead takes time to effectively reduce negative feelings toward relationship partners (McCullough et al., 2003). Longitudinal designs that also investigate forgiveness toward close others from our social life are needed to determine how forgiveness may update our real-world memories of interpersonal conflict. One possibility is that the same neural mechanisms observed here apply to forgiveness in everyday life. Alternatively, given that the DMPFC is often associated with mentalizing about strangers, whereas the anterior MPFC is associated with mentalizing about close others (Krienen et al., 2010), it is possible that forgiveness updates memories for maltreatment from

strangers and close others through different subcomponents of the prefrontal cortex. Future research is needed to arbitrate between these possibilities. Our approach to measuring changes in multivariate patterns over time in response to strangers' behaviors provides a backbone for future research to examine related questions between relationship partners.

## Limitations

Certain limitations of the present work are worth highlighting. First, receiving negative images from a stranger to view in an fMRI study is a relatively artificial interpersonal experience. We developed our paradigm to mirror prior research on cognitive reappraisal, which also uses highly negative images to generate an emotional response and compares neural responses when attempting to reappraise (vs. look at) the images (Buhle et al., 2014). By staying true to this paradigm, our results can be more easily compared and contrasted with prior neuroimaging results. However, the cost comes with less external validity and future research using more ecologically valid paradigms will be useful.

Second, the sample size is relatively small. This concern is slightly offset by the fact that (a) our primary results come from trial-by-trial multivariate analyses within subjects, which are more robust to small samples than either univariate or between-subject analyses (Nee, 2019; Rosenberg & Finn, 2022), and (b) the multivariate neural results were replicated with nonparametric permutation testing that is robust to small samples. That said, it is important to replicate the present study in future work. Such replications could take the form of larger scale fMRI studies to directly replicate the present findings or the form of follow-up behavioral studies that examine the same premises at different levels of analysis. For instance, and as noted above, the experimental paradigm could be adapted such that participants provide written responses of their reactions to each experimental phase (i.e., encoding, manipulation [forgive vs. look], and reconsideration). Natural language processing methods could be applied to the written responses to test whether, after a delay (e.g., 24 hr), participants have incorporated the wrongdoer's perspective and/or certain episodic details of the apology into their reconsideration of the event. Prior works on forgiveness and memory updating are consistent with these ideas (Fernández-Miranda et al., 2025; Speer et al., 2021), though they have not yet been demonstrated in the semantic content of memories. While not a direct replication of the DMPFC or hippocampus findings, this approach could conceptually replicate them with large samples and offer additional insight into what aspects of mentalizing and episodic memory are updated through forgiveness.

Third, although our primary prediction was for increased multivariate pattern similarity between the forgiveness and reconsideration phase for effectively forgiven trials (which we found support for), we did not observe decreases in multivariate pattern similarity between the reconsideration and initial encoding phase for effectively forgiven trials. This may reflect the fact that both the stimuli and task (i.e., "rate the images given to you") during encoding and reconsideration were identical, and these similarities may have overshadowed our ability to detect subtle representational change. In other words, for effective forgiveness trials, participants had incorporated information from the moment of forgiveness, but this update was not detectable in the reconsideration versus encoding comparison. In everyday life, it is rare that the context of later

reconsideration of a wrongdoing is identical to initial encoding, and thus, more naturalistic paradigms may be able to detect representational change between reconsideration and initial encoding.

## Conclusion

Around the world, people are encouraged to forgive wrongdoers to help themselves move on from the event and restore social connection. While extensive social psychology research points to the benefits of forgiveness for the victim and transgressor (Gismero-González et al., 2020; Griffin et al., 2015; Webb & Toussaint, 2020), the underlying mechanisms underpinning forgiveness are not fully understood. To begin to fill this gap, we showed that information from the moment of forgiveness becomes incorporated into the memory of the transgression. This incorporation occurred in two brain regions—a portion of DMPFC reliably associated with mentalizing (i.e., inferring people's mental states, feelings, and intentions) and a portion of the hippocampus reliably associated with episodic memory. Overall, our results suggest that rather than "forgive and forget," we may "forgive and update" our representations of negative experiences, a process that may foster reconciliation and connection.

## References

- Agren, T., Engman, J., Frick, A., Björkstrand, J., Larsson, E. M., Furmark, T., & Fredrikson, M. (2012). Disruption of reconsolidation erases a fear memory trace in the human amygdala. *Science*, *337*(6101), 1550–1552. <https://doi.org/10.1126/science.1223006>
- Alberini, C. M., & Ledoux, J. E. (2013). Memory reconsolidation. *Current Biology*, *23*(17), R746–R750. <https://doi.org/10.1016/j.cub.2013.06.046>
- Amaya, S. (2019). Forgiving as emotional distancing. *Social Philosophy and Policy*, *36*(1), 6–26. <https://doi.org/10.1017/S0265052519000311>
- American Psychological Association. (2018). Forgiveness. In *APA dictionary of psychology*. Retrieved July 31, 2024, from <https://dictionary.apa.org/forgiveness>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901. <https://doi.org/10.1016/j.neuropsychologia.2007.06.015>
- Báez-Mendoza, R., Mastrobattista, E. P., Wang, A. J., & Williams, Z. M. (2021). Social agent identity cells in the prefrontal cortex of interacting groups of primates. *Science*, *374*(6566), Article eabb4149. <https://doi.org/10.1126/science.abb4149>
- Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala-frontal connectivity during emotion regulation. *Social Cognitive and Affective Neuroscience*, *2*(4), 303–312. <https://doi.org/10.1093/scan/nsm029>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Behzadi, Y., Restom, K., Liao, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042>
- Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, *11*(1), Article 3451. <https://doi.org/10.1038/s41467-020-17287-1>

- Billingsley, J., & Losin, E. A. R. (2017). The neural systems of forgiveness: An evolutionary psychological perspective. *Frontiers in Psychology*, 8, Article 737. <https://doi.org/10.3389/fpsyg.2017.00737>
- Blustein, J. (2014). *Forgiveness and remembrance: Remembering wrongdoing in personal and public life*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199329397.001.0001>
- Boleyn-Fitzgerald, P. (2002). What should “forgiveness” mean? *The Journal of Value Inquiry*, 36(4), 483–498. <https://doi.org/10.1023/A:1021929531830>
- Bono, G., McCullough, M. E., & Root, L. M. (2008). Forgiveness, feeling connected to others, and well-being: Two longitudinal studies. *Personality and Social Psychology Bulletin*, 34(2), 182–195. <https://doi.org/10.1177/0146167207310025>
- Brietzke, S. C., Barbarossa, K., & Meyer, M. L. (2024). Get out of my head: Social evaluative brain states carry over into post-feedback rest and influence remembering how others view us. *Cerebral Cortex*, 34(7), Article bhae280. <https://doi.org/10.1093/cercor/bhae280>
- Broom, T. W., & Wagner, D. D. (2023). The boundary between real and fictional others in the medial prefrontal cortex is blurred in lonelier individuals. *Cerebral Cortex*, 33(16), 9677–9689. <https://doi.org/10.1093/cercor/bhad237>
- Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., Weber, J., & Ochsner, K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex*, 24(11), 2981–2990. <https://doi.org/10.1093/cercor/bht154>
- Chavez, R. S., & Wagner, D. D. (2020). The neural representation of self is recapitulated in the brains of friends: A round-robin fMRI study. *Journal of Personality and Social Psychology*, 118(3), 407–416. <https://doi.org/10.1037/pspa0000178>
- Collier, E., & Meyer, M. L. (2020). Memory of others’ disclosures is consolidated during rest and associated with providing support: Neural and linguistic evidence. *Journal of Cognitive Neuroscience*, 32(9), 1672–1687. [https://doi.org/10.1162/jocn\\_a\\_01573](https://doi.org/10.1162/jocn_a_01573)
- Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., & Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bulletin & Review*, 28(6), 1796–1810. <https://doi.org/10.3758/s13423-021-01978-x>
- Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10(4–5), 171–178. [https://doi.org/10.1002/\(SICI\)1099-1492\(199706/08\)10:4/5<171::AID-NBM453>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L)
- Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, 136(1), 87–102. <https://doi.org/10.1037/a0017937>
- Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, 24(8), 1742–1752. [https://doi.org/10.1162/jocn\\_a\\_00233](https://doi.org/10.1162/jocn_a_00233)
- Enright, R. D., & North, J. (Eds.). (1998). *Exploring forgiveness*. University of Wisconsin Press.
- Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). *fMRIPrep* (Version 23.1.3) [Computer software]. <https://doi.org/10.5281/zenodo.852659>
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). *fMRIPrep*: A robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1), 111–116. <https://doi.org/10.1038/s41592-018-0235-4>
- Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., Berleant, S., Ellis, D. G., Pinsard, B., Madison, C., Waskom, M., Notter, M. P., Clark, D., Manhães-Savio, A., Clark, D., Jordan, K., Dayan, M., Halchenko, Y. O., Loney, F., ... Ghosh, S. (2025). *nipy/nipype: 1.8.6 (1.8.6)*. Zenodo. <https://doi.org/10.5281/zenodo.15054147>
- FeldmanHall, O., Montez, D. F., Phelps, E. A., Davachi, L., & Murty, V. P. (2021). Hippocampus guides adaptive learning during dynamic social interactions. *The Journal of Neuroscience*, 41(6), 1340–1348. <https://doi.org/10.1523/JNEUROSCI.0873-20.2020>
- Fernández-Miranda, G., Stanley, M., Murray, S., Faul, L., & De Brigard, F. (2025). The emotional impact of forgiveness on autobiographical memories of past wrongdoings. *Journal of Experimental Psychology: General*, 154(8), 2179–2200. <https://doi.org/10.1037/xge0001787>
- Fourie, M. M., Hortensius, R., & Decety, J. (2020). Parsing the components of forgiveness: Psychological and neural mechanisms. *Neuroscience and Biobehavioral Reviews*, 112, 437–451. <https://doi.org/10.1016/j.neubiorev.2020.02.020>
- Gilead, M., Boccagno, C., Silverman, M., Hassin, R. R., Weber, J., & Ochsner, K. N. (2016). Self-regulation via neural simulation. *Proceedings of the National Academy of Sciences of the United States of America*, 113(36), 10037–10042. <https://doi.org/10.1073/pnas.1600159113>
- Gismero-González, E., Jódar, R., Martínez, M. P., Carrasco, M. J., Cagigal, V., & Prieto-Urúa, M. (2020). Interpersonal offenses and psychological well-being: The mediating role of forgiveness. *Journal of Happiness Studies*, 21(1), 75–94. <https://doi.org/10.1007/s10902-018-00070-x>
- Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, 5, Article 13. <https://doi.org/10.3389/fninf.2011.00013>
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>
- Griffin, B. J., Worthington, E. L., Lavelock, C. R., Wade, N. G., & Hoyt, W. T. (2015). Forgiveness and mental health. In L. Toussaint, E. Worthington, & D. Williams (Eds.), *Forgiveness and health* (pp. 77–90). Springer. [https://doi.org/10.1007/978-94-017-9993-5\\_6](https://doi.org/10.1007/978-94-017-9993-5_6)
- Guthrie, T. D., Benadjaoud, Y. Y., & Chavez, R. S. (2022). Social relationship strength modulates the similarity of brain-to-brain representations of group members. *Cerebral Cortex*, 32(11), 2469–2477. <https://doi.org/10.1093/cercor/bhab355>
- Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, 2(3), 289–293. <https://doi.org/10.1038/6404>
- He, Z., Cohen, N., Shu, J., Bo, K., Wager, T. D., & Ochsner, K. N. (2025). Comparing the neural bases of self- and social-reappraisal. *Cerebral Cortex*, 35(8), bhaf189. <https://doi.org/10.1093/cercor/bhaf189>
- Hitti, F. L., & Siegelbaum, S. A. (2014). The hippocampal CA2 region is essential for social memory. *Nature*, 508(7494), 88–92. <https://doi.org/10.1038/nature13028>
- Inagaki, T. K., Alvarez, G. M., Orehek, E., Ferrer, R. A., Manuck, S. B., Abaya, N. M., & Muscatell, K. A. (2023). Support-giving is associated with lower systemic inflammation. *Annals of Behavioral Medicine*, 57(6), 499–507. <https://doi.org/10.1093/abm/kaac059>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841. <https://doi.org/10.1006/nimg.2002.1132>
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/10.1016/S1361-8415(01)00036-6)
- Jimenez, C. A., & Meyer, M. L. (2024). The dorsomedial prefrontal cortex prioritizes social learning during rest. *Proceedings of the National Academy of Sciences of the United States of America*, 121(12), Article e2309232121. <https://doi.org/10.1073/pnas.2309232121>
- Johnson, J. D., McDuff, S. G., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, 63(5), 697–708. <https://doi.org/10.1016/j.neuron.2009.08.011>

- Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, 31(8), 1169–1180. <https://doi.org/10.3758/BF03195800>
- Kensinger, E. A., & Ford, J. H. (2021). Guiding the emotion in emotional memories: The role of the dorsomedial prefrontal cortex. *Current Directions in Psychological Science*, 30(2), 111–119. <https://doi.org/10.1177/0963721421990081>
- Kessler, H., Blackwell, S. E., & Kehyayan, A. (2017). Reconsolidation and trauma memory. In N. Axmacher & B. Rasch (Eds.), *Cognitive neuroscience of memory consolidation* (pp. 369–379). Springer. [https://doi.org/10.1007/978-3-319-45066-7\\_22](https://doi.org/10.1007/978-3-319-45066-7_22)
- Kredlow, M. A., Unger, L. D., & Otto, M. W. (2016). Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. *Psychological Bulletin*, 142(3), 314–336. <https://doi.org/10.1037/bul0000034>
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, Article 4. <https://doi.org/10.3389/neuro.06.004.2008>
- Krienen, F. M., Tu, P. C., & Buckner, R. L. (2010). Clan mentality: Evidence that the medial prefrontal cortex responds to close others. *The Journal of Neuroscience*, 30(41), 13906–13915. <https://doi.org/10.1523/JNEUROSCI.2180-10.2010>
- 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. <https://doi.org/10.18637/jss.v082.i13>
- Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, 1(1), 76–85. <https://doi.org/10.1137/0701007>
- 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*, 38, Article e1. <https://doi.org/10.1017/S0140525X14000041>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-8). University of Florida.
- Lee, J. L. C., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in Cognitive Sciences*, 21(7), 531–545. <https://doi.org/10.1016/j.tics.2017.04.006>
- Lichtenfeld, S., Buechner, V. L., Maier, M. A., & Fernández-Capo, M. (2015). Forgive and forget: Differences between decisional and emotional forgiveness. *PLOS ONE*, 10(5), Article e0125561. <https://doi.org/10.1371/journal.pone.0125561>
- Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience and Biobehavioral Reviews*, 99, 311–328. <https://doi.org/10.1016/j.neubiorev.2018.12.021>
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361–366. <https://doi.org/10.1101/lm.94705>
- McCauley, T. G., Billingsley, J., & McCullough, M. E. (2022). An evolutionary psychology view of forgiveness: Individuals, groups, and culture. *Current Opinion in Psychology*, 44, 275–280. <https://doi.org/10.1016/j.copsyc.2021.09.021>
- McCullough, M. E. (2001). Coping with stress: Effective people and processes. In C. R. Snyder (Ed.), *Coping with stress: Effective people and processes* (pp. 93–113). Oxford University Press. <https://doi.org/10.1093/med:psych/9780195130447.001.0001>
- McCullough, M. E., Fincham, F. D., & Tsang, J. A. (2003). Forgiveness, forbearance, and time: The temporal unfolding of transgression-related interpersonal motivations. *Journal of Personality and Social Psychology*, 84(3), 540–557. <https://doi.org/10.1037/0022-3514.84.3.540>
- McCullough, M. E., Pedersen, E. J., Tabak, B. A., & Carter, E. C. (2014). Conciliatory gestures promote forgiveness and reduce anger in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 111(30), 11211–11216. <https://doi.org/10.1073/pnas.1405072111>
- McCullough, M. E., Worthington, E. L., Jr., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology*, 73(2), 321–336. <https://doi.org/10.1037/0022-3514.73.2.321>
- McGaugh, J. L. (2002). Memory consolidation and the amygdala: A systems perspective. *Trends in Neurosciences*, 25(9), 456–461. [https://doi.org/10.1016/S0166-2236\(02\)02211-7](https://doi.org/10.1016/S0166-2236(02)02211-7)
- Meira, T., Leroy, F., Buss, E. W., Oliva, A., Park, J., & Siegelbaum, S. A. (2018). A hippocampal circuit linking dorsal CA2 to ventral CA1 critical for social memory dynamics. *Nature Communications*, 9(1), Article 4163. <https://doi.org/10.1038/s41467-018-06501-w>
- Monfils, M. H., & Holmes, E. A. (2018). Memory boundaries: Opening a window inspired by reconsolidation to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry*, 5(12), 1032–1042. [https://doi.org/10.1016/S2215-0366\(18\)30270-0](https://doi.org/10.1016/S2215-0366(18)30270-0)
- Monsa, R., Dafni-Merom, A., & Arzy, S. (2024). What makes an event significant: An fMRI study on self-defining memories. *Cerebral Cortex*, 34(7), Article bhac303. <https://doi.org/10.1093/cercor/bhac303>
- Nee, D. E. (2019). fMRI replicability depends upon sufficient individual-level data. *Communications Biology*, 2(1), Article 130. <https://doi.org/10.1038/s42003-019-0378-6>
- Noreen, S., Bierman, R. N., & MacLeod, M. D. (2014). Forgiving you is hard, but forgetting seems easy: Can forgiveness facilitate forgetting? *Psychological Science*, 25(7), 1295–1302. <https://doi.org/10.1177/0956797614531602>
- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242–249. <https://doi.org/10.1016/j.tics.2005.03.010>
- Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*, 23(2), 483–499. <https://doi.org/10.1016/j.neuroimage.2004.06.030>
- Ochsner, K. N., Silvers, J. A., & Buhle, J. T. (2012). Functional imaging studies of emotion regulation: A synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, 1251(1), E1–E24. <https://doi.org/10.1111/j.1749-6632.2012.06751.x>
- Park, J. (2005). Effect of arousal and retention delay on memory: A meta-analysis. *Psychological Reports*, 97(2), 339–355. <https://doi.org/10.2466/pr0.97.2.339-355>
- Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, 144, 74–82. <https://doi.org/10.1016/j.neuroimage.2016.08.051>
- Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, 17(5), 230–240. <https://doi.org/10.1016/j.tics.2013.03.005>
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, 84, 320–341. <https://doi.org/10.1016/j.neuroimage.2013.08.048>
- R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>
- Reeck, C., Ames, D. R., & Ochsner, K. N. (2016). The social regulation of emotion: An integrative, cross-disciplinary model. *Trends in Cognitive Sciences*, 20(1), 47–63. <https://doi.org/10.1016/j.tics.2015.09.003>
- Ridderinkhof, K. R., van den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, 56(2), 129–140. <https://doi.org/10.1016/j.bandc.2004.09.016>

- Rosenberg, M. D., & Finn, E. S. (2022). How to establish robust brain-behavior relationships without thousands of individuals. *Nature Neuroscience*, 25(7), 835–837. <https://doi.org/10.1038/s41593-022-01110-9>
- Sahi, R. S., Gaines, E. M., Nussbaum, S. G., Lee, D., Lieberman, M. D., Eisenberger, N. I., & Silvers, J. A. (2025). You changed my mind: Immediate and enduring impacts of social emotion regulation. *Emotion*, 25(2), 330–339. <https://doi.org/10.1037/emo0001284>
- Sahi, R. S., He, Z., Silvers, J. A., & Eisenberger, N. I. (2023). One size does not fit all: Decomposing the implementation and differential benefits of social emotion regulation strategies. *Emotion*, 23(6), 1522–1535. <https://doi.org/10.1037/emo0001194>
- Sahi, R. S., Ninova, E., & Silvers, J. A. (2021). With a little help from my friends: Selective social potentiation of emotion regulation. *Journal of Experimental Psychology: General*, 150(6), 1237–1249. <https://doi.org/10.1037/xge0000853>
- Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64, 240–256. <https://doi.org/10.1016/j.neuroimage.2012.08.052>
- Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, 54(3), 182–203. <https://doi.org/10.1037/0003-066X.54.3.182>
- Schafer, M., & Schiller, D. (2018). The hippocampus and social impairment in psychiatric disorders. *Cold Spring Harbor Symposia on Quantitative Biology* (Vol. 83, pp. 105–118). Cold Spring Harbor Laboratory Press. <https://doi.org/10.1101/sqb.2018.83.037614>
- Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., Ledoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, 463(7277), 49–53. <https://doi.org/10.1038/nature08637>
- Sekeres, M. J., Moscovitch, M., & Winocur, G. (2017). Mechanisms of memory consolidation and transformation. In N. Axmacher & B. Rasch (Eds.), *Cognitive neuroscience of memory consolidation* (pp. 17–44). Springer. [https://doi.org/10.1007/978-3-319-45066-7\\_2](https://doi.org/10.1007/978-3-319-45066-7_2)
- Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, 680, 39–53. <https://doi.org/10.1016/j.neulet.2018.05.006>
- Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage*, 82, 403–415. <https://doi.org/10.1016/j.neuroimage.2013.05.081>
- Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron*, 60(2), 378–389. <https://doi.org/10.1016/j.neuron.2008.09.023>
- Shu, J., Bolger, N., & Ochsner, K. N. (2021). Social emotion regulation strategies are differentially helpful for anxiety and sadness. *Emotion*, 21(6), 1144–1159. <https://doi.org/10.1037/emo0000921>
- Silvers, J. A., & Guassi Moreira, J. F. (2019). Capacity and tendency: A neuroscientific framework for the study of emotion regulation. *Neuroscience Letters*, 693, 35–39. <https://doi.org/10.1016/j.neulet.2017.09.017>
- Smith, A. S., Williams Avram, S. K., Cymerblit-Sabba, A., Song, J., & Young, W. S. (2016). Targeted activation of the hippocampal CA2 area strongly enhances social memory. *Molecular Psychiatry*, 21(8), 1137–1144. <https://doi.org/10.1038/mp.2015.189>
- Speer, M. E., Ibrahim, S., Schiller, D., & Delgado, M. R. (2021). Finding positive meaning in memories of negative events adaptively updates memory. *Nature Communications*, 12(1), Article 6601. <https://doi.org/10.1038/s41467-021-26906-4>
- Spunt, B. (2016). *Easy-optimize-x: Formal release for archiving on Zenodo* (Version 1.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.58616>
- Tabak, B. A., McCullough, M. E., Luna, L. R., Bono, G., & Berry, J. W. (2012). Conciliatory gestures facilitate forgiveness and feelings of friendship by making transgressors appear more agreeable. *Journal of Personality*, 80(2), 503–536. <https://doi.org/10.1111/j.1467-6494.2011.00728.x>
- Takaku, S. (2001). The effects of apology and perspective taking on interpersonal forgiveness: A dissonance-attribution model of interpersonal forgiveness. *The Journal of Social Psychology*, 141(4), 494–508. <https://doi.org/10.1080/00224540109600567>
- Tan, K. M., Daitch, A. L., Pinheiro-Chagas, P., Fox, K. C. R., Parvizi, J., & Lieberman, M. D. (2022). Electroencephalographic evidence of a common neurocognitive sequence for mentalizing about the self and others. *Nature Communications*, 13(1), Article 1919. <https://doi.org/10.1038/s41467-022-29510-2>
- Thorpe, J. N., Gasser, C., Blessing, E., & Davachi, L. (2022). Data-driven clustering of functional signals reveals gradients in processing both within the anterior hippocampus and across its long axis. *The Journal of Neuroscience*, 42(39), 7431–7441. <https://doi.org/10.1523/JNEUROSCI.0269-22.2022>
- Tompary, A., & Davachi, L. (2024). Integration of overlapping sequences emerges with consolidation through medial prefrontal cortex neural ensembles and hippocampal-cortical connectivity. *eLife*, 13, Article e84359. <https://doi.org/10.7554/eLife.84359>
- Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3), 374–391. <https://doi.org/10.1002/hipo.450040319>
- vanOyen Witvliet, C., & Luna, L. R. (2018). Forgiveness and well-being. In D. S. Dunn (Ed.), *Positive psychology: Established and emerging issues* (pp. 131–152). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781315106304-9>
- Vazdarjanova, A., & McGaugh, J. L. (1999). Basolateral amygdala is involved in modulating consolidation of memory for classical fear conditioning. *The Journal of Neuroscience*, 19(15), 6615–6622. <https://doi.org/10.1523/JNEUROSCI.19-15-06615.1999>
- Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, 59(6), 1037–1050. <https://doi.org/10.1016/j.neuron.2008.09.006>
- Webb, J. R., & Toussaint, L. L. (2020). Forgiveness, well-being, and mental health. In E. L. Worthington, Jr. & N. G. Wade (Eds.), *Handbook of forgiveness* (2nd ed., pp. 188–197). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781351123341-18>
- Wu, S., & Meyer, M. L. (2025). *Forgiveness updates interpersonal memories to be less negative*. <https://osf.io/bujms>
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665–670. <https://doi.org/10.1038/nmeth.1635>
- Yu, C., Huang, S., Howard, C. M., Hovhannisyan, M., Clarke, A., Cabeza, R., & Davis, S. W. (2025). Subsequent memory effects in cortical pattern similarity differ by semantic class. *Journal of Cognitive Neuroscience*, 37(1), 155–166. [https://doi.org/10.1162/jocn\\_a\\_02238](https://doi.org/10.1162/jocn_a_02238)
- Zaki, J., & Williams, W. C. (2013). Interpersonal emotion regulation. *Emotion*, 13(5), 803–810. <https://doi.org/10.1037/a0033839>
- Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews*, 41(1), 88–123. [https://doi.org/10.1016/S0165-0173\(02\)00248-5](https://doi.org/10.1016/S0165-0173(02)00248-5)

Received February 20, 2025

Revision received September 23, 2025

Accepted September 25, 2025 ■