Supplementary Materials

The primary manuscript excluded participants who did not believe the targets were real people. Here we report with the full sample (n=37) the same primary analyses reported in the manuscript as statistically significant. For the behavioral analyses, the primary takeaway is that all of the significant results demonstrating preferentially reduced negative affect for the forgiveness vs. look condition reported in the primary paper remain statistically significant in the full sample. For the neuroimaging analyses, we also examined if the primary results – testing for significant changes in neural pattern similarity between manipulation and reconsideration for the forgiven vs. looked at target as a function of affective rating change – remain significant; these analyses remain significant for the left posterior Shen parcellation hippocampus ROI (parcel number 229), marginal for the rDMPFC Mentalizing ROI, and not significant but trending in the same direction for the rDMPFC Shen parcellation ROI (parcel number 10). Overall, these results suggest that being instructed to forgive is sufficient to show behavioral markers of forgiveness (i.e. reduced negative affect) and possibly change the episodic details of a memory to match those considered during forgiveness. However, updating representations in the rDMPFC, a region implicated in mentalizing about other people's thoughts, feelings, and intentions, may require the forgiver to believe the person they are forgiving is authentic. For the interested reader, the full details of the relevant results from the full sample are below.

Changes in Affect Ratings in the Forgiving and Looking Conditions

When examining the full sample, the ratings made in the scanner yielded the same pattern of results. The affective ratings made during fMRI scans indicated that the forgiveness condition was more effective than the look condition in attenuating negative affect. A 2 (target: forgiven vs. looked at) × 3 (experimental phase: encoding, manipulation, reconsideration) repeated-measures ANOVA indicated that the interaction for the negative image ratings was significant ($F(1.73, 62.18) = 8.55$, $p < .001$, $\eta_g^2 = 0.01$). Within the forgiven target condition, we found that forgiving during Day 1 manipulation reduced negativity ratings relative to encoding ($t(36) = 5.27$, $p < .001$). Moreover, the ratings of negative images on Day 2 reconsideration were still less negative than during encoding on Day 1 ($t(36) = 4.09$, $p < .001$). reconsideration ratings were also more negative than forgiveness manipulation ratings ($t(36) = -2.80$, $p = .008$), suggesting that while the negativity of the memory trace remains attenuated on Day 2, it is not as attenuated as during the moment of forgiveness on Day 1.

Ratings for the looked at target were also less negative during the look manipulation vs. encoding on Day 1 ($t(36) = 3.05$, $p = .004$), less negative on Day 2 reconsideration vs. Day 1 encoding ($t(36) = 2.31$, $p = .027$), but showed no difference in negativity on Day 2 reconsideration vs. Day 1 manipulation ($t(36) = -0.94$, $p = .354$).

We next again 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. It is noteworthy that on Day 1 encoding, negative trials paired with the forgiven target were still rated as less negative than those paired with the looked at target ($t(36)=4.74$, $p < .001$). We therefore tested whether differences in affect ratings for the forgiven vs. 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 (β = -0.22, t(4301.10) = -4.85, p < .001). Additionally, the negative trials shown with the forgiven target were rated as less negative than the trials shown with the looked at target during reconsideration, while controlling for baseline differences during encoding (β = -0.12, t(4305.06) = -2.57, p = .010).

With respect to manipulation check analyses, for the forgiven target, less negative ratings of the IAPs images during forgiveness again correlated with greater liking of the forgiven target (r(35) = 0.47, p = .002, one-tailed). However, unlike the primary results, negative IAPS image ratings did not correlate with less revengeful image selections (r(35) = 0.13, p = 0.218, one-tailed).

Changes in Neural Pattern Similarity for Forgiven vs. Looked at Targets

Next we assessed whether the statistically significant neural pattern similarity results reported in the primary manuscript persisted in the full sample. The multivariate analyses testing whether the neural pattern from the moment of forgiveness becomes incorporated into forgiven experiences to make them less negative showed that one of our targeted ROIs demonstrated our prediction in the full sample. The posterior left hippocampus from the Shen parcellation (parcel number 229) showed a significant change in neural pattern similarity between manipulation and reconsideration for the forgiven vs. looked at target as a function of affective rating change (β = -0.05, t(2166.06) = -3.44, p < .001, 95% CI [-0.080, -0.021]). The same analysis was statistically marginal for the rDMPFC Neurosynth mentalizing ROI (β = -0.03, t(2163.84) = -1.85, p = .064, 95% CI [-0.068, 0.004]). For the rDMPFC Shen ROI, this analysis was not statistically significant though it showed the same overall pattern: β = -0.03, t(2160.40) = -1.61, p = .106, 95% CI [-0.063, 0.005]).

The follow-up simple slope analysis for the posterior left hippocampus showed that the forgiveness and look slopes were different when affective rating changes were relatively low (change value of -3: β = 0.13, t(2165.22) = 2.94, p = .003; change value of -2: β = 0.08, t(2164.08) = 2.63, p = .009; change value of -1: β = 0.03, t(2158.48) = 1.65, p = .098), equivalent (change value of 0: β = -0.023, t(2137.55) = -2.20, p = .028), and high (change value of 1: β = -0.07, t(2160.07) = -3.85, p < .001; change value of 2: β = -0.12, t(2164.21) = -3.83, p < .001).