Emotion
2019, Vol. 19, No. 1, 132–145

1528-3542/19/$12.00 http://dx.doi.org/10.1037/emo0000430

People are Better at Maintaining Positive Than Negative Emotional States

Christian E. Waugh, Kristin E. Running,
and Olivia C. Reynolds
Wake Forest University

Ian H. Gotlib
Stanford University

Determining how people maintain positive and negative emotional states is critical to understanding emotional dynamics, individual differences in emotion, and the instrumental value of emotions. There has been a surge in interest in tasks assessing affective working memory that can examine how people maintain stimulus-independent positive and negative emotional states. In these tasks, people are asked to maintain their emotional state that was induced by an initial stimulus in order to compare that state with the state induced by a subsequent stimulus. It is unclear, however, whether measures of accuracy in this task actually reflect the success of maintaining the initial emotional state. In a series of studies, we introduce an idiographic metric of accuracy that reflects the success of emotional maintenance and use that metric to examine whether people are better at maintaining positive or negative emotional states. We demonstrate that people are generally better at maintaining positive emotional states than they are at maintaining negative emotional states (Studies 1–3). We also show that this effect is not due to decay or to spontaneous interference processes (Studies 2–3), retroactive interference processes (Studies 4–5), or reduced engagement with the initial emotional state (Study 5). Although the mechanism underlying this effect is not yet clear, our results have important implications for understanding emotional maintenance and the possible functions of positive and negative emotions.

Keywords: maintenance, working memory, positive emotion, negative emotion, dynamics

Supplemental materials: http://dx.doi.org/10.1037/emo0000430.supp

Emotions are part of the dynamic interplay between people and their environments; consequently, the duration and timing of emotions are often due to the duration and timing of the eliciting stimuli (Verduyn, Delvaux, Van Coillie, Tuerlinckx, & Van Mechelen, 2009). Emotional dynamics are not bounded inexorably by the dynamics of the stimuli, however; indeed, they often endure long after a stimulus has passed. Researchers have found that people strategically maintain emotional states to enhance performance on subsequent emotion-congruent tasks (Tamir, Mitchell, & Gross, 2008) and that individuals diagnosed with depression are characterized by persistent negative emotional states (Gilboa & Gotlib, 1997). Thus, maintaining an emotional state, either automatically or strategically, is an important aspect of both normal and psychopathological emotional functioning.

Like the verbal and visual material that investigators have demonstrated can be maintained in working memory, emotions also convey information (Schwarz & Clore, 1983) that can be maintained in the service of operating on it at a later point (Mikels, Reuter-Lorenz, Beyer, & Fredrickson, 2008). For example, Mikels et al. (2008) demonstrated that people can hold their emotional response to an image in mind in order to compare it to their response to a subsequent image. Mikels et al. also showed that emotional working memory involves a different system than does cognitive working memory—intervening cognitive distracters did not interfere with the maintenance of emotional information. Further supporting the independence of cognitive and emotional working memory, investigators have found that people with severe anterograde amnesia are able to maintain emotional states beyond their episodic memory of the eliciting stimulus (Feinstein, Duff, & Tranel, 2010).

In the same way that assessing performance on cognitive working memory tasks has proven critical for understanding the cognitive mechanisms underlying working memory (Baddeley, 2012), individual differences in executive functioning (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010), and working memory deficits in psychopathology (Joormann, Levens, & Gotlib, 2011), it is likely that assessing performance on emotional working memory tasks will be important in increasing our understanding of affective working memory. Indeed, Levens and Gotlib (2010) found that depressed individuals were slower than were their nondepressed counterparts at disengaging from sad faces on a

This article was published Online First March 22, 2018.
Christian E. Waugh, Kristin E. Running, and Olivia C. Reynolds, Department of Psychology, Wake Forest University; Ian H. Gotlib, Department of Psychology, Stanford University.

Kristin E. Running is now at the Department of Educational Psychology, University of Minnesota.

This research was supported by Grants MH 59259 and MH74849 from the National Institute of Mental Health to Ian H. Gotlib. We thank Maria Lemus, Arkadiy Maksimovskiy, and Henry Valk for their help in recruiting and running the participants in this study and for creating stimuli and preprocessing the data. We also thank Joseph Mikels for discussions about the design of this project.

Correspondence concerning this article should be addressed to Christian E. Waugh, Department of Psychology, Wake Forest University, Winston Salem, NC 27109. E-mail: waughce@wfu.edu

2-back task. In addition, Gard et al. (2011) found that participants with schizophrenia exhibited less accurate affective working memory than did participants without schizophrenia.

Although studies of affective working memory to date are promising, they have not yet assessed *the ability to maintain a stimulus-independent emotional state.* Most studies of affective working memory involve holding in mind verbal or visual information that is emotionally charged (e.g., negative words; Kensinger & Corkin, 2003; angry facial expressions; Jackson, Linden, & Raymond, 2014); however, performance is always operationalized as the ability to hold in mind the verbal or visual information, not the emotional state itself. Thus, these studies focus on how emotion can influence cognitive working memory and not on working memory for stimulus-independent emotional states. Indeed, this prompted Mikels et al. (2008) to develop their stimulus-independent emotional state task in which they compared emotional states evoked from a first image with the emotional states evoked by a second image. The metric of performance these investigators and others using this task have employed is normative concordance, that is, the rate at which participants' selection of which emotional state was higher in intensity matched the "correct selection" according to previously normed intensity values of the images. In this case, normative concordance is only a rough performance metric because it is not clear whether errors are due to the failure to maintain the emotional state or to a difference between a particular individual's responses to this emotional image and the normed responses.

To provide a better assessment of affective working memory performance, researchers have recently calculated an idiographic accuracy metric (Gard et al., 2011; Waugh, Lemus, & Gotlib, 2014). In these studies participants provide ratings of their emotional responses to images independent of the maintenance task, and these ratings are then used to determine the "correct" selection of which emotional state should have been selected as the more intense during the maintenance task. Although this type of metric has advantages over the normative concordance measures, there are nevertheless problems with using it as a measure of affective working memory performance. Unlike other delayed match-to-sample tasks, in this affective working memory task participants have to generate responses to both the target and probe stimuli to subsequently be able to compare them. This leaves open the possibility that errors on this task are not due only to improperly generating/maintaining the initial emotional state, but also to some transformation happening during the generation of the second emotional state. For example, it is not yet clear if the act of maintaining an emotional state changes subsequent responses to emotional stimuli.

In the current studies, we attempt to address this problem. First, in Studies 1–3, in addition to computing an overall idiographic accuracy score, we follow Waugh et al. (2014) and also calculate separate idiographic accuracy scores for trials in which the correct selection of the more intense emotional state (idiographically) was supposed to be the first emotional state, and for trials in which the correct selection was supposed to be the second emotional state. More errors when the first emotional state should have been selected could reflect either the failure to maintain the intensity of the initial emotional state and/or some augmentation of the second emotional state. Alternatively, more errors when the second emotional state should have been selected could reflect either an elaboration of the initial emotional state or the attenuation of the second emotional state. Second, in Studies 2–3, we vary the duration of the maintenance period. An increase in errors from short to long delays may reflect decay/interference processes that affect the ability to maintain the first emotional state. On the other hand, a decrease in errors from short to long delays may be due to the first emotional state having a greater impact on the second emotional state. Third, in Studies 4 and 5, participants maintain their first and second emotional responses in the service of rating them (not comparing them to each other), with proper safeguards in place to limit verbal and spatial rehearsal. In these studies, the second images of each are either emotional or neutral, allowing us to test whether there is retroactive interference of experiencing the second emotional state on the maintenance of the first emotional state.

The primary goal of these studies is to use these novel performance metrics to test whether participants exhibit different affective working memory performance for negative emotional states than for positive emotional states. Previous studies that have attempted to answer this question have typically compared working memory for positive and negative emotional verbal or visual information. The findings from these studies are mixed; whereas some studies show enhanced working memory for negatively valenced information (e.g., anger faces in a delayed match-to-sample task; Jackson et al., 2014), other investigations show enhanced working memory for positively valenced information (e.g., happy faces in a 2-back task; Levens & Gotlib, 2010; and positive images in a delayed match-to-sample; Perlstein, Elbert, & Stenger, 2002), and still other studies show no enhancement in working memory for either positive or negative information (Kensinger & Corkin, 2003). As we noted above, however, these studies do not assess stimulus-independent affective working memory; moreover, the studies that do (Gard et al., 2011; Mikels et al., 2008) have not determined whether or not differences in positive and negative affective working memory were due to the generation/maintenance of the initial emotional state.

In a recent functional MRI study, we adapted Mikels et al.'s (2008) emotional working memory task to examine the neural correlates underlying successful maintenance of emotion (Waugh et al., 2014). In that study, we separately calculated the idiographic accuracy for when the first or the second emotional state "should" have been selected as the more intense state. We found that participants showed poorer idiographic accuracy when the first negative emotional state should have been selected than when the second negative state should have been selected, and equivalent idiographic accuracy across first and second positive emotional states. This study provided preliminary evidence that it may be more difficult to maintain the intensity of a negative than of a positive emotional state. We found additional evidence of this formulation in the neuroimaging results: when maintaining positive but not negative emotional states, participants exhibited greater duration of activation in the rostral medial frontal cortex (rMFC), a region that tracks the duration of emotional events (Lindquist, Waugh, & Wager, 2007; Waugh, Hamilton, Chen, Joormann, & Gotlib, 2012; Waugh, Hamilton, & Gotlib, 2010).

In the current studies, we used the strategies outlined above in order to yield more accurate estimates of affective working memory performance to test the formulation that people are better at maintaining positive than negative emotional states.

Study 1

The primary purpose of Study 1 was to replicate Waugh et al.'s (2014) study outside of the MRI scanning environment. To induce emotional maintenance, participants maintained the intensity of their emotional response to one image during a delay period and compared it with the intensity of their emotional response to a subsequent image. There was also a "nonmaintenance" of emotion condition in which participants saw both images, but only rated the second image. In addition, we assessed various psychophysiological responses throughout the maintenance task, such as facial electromyography, skin conductance, and the startle response. This nonmaintenance condition and psychophysiology responses were originally designed to allow us to separate mechanisms involved specifically with the maintenance of emotion from those involved in emotional responses more generally (Waugh et al., 2014); however, because this study focused on affective working memory performance, the findings from this condition and psychophysiological responses are not relevant to this article and are not reported here (see supplementary materials for information about the psychophysiological data). If this study replicates Waugh et al., then we should find that participants exhibit poorer idiographic accuracy when trying to maintain the initial negative emotional state than when trying to maintain the second negative emotional state; in contrast, we should not see this accuracy difference for positive emotion maintenance trials.

Method

Participants. Forty participants (20 female) were recruited from the San Francisco Bay Area community through postings on Internet classifieds (e.g., craigslist.org) to participate in a study on information processing. To be eligible, participants had to be between the ages of 18 and 55 years (M = 37.03 years, SD = 13.77) and must not have had any current or past cardiovascular health problems. This sample size was chosen to be roughly 150% of the sample size from Waugh et al.'s (2014) study to allow for possible inflation of the original effect size.

Emotional maintenance task. Each trial consisted of five parts: first picture (2 s), one of two instruction slides (1 s), a black screen that served either as a forgetting or retention interval (8 s), second picture (2 s), and a screen prompting the participant to enter his or her response. All of the pictures were taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997). For each positive and negative emotion trial, participants were given one of two instructions after viewing the first picture. In the nonmaintain condition, the participants' task was to rate the intensity of the second picture; they did not rate the first picture, although they were not explicitly instructed to forget it. The instruction slide had one square (representing the second picture) and a question mark. During the rating slide, participants reported the intensity of their emotional response to that second picture as *low* (1) or *high* (2). In the maintain condition, participants' task was to compare the intensity of their emotional response to the first picture to the intensity of their emotional response to the second picture. The instruction slide had two squares (representing the two pictures) and a question mark between the squares. During the rating slide, participants reported the intensity of their emotional response as being *lower* (1) or *higher* (2) during the second picture than during the first picture. For neutral trials, participants received only the nonmaintain instructions.

Participants received 20 trials of each type: maintain-negative, maintain-positive, nonmaintain-negative, nonmaintain-positive, and nonmaintain-neutral, for a total of 100 trials separated into two blocks with a 1-min rest between blocks. Because this task was originally designed for a functional MRI experiment, for each trial type there were four "catch" trials in which participants did not see the second picture and did not make any ratings, but instead saw a gray screen. To ensure there were enough trials in which each emotional state (first or second) was more intense than the other, for half of the emotion trials the first picture was normed as being more emotionally intense than the second picture, and vice versa for the other half of the emotion trials. The pictures were selected by first omitting IAPS erotic pictures and then equating the four emotional trial types on normed ratings of intensity of emotional valence (intensity for positive pictures was reverse-coded for comparability): maintain-positive = 3.65, nonmaintain-positive = 3.68, maintain-negative = 3.64, nonmaintain-negative = 3.60, $F(1, 76) = .15$, $p = .699$, and then including less intense neutral pictures: nonmaintain-neutral = 4.22, $F(1, 98) = 50.23$, $p < .001$ (vs. four other conditions). On half of the trials within each of the emotion conditions the second image was normed as more intense ($M = 4.12$, $SD = .47$) than the first image ($M = 3.16$, $SD = .43$), $t(39) = 14.86$, $p < .001$, and on the other half, the second image was normed as less intense ($M = 3.13$, $SD = .41$) than the first image ($M = 4.16$, $SD = .43$), $t(39) = 12.37$, $p < .001$.

Startle probes. On a randomly selected 80% of the trials (16 of the 20 trials for each trial type), participants heard a startle probe through the headphones that consisted of instantaneous-rise 50-ms bursts of white noise at 95 dB (Blumenthal et al., 2005). For half of these startle trials, the "response" startle probes occurred 1.5 s into the first picture of each pair. For the other half of these trials, the "delay period" startle probes occurred 4 s into the interstimulus interval between the first and second pictures. We present this information because it is critical for understanding the participants' experience; however, we do not present findings from the startle probes (see online supplemental materials).

Posttask ratings. After the emotional maintenance task, participants viewed the pictures again. In one viewing block, participants rated the emotional intensity ("How intense was your emotional response to this image?") of all the pictures on a visual analogue scale that stretched the width of the monitor (from 0 to 1,024 pixels) anchored by *low emotional intensity* on the left, *moderate emotional intensity* in the middle, and *high emotional intensity* on the right. In a separate block, participants rated the self-relevance ("How much did you personally associate with this image?") of all the pictures on the same visual analogue scale (except that the anchors specified low, moderate, and high "personal association"). These self-relevance ratings are not presented here. The order of the rating blocks was counterbalanced.

Procedure. Participants were run individually. After the participant signed the informed-consent forms, the experimenter attached the sensors and placed the headphones on the participant. The participant completed filler questionnaires during a 10-min acclimation period. Next, the participants rested quietly for a 5-min baseline period. The experimenter then explained the emotional maintenance task to the participant, introduced the startle probes, and administered five example startle probes. The participant was instructed to ignore the startle probes throughout the task.

After the emotional maintenance task, the participant completed the ratings task. After she or he completed the ratings task and any remaining questionnaires, the participant was debriefed, paid ($25/hr), and thanked for her or his participation. This procedure was approved by the Stanford University Institutional Review Board.

Affective working memory performance. To assess how well participants can maintain their emotional states, we calculated idiographic metrics of performance (Gard et al., 2011; Mikels et al., 2008; Waugh et al., 2014). For the overall idiographic accuracy metric, we calculated the percentage of maintain trials for which participants' responses (e.g., "higher") when comparing the emotional intensity of the two pictures matched the differences between the pictures in their posttask ratings (e.g., the second picture was rated as eliciting a more intense emotional response than the first). Trials were only counted if participants' two emotional responses were more than 100 pixels apart on the scale that was 1,024 pixels in length (ratings that were closer than 100 pixels were considered "equivalent" and the trials were omitted from the analyses). This also allowed us to match more closely the subsequent studies in which we use a 1–10 scale. Of note, excluding the "equivalent" responses significantly improved idiographic accuracy from .70 to .75, $t(39) = 5.48$, $p < .001$.

We also calculated separately the idiographic accuracy for the trials in which participants' posttask ratings indicated that they *should* have selected the first picture as being more intense than the second (1gtr2) and for the trials in which their posttask ratings indicated that they should have selected the second picture as more intense than the first (2gtr1). One participant was dropped from this analysis due to not having at least 1 trial that could be included when separated by order.

Results

First, we examined participants' overall idiographic accuracy of their choice of which picture in each pair elicited a more intense emotional response. Participants exhibited high idiographic (M = .745, SE = .023) accuracy, which was greater than can be expected by chance (.5), $t(39) = 10.97$, $p < .001$. Participants exhibited equivalent idiographic accuracy when comparing negative emotional states (M = .744, SE = .028) and when comparing positive emotional states (M = .745, SE = .028), $t(39) = .32$, $p = .748$.

Next, we examined the idiographic accuracy separately for 1gtr2 trials and 2gtr1 trials. We conducted a 2 (Valence: positive, negative) × 2 (Order: 1gtr2, 2gtr1) repeated-measures analysis of variance (ANOVA) on the idiographic accuracy. The only significant effect was a main effect of Order, $F(1, 38) = 20.51$, $p < .001$; there was not a significant effect of Valence, $F(1, 38) = .49$, $p = .49$, or a significant interaction of Valence × Order, $F(1, 38) = 1.4$, $p = .245$ (see Figure 1). When comparing negative pictures, participants exhibited greater accuracy when the second picture should have been selected as the more intense picture (M = .825, SE = .035) than when the first picture should have been selected as the more intense picture (M = .632, SE = .043), $t(38)$, = 4.02, $p < .001$. There was a similar, but only marginally significant, pattern when participants compared positive pictures: they exhibited greater accuracy when the second picture should have been selected as the more intense picture (M = .803, SE = .040) than when the first picture should have been selected (M = .700, SE = .041), $t(38) = 1.98$, $p = .055$.

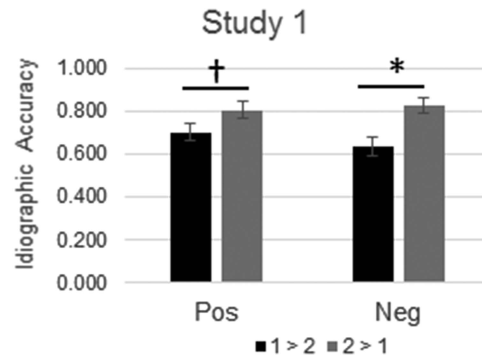

Figure 1. Study 1 idiographic accuracy for positive (Pos) and negative (Neg) trials in which the first picture was rated (during the pretask) as eliciting a more intense emotional response (1 > 2) and trials in which the second picture was rated as eliciting a more intense emotional response (2 > 1). Error bars are standard error of the mean. * $p < .05$. † $p < .1$.

Discussion

In Study 1, we examined idiographic accuracy separately for when the first or second emotional states should have been selected as the most intense. Across both positive and negative images, participants were more accurate when the second emotional state should have been selected as more intense than when the first emotional state should have been selected as more intense. This provides initial evidence that participants may find it difficult to hold in mind the initial emotional state, perhaps causing it to decrease in intensity over the delay period so that it is then judged as being less intense when comparing it to the recent experience of the second emotional state. Notably, there was not a significant difference between idiographic accuracy for negative and positive emotional states, which does not replicate Waugh et al.'s (2014) study. Therefore, we conducted subsequent studies to determine which of these two findings is replicable.

Study 2

In Study 2, we again had participants perform the affective maintenance task with both maintain and nonmaintain conditions (to be consistent with previous versions of this task). In this study, however, we varied the delay period. In studies of cognitive working memory, individuals tend to perform better when holding in mind information over short delays than over long delays (Barrouillet & Camos, 2012), presumably due to a number of processes, such as decay (Barrouillet & Camos, 2012) and interference (Lewandowsky & Oberauer, 2009). In conjunction with the separate idiographic accuracy metrics for the first and second emotional states, a delay period may elucidate whether this relatively lower accuracy is due to decay and/or to nontask interference processes (such as spontaneously engaging in other emotion-related processes), which would be reflected by an increase in errors from short to long delays.

Method

Participants. Seventy-three participants (38 female) were recruited from Amazon's Mechanical Turk website to participate in a study on affective maintenance. To be eligible, participants had to be over the age of 18 ($M = 37.58$ years, $SD = 12.30$), speak fluent English, and reside in the United States. We also required the Mechanical Turk users to have at least a 95% completion and approval rating on their previously completed Mechanical Turk tasks to ensure we recruited engaged participants (Buhrmester, Kwang, & Gosling, 2011). Three participants were excluded from data analyses for lack of response diversity (they reported the same emotional intensity for all the pictures in at least one trial type) leaving 70 participants. This sample size was originally selected so as to achieve 90% power to detect an effect size of .41 from Study 1, but that original effect size was from an analysis not included here. Fortunately, using the new effect size of .61 (difference between acc1gtr2 and acc2gtr1 for negative emotion trials) allowed us to have over 99% power at $\alpha = .05$ (two-tailed).

Emotional maintenance task. The maintenance task was similar to that in Study 1 except that each trial had either a short delay (2 s) or a long delay (10 s) between the instruction screen and the second image. There were also some additional small changes. We changed the instruction screen to RATE NEXT for the nonmaintenance trials and COMPARE for the maintenance trials, there were no neutral trials, and there were no startle probes. Participants received 10 trials of each type: maintain-negative with short delay, maintain-negative with long delay, maintain-positive with short delay, maintain-positive with long delay, nonmaintain-negative with short delay, nonmaintain-negative with long delay, nonmaintain-positive with short delay, and nonmaintain-positive with long delay, for a total of 80 trials. We used the image set from Study 1, but reassigned images within each original trial type (e.g., maintain-negative) to the short delay versus long delay sub trial types. We also added images to reach 10 trials of each type. Again, we ensured that there was an equivalent number of trials within each trial type in which the first image was normed to be of greater valence than the second image and vice versa. All trial types had images of equivalent normed intensity, all ts $< .82$.

Posttask ratings. As in Study 1, after the emotional maintenance task, participants viewed the pictures again and rated the emotional intensity and personal relevance (not presented here). In this study, however, they rated them on a scale of 1 (*low emotional intensity*) to 10 (*high emotional intensity*).

Procedure. Participants completed the task online through their Mechanical Turk accounts. After the participants electronically signed the informed-consent forms, they then completed a short demographics measure about their age, race, ethnicity, gender, education level, and family income level. They then moved on to the emotional maintenance task. After completing the emotional-maintenance task, participants completed the posttask ratings of the pictures. After participants completed these ratings, they were debriefed and thanked for their participation. They received a compensation code to be entered into the Mechanical Turk website. The first 48 participants were paid $1.50 through Mechanical Turk for completing this survey in the allotted 60 min. We decided to increase the amount of time participants could take to complete the survey after the first 48 participants; therefore, the remaining 25 participants were paid $3.00 for completing the survey within 90 min. The Wake Forest University Institutional Review Board approved this procedure.

Behavioral performance. To assess how well participants can maintain their emotional states, we calculated the same metrics of idiographic accuracy as we did in Study 1. Pairs of images with ratings that were the same were considered "too close to call" and omitted. Six participants were dropped from the acc1gtr2 and acc2gtr1 idiographic accuracy analyses due to not having at least 1 trial that could be included when separated by order.

Results

First, we examined participants' overall idiographic accuracy of their choice of which picture in each pair elicited a more intense emotional response. Participants exhibited high idiographic accuracy ($M = .741$, $SE = .013$), which was greater than would be expected by chance (.5), $t(63) = 18.80$, $p < .001$. A 2 (Valence: positive, negative) $\times$ 2 (Delay: short, long) repeated-measures ANOVA conducted on idiographic accuracy yielded no significant main effects or interactions, Fs < 1.87, ps $> .177$.

Next, we examined the idiographic accuracy separately for 1gtr2 trials and 2gtr1 trials. We conducted a 2 (Valence: positive, negative) $\times$ 2 (Delay: short, long) $\times$ 2 (Order: 1gtr2, 2gtr1) repeated-measures ANOVA on the idiographic accuracy. The main effects of Valence, $F(1, 63) = .22$, $p = .643$, Order, $F(1, 63) = .25$, $p = .619$, and Delay, $F(1, 63) = 3.30$, $p = .074$, were not significant. There was, however, a significant interaction of Valence $\times$ Order, $F(1, 63) = 24.83$, $p < .001$ (see Figure 2). Consistent with Study 1, when comparing negative emotional states, participants exhibited greater accuracy when the second state should have been selected as more intense ($M = .827$, $SE = .021$) than when the first state should have been selected as more intense ($M = .696$, $SE = .029$), $t(63)$, $= 2.72$, $p = .008$. Participants exhibited the opposite pattern when comparing positive pictures: they exhibited greater accuracy when the first picture should have been selected as the more intense picture ($M = .806$, $SE = .022$) than when the second picture should have been selected ($M = .697$, $SE = .028$), $t(63) = 2.27$, $p = .027$. In addition, accuracy for acc1gtr2 trials for positive emotional states was greater than those for negative emotional states, $t(63) = 3.06$, $p = 003$, and vice versa for acc2gtr1 trials, $t(63) = 4.64$, $p < .001$. The interactions of Valence $\times$ Delay, F(1,

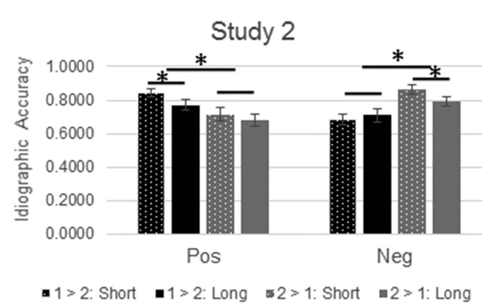

Figure 2. Study 2 idiographic accuracy for positive (Pos) and negative (Neg) trials in which the first picture was rated (during the pretask) as eliciting a more intense emotional response (1 > 2) and trials in which the second picture was rated as eliciting a more intense emotional response (2 > 1). Short and long refer to trials on which there was a short and long delay period, respectively. Error bars are standard error of the mean. * $p < .05$.

63) = .54, p = .463, Delay × Order, F(1, 63) = .75, p = .39, and Valence × Delay × Order, F(1, 63) = 3.22, p = .078, were not significant.

To test our specific hypotheses regarding the effect of the delay period on accuracy when participants should have selected the first emotional state as more intense, we conducted pairwise t tests on 1gtr2 trials comparing idiographic accuracy for short and long delay periods. When maintaining the first negative emotional state that should have been indicated as being more intense, participants were comparably accurate on short (M = .684, SE = .033) and long (M = .709, SE = .036) trials, t(63) = .62, p = .536. When maintaining the initial positive emotional states, however, participants were more accurate on short trials (M = .840, SE = .024) than they were on long trials (M = .772, SE = .030), t(63) = 2.22, p = .03.

Discussion

Consistent with Study 1 and with Waugh et al.'s (2014) findings, participants again exhibited poor performance when trying to maintain an initial negative emotional state that should have been more intense than the second emotional state. This finding provides further evidence that participants have difficulty maintaining negative emotional states after the offset of the eliciting stimulus. Notably, this poor performance was not influenced by the length of the delay period, suggesting that it was not due to decay or spontaneous interference but rather to some other process. We test possible alternative processes in Study 4.

When maintaining positive emotional states, participants exhibited the opposite pattern, exhibiting greater idiographic accuracy on trials in which the initial states should have been more intense than on trials in which the second positive emotional state should have been more intense. One must use caution, however, when interpreting these findings given that they are inconsistent with both Waugh et al. (2014), who showed equivalent accuracy for first and second emotional states, and Study 1, which showed marginally greater accuracy when the second emotional state was more intense. Thus, we have not yet found a reliable pattern of positive emotional working memory accuracy. Notably, participants exhibited greater accuracy for the first positive emotional state when the delay period was short than when it was long, suggesting that, unlike negative emotion maintenance, decay or spontaneous interference processes are operating when individuals are maintaining positive emotional states.

Study 3

The primary purpose of Study 3 was to replicate the findings from Study 2 concerning negative emotional working memory with delay and to investigate further the seemingly mixed findings regarding positive emotional working memory. In addition, one significant change in this study was that participants provided their emotional ratings of the images before the maintenance task instead of after the task. We made this change to account for the possibility in the previous studies that maintaining (or not) emotional states induced by these images may differentially change participants' emotional responses when rating those images again later. If that were to occur, then those ratings would be less useful as reference points for determining the idiographic accuracy during the maintenance task.

Method

Participants. Sixty-seven participants (27 female) were recruited from Amazon's Mechanical Turk website to participate in a study on affective maintenance. To be eligible, participants had to be over the age of 18 (M = 35.73 years, SD = 11.46), speak fluent English, and reside in the United States. We also required the Mechanical Turk users to have at least a 95% completion and approval rating on their previously completed Mechanical Turk tasks to ensure we recruited engaged participants. This sample size was determined in the same way that the sample size from Study 2 was determined. Four participants were excluded from overall idiographic accuracy data analysis (n = 63) and an additional six participants were excluded from the idiographic accuracy by order analysis (n = 57) due to not having at least one trial in a given trial type that could be included to calculate accuracy.

Procedures. The tasks, questionnaires and procedures were the same as in Study 2 with two notable exceptions. First, we omitted the personal association ratings because they were not relevant to our study, and second, participants rated their emotional responses to the pictures before the emotional maintenance task rather than after the task. All participants were paid $3.00 through Mechanical Turk for completing this survey.

Results

First, we examined participants' idiographic accuracy of their choice of which picture in each pair elicited a more intense emotional response. Overall, participants exhibited idiographic accuracy (M = .654, SE = .013) that was greater than would be expected by chance (.5), t(56) = 11.63, $p < .001$. A 2 (Valence: positive, negative) × 2 (Delay: short, long) repeated-measures ANOVA conducted on idiographic accuracy yielded a significant main effect of Delay, F(1, 56) = 20.64, $p < .001$, such that participants exhibited more idiographic accuracy when comparing pictures over a short delay (M = .730, SE = .018) than over a long delay (M = .632, SE = .018); neither the main effect of Valence, F(1, 56) = .46, p = .499, nor the interaction of Valence × Delay, F(1, 56) = .032, p = .859, was significant.

We then conducted a 2 (Valence: positive, negative) × 2 (Delay: short, long) × 2 (Order: 1gtr2, 2gtr1) repeated-measures ANOVA on the idiographic accuracy. There was a significant main effect of Delay, F(1, 56) = 11.17, p = .001, similar to when we examined overall idiographic accuracy. As in Study 2, there was a significant interaction of Valence × Order, F(1, 56) = 11.68, p = .001 (see Figure 3). When comparing negative pictures, participants exhibited greater accuracy when the second picture should have been selected as the more intense picture (M = .768, SE = .026) than when the first picture should have been selected as the more intense picture (M = .637, SE = .026), t(56), = 3.97, $p < .001$. Alternatively, when comparing positive pictures participants exhibited the opposite pattern (1gtr2: M = .719, SE = .024; 2gtr1: M = .672, SE = .034) although it was not significant as it was in Study 2, t(56) = 1.07, p = .285. In addition, accuracy for 1gtr2 trials for positive emotional states was greater than those for negative emotional states, t(56) = 2.13, p = 038, and vice versa for 2gtr1 trials, t(56) = 3.00, p = .004. There were not significant interactions of Valence × Delay, F(1, 56) = .005, p = .944, Delay × Order, F(1, 56) = .12, p = .727, or of Valence × Delay × Order, F(1, 56) = .11, p = .746.

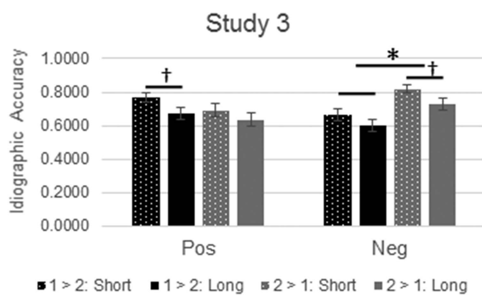

Figure 3. Study 3 idiographic accuracy for positive (Pos) and negative (Neg) trials in which the first picture was rated (during the pretask) as eliciting a more intense emotional response ($1 > 2$) and trials in which the second picture was rated as eliciting a more intense emotional response ($2 > 1$). Short and long refer to trials on which there was a short and long delay period, respectively. Error bars are standard error of the mean. * $p < .05$. † $p < .1$.

To test our specific hypotheses regarding the effect of the delay period on accuracy when participants should have selected the first emotional state as more intense, we conducted pairwise *t* tests on 1gtr2 trials comparing idiographic accuracy for short and long delay periods. When maintaining the first negative emotional state that should have been indicated as being more intense, participants were again comparably accurate on short ($M = .673$, $SE = .034$) and long ($M = .602$, $SE = .034$) trials, $t(56) = 1.62$, $p = .110$. When maintaining positive emotional states, however, participants were marginally more accurate on short trials ($M = .763$, $SE = .032$) than they were on long trials ($M = .673$, $SE = .034$), $t(56) = 1.95$, $p = .056$.

Discussion

The results from Study 3 largely replicate the findings from Study 2 to show that people are generally less effective at maintaining their initial negative emotional state than they are at maintaining their initial positive emotional state. Notably, this finding was supported even though we had participants independently rate the emotional images before the maintenance task instead of after the task as in the previous studies, which further corroborates our use of these ratings in determining the accuracy of affective working memory. Again, however, it is not clear what mechanism is responsible for this relatively poorer working memory accuracy for the initial negative emotional state given that accuracy was equivalent across both short and long delay trials. We test another candidate mechanism in Study 4.

Interim Meta-Analysis

At this point, our lab had conducted four studies (Current Studies 1–3 and Waugh et al., 2014) using the affective maintenance paradigm in which participants hold in mind the emotional state induced by one image to compare it to the emotional state induced by a second image. To collate these findings, we conducted an internal meta-analysis. We computed effect sizes (*d*) of idiographic accuracy for (a) each pairwise comparison within valence (negative, positive) for 2gtr1 trials versus 1gtr2 trials and (b) each pairwise comparison across valence (negative vs. positive) for 1gtr2 and 2gtr1 trials separately. Effect sizes and variances were calculated with the correction for correlated designs (Dunlap, Cortina, Vaslow, & Burke, 1996) and entered into the R program metafor (Viechtbauer, 2010). We specified random effects models using the Hunter–Schmidt method (Hunter & Schmidt, 2004) along with a Knapp and Hartung adjustment of the standard error (Knapp & Hartung, 2003).

Results and Discussion

Consistent with the pattern of findings thus far, Valence (negative, positive) moderated the difference in accuracy rates for 1gtr2 versus 2gtr1 trials, $d = -.894$, $SE = .211$, $z = -4.229$, $p = .005$, 95% confidence interval (CI) [−1.411, −.377] (see Figure 4). When maintaining negative emotional states, participants exhibited significantly greater accuracy on 2gtr1 trials than on 1gtr2 trials, $d = .706$, $SE = .196$, $p < .001$, 95% CI [.547, .865]. Alternatively, when maintaining positive emotional states there was no difference in accuracy for 2gtr1 trials and 1gtr2 trials, $d = -.165$, $SE = .203$, $p = .476$, 95% CI [−.812, .482]. Comparing valence within accuracy types revealed that on 1gtr2 trials, negative emotional accuracy was significantly poorer than positive emotional accuracy, $d = -.443$, $SE = .071$, $p = .008$, 95% CI [−.670, −.216]. Alternatively, on 2gtr1 trials negative emotional accuracy was marginally better than positive emotional accuracy, $d = .465$, $SE = .154$, $p = .057$, 95% CI [−.025, .954].

These findings, collated across four studies, strongly support our main conclusion that when holding in mind two emotional states, people have more difficulty maintaining the initial negative emotional state than they do the second negative emotional state, whereas they do not have the same difficulty when maintaining positive emotional states. In Studies 4 and 5, we use a slightly different paradigm to test candidate mechanisms for these effects.

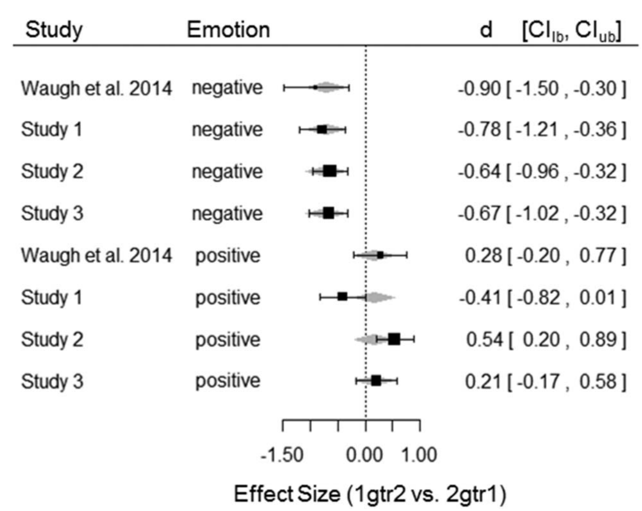

Figure 4. Results from internal meta-analysis of Studies 1–3 and Waugh et al. (2014). Effect sizes are calculated from the difference in idiographic accuracy on trials in which the first picture should have been chosen as eliciting a more intense emotional response (1gtr2) versus trials on which the second picture should have been chosen as eliciting a more intense emotional response (2gtr1). d = Cohen's d adjusted for correlated designs. CI = 95% confidence interval lower bound and upper bound. Error bars are standard error of the mean.

Study 4

In Study 4, we examine an alternative candidate mechanism for this relatively poorer working memory accuracy for the initial negative emotional state than for the second negative emotional state and initial positive emotional state. So far, we have found no evidence that this poor accuracy is due to decay and/or nontask related interference processes given that accuracy for the initial negative state did not change based on duration of delay period. In the present study, we test the possibility that the second negative emotional state may be retroactively interfering with the maintenance of the initial negative emotional state.

To test this mechanism, we no longer had participants compare the emotional intensity of the first and second emotional states in which it is impossible to disentangle the effects of the first and second emotional states. Instead, participants were asked to hold in mind their emotional intensity of each emotional state in order to rate that intensity at the end of the trial. We also added trials in which the second image was neutral. If the relatively poorer working memory accuracy for initial negative states is due to retroactive interference by the second emotional state, then we should expect to find relatively lower intensity ratings of the initial emotional state when followed by an emotional image than when followed by a neutral image.

Method

Participants. Sixty-nine participants (33 female) were recruited from Amazon's Mechanical Turk website to participate in a study on affective maintenance. To be eligible, participants had to be over the age of 18 (M = 30.59 years, SD = 14.45), speak fluent English, and reside in the United States. We also required the Mechanical Turk users to have at least a 95% completion and approval rating on their previously completed Mechanical Turk tasks to ensure we recruited engaged participants. Sample size was determined to be similar to that used in Studies 2 and 3. Four participants were excluded from data analysis due to not providing ratings for all the trial types, and one participant was excluded for providing nonvariable ratings (n = 64).

Emotional maintenance task. On each trial, participants viewed the initial image (2 s), a black screen during the delay period (7 s), a second image (2s), and a rating screen (5 s). Participants were instructed to remember their emotional state when viewing each image in order to rate that emotional state at the end of the trial. The participants were also instructed that they would not know beforehand which emotional state they would rate, so it was vital for them to maintain both emotional states.

At the end of each trial, participants rated the emotional intensity induced by either the first or second image. To limit verbal rehearsal, the participants made ratings on a visual analogue scale that varied from light gray (1 = *low emotional intensity*) to black (10 = *high emotional intensity*). To limit the use of spatial working memory, the rating scales were oriented either vertically or horizontally across the screen, which was randomized throughout the survey.[1]

As with the previous studies, all of the images were taken from the IAPS (Lang et al., 1997) in addition to selected Corel images (Mikels et al., 2005). Norms for the emotional images were taken from Mikels et al. (2005) and, for the neutral images, from Lang et al. (1997). Trials were counterbalanced to have equivalent normed intensity across trial types. Participants received 13 trials of each type: negative–negative rating first (M = 3.57, SD = .55), negative–negative rating second (M = 3.40, SD = .75), negative–neutral rating first (M = 3.63, SD = .80), negative–neutral rating second (M = 5.59, SD = .75),[2] positive–positive rating first (M = 3.69, SD = .84), positive–positive rating second (M = 3.74, SD = .64), positive–neutral rating first (M = 3.56, SD = .69), and positive–neutral rating second (M = 5.68, SD = .83), for a total of 104 trials. Within each valence, the negative image ratings were equivalent to each other, ts $<$ .64, and positive image ratings were equivalent to each other, ts $<$ .43.

Pretask ratings. Before the emotional maintenance task, participants viewed and rated all the images. To simulate the emotional maintenance task, participants viewed the image for 2 s before the survey auto-advanced to the rating. The participants then had 5 s to rate the emotional intensity of the image before the survey autoadvanced to the next image. As before, the rating consisted of the question, "How intense was your emotional response to this image?" This time, however, in order to eliminate any differences caused by the orientation of the scale, participants' responses were recorded on a vertical or horizontal scale to correspond to the scale used for that particular image later in the emotional maintenance task.

Procedure. Participants completed the survey online through their Mechanical Turk accounts. After the participants electronically signed the informed-consent forms, they completed a short demographics measure assessing their age, race, ethnicity, gender, education level, and family income level. Before completing the emotional-maintenance task, participants completed ratings of all the images shown in the emotional maintenance task.

After participants completed the emotional maintenance task, they were asked if they would like to complete a personality questionnaire. Regardless of whether they completed the personality questionnaire, participants were debriefed and thanked for their participation. They received a compensation code to be entered into the Mechanical Turk website. All participants were paid $3.00 through Mechanical Turk for completing this survey. The first 48 participants were allotted 1 hr 30 min to complete the survey. After further review, the allotted time was adjusted to 2 hr for the remaining participants. The Wake Forest University Institutional Review Board approved this procedure.

Results

Because we used the pretask ratings as references for the maintenance task ratings, we created difference scores for each trial type that represented the change in ratings from the pretask to the maintenance task (maintenance task – pretask ratings).

To test whether the remembered intensity of the first emotional state was affected by experiencing a subsequent negative emotional state, we conducted a 2 (Valence: positive, negative) $\times$ 2 (Second State: emotional, neutral) $\times$ 2 (Rating Orientation: horizontal, vertical) within-subjects ANOVA on the ratings from the

[1] Due to an error, horizontal and vertical ratings were not evenly distributed across the different trial types, with trial types having between four and nine of one and nine and four of the other.

[2] To balance the trial types, we added three neutral Corel images, but these have not been normed yet.

first emotional state of each pair. There were no significant main effects of Valence, $F(1, 63) = 2.06$, $p = .16$, Second State, $F(1, 63) = .01$, $p = .941$, or Rating Orientation, $F(1, 63) = .17$, $p = .68$. Importantly, there was a significant interaction of Valence × Second State, $F(1, 63) = 8.32$, $p = .005$ (Figure 5A). When participants were maintaining negative emotional states, there was an equivalent decrease in intensity (from pretask to maintenance task) when those initial states were followed by negative images ($M = -.504$, $SE = .135$) as when they were followed by neutral images ($M = -.717$, $SE = .135$), $t(63) = 1.63$, $p = .107$. Alternatively, when participants were maintaining positive emotional states, there was a greater intensity decrease when those initial states were followed by positive images ($M = -.528$, $SE = .116$) as when they were followed by neutral images ($M = -.302$, $SE = .097$), $t(63) = 2.62$, $p = .011$. These findings suggest that there was retroactive interference of the second positive image on maintaining the initial positive emotional state, but not when maintaining negative emotional states.

To determine whether these decreases in emotional intensity from the pretask to the maintenance task were due to poor maintenance of the emotional state or to habituation to the images, we next examined if there were changes in emotional intensity for the second picture of each pair as well. We conducted a 2 (Valence: positive, negative) × 2 (Picture Order: first, second) × 2 (Rating Type: horizontal, vertical) within-subjects ANOVA. The only significant effect was a main effect of Picture Order, $F(1, 63) = 48.30$, $p < .001$: There was a significant decrease in intensity from the pretask to the maintenance task for the initial emotional states ($M = -.516$, $SE = .097$), $t(63) = 5.32$, $p < .001$, whereas there was no significant change in intensity for the second emotional states ($M = .117$, $SE = .090$), $t(63) = 1.30$, $p = .198$ (Figure 5B). Clearly, this decrease in reported intensity for the initial emotional states was due to poor maintenance and not only to habituation.

Discussion

In this study, we showed that the previously identified deficit in maintaining negative emotional states does not seem to be due to retroactive interference of the viewing the second emotional image and the emotions induced by this image. Participants reported a decrease in the intensity of their initial negative emotions (which was not due to habituation) regardless of whether that emotional state was followed by another negative emotional state or by a neutral state. When maintaining the initial positive emotional state, however, viewing that second positive emotional image did seem to interfere with the maintenance of the initial state.

Study 5

So far, we have not been able to identify the mechanism that underlies people's relatively poor maintenance of negative emotional states. In this study, we test one final mechanism. Sometimes, the maintenance of information can be disrupted by poor initial encoding of that information (Tas, Luck, & Hollingworth, 2016). In the case of emotional working memory, participants may not be fully engaging in and encoding the initial negative emotional state, which then affects their maintenance of that state. One reason for why they might not engage in the initial emotional state is that they know that they will certainly (Studies 1–3), or possibly (Study 4), see another negative image in a few seconds. Therefore, preparing for a future negative stimulus may hinder engagement with the present stimulus. We test this possibility in Study 5 by using the same paradigm from Study 4 except that we alert participants before they view the first image whether the second image will be emotional or neutral. If preparing for a future negative stimulus hinders the encoding of the present stimulus, then we should see the previously found decrease in intensity on those trials in which participants are told that the second image will be negative, but not when they are told that the second image will be neutral.

Method

Participants. Seventy-one participants (37 female) were recruited from Amazon's Mechanical Turk website to participate in a study on affective maintenance. To be eligible, participants had to be over the age of 18 ($M = 33.94$ years, $SD = 8.06$), speak fluent English, and reside in the United States. We also required the Mechanical Turk users to have at least a 95% completion and approval rating on their previously completed Mechanical Turk tasks to ensure we recruited engaged participants. Sample size was determined to be similar to that used in Studies 2–4. Eight participants were excluded from data analysis due to not providing ratings for all the trial types ($n = 63$).

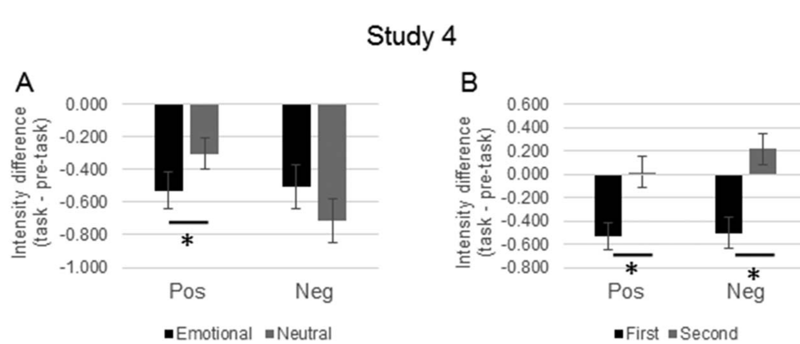

Figure 5. Study 4 emotion intensity differences (pretask ratings—maintenance task ratings) for positive (Pos) and negative (Neg) trials in which (A) the first emotional image is followed by either a similar emotional image or a neutral image and (B) the first emotional image of each pair versus the second emotional image of each pair. Error bars are standard error of the mean. * $p < .05$.

Emotional maintenance task. The maintenance task was similar to that used in Study 4 except that participants saw a screen (2 s) before each trial that indicated whether the two images on that trial "represent similar emotions" (*similar*) or "represent different emotions" (*different*). In this way, once participants viewed the first image, participants would know whether the second image was going to be negative (first image is negative + similar), positive (first image is positive + similar), or neutral (either positive or negative image + different). The other change was that we balanced the number of rating types (horizontal vs. vertical) among each of the trial types (six each).

Pretask ratings. The pretask ratings were the same as in Study 4.

Procedure. Participants completed the survey online through their Mechanical Turk accounts. After the participants electronically signed the informed-consent forms, they then completed a short demographics measure assessing their age, race, ethnicity, gender, education level, and family income level. Before completing the emotional-maintenance task, participants completed ratings of all the images shown in the emotional maintenance task.

After participants completed the emotional maintenance task, the participants were asked if they would like to complete a personality questionnaire. Regardless of whether they completed the personality questionnaire, participants were debriefed and thanked for their participation. They received a compensation code to be entered into the Mechanical Turk website. All participants were paid $3.00 through Mechanical Turk for completing this survey. The Wake Forest University Institutional Review Board approved this procedure.

Results

As in Study 4, we created difference scores for each trial type that represented the change in ratings from the pretask to the maintenance task (maintenance task – pretask ratings).

To test whether the remembered intensity of the first emotional state was affected by anticipating and then experiencing a subsequent negative emotional (vs. neutral) state, we conducted a 2 (Valence: positive, negative) × 2 (Second State: emotional, neutral) × 2 (Rating Orientation: horizontal, vertical) within-subjects ANOVA on the ratings from the first emotional state of each pair. There were no significant main effects. Unlike in Study 4, there was also not a significant interaction of Valence × Second State, $F(1, 62) = .00$, $p = .991$ (Figure 6A). Importantly, when participants were maintaining negative emotional states, there was still an equivalent decrease in intensity (from pretask to maintenance task) when those initial states were followed by negative images ($M = -.538$, $SE = .128$) as when they were followed by neutral images ($M = -.676$, $SE = .153$), $t(62) = 1.19$, $p = .239$. Different from Study 1, however, was that this pattern also held for ratings of positive images that were followed by positive images ($M = -.414$, $SE = .148$) and positive images that were followed by neutral images ($M = -.550$, .107), $t(62) = .95$, $p = .343$.

To determine whether these decreases in emotional intensity from the pretask to the maintenance task were due to poor maintenance of the emotional state or to habituation to the images, we next examined these changes (or not) in emotional intensity for the second picture of each pair as well. We conducted a 2 (Valence: positive, negative) × 2 (Picture Order: first, second) × 2 (Rating Type: horizontal, vertical) within-subjects ANOVA. There was again a significant main effect of Picture Order, $F(1, 62) = 15.98$, $p < .001$, but this time it was qualified by a significant interaction of Picture Order × Valence, $F(1, 62) = 4.49$, $p = .038$ and a significant interaction of Picture Order × Rating Type, $F(1, 62) = 10.69$, $p = .002$. We first explored the interaction of Picture Order × Valence (Figure 6B). As in Study 4, participants generally reported a greater decrease in intensity from pretask to maintenance task for the first images of each pair ($M = -.476$, $SE = .108$) than for the second images of each pair ($M = .027$, $SE = .112$). Unlike Study 4, however, for negative images this decrease in intensity for the first image ($M = -.538$, $SE = .128$) relative to the second image ($M = .203$, $SE = .138$), $t(62) = 5.26$, $p < .001$, was greater than the relative differences in decreased intensity for the first ($M = -.414$, $SE = .148$) and second ($M = -.148$, $SE = .121$) positive images, $t(62) = 1.39$, $p = .172$. The interaction of Picture Order × Rating Type seemed to be due to a greater decrease in ratings for those first images that were rated on a horizontal scale ($M = -.685$, $SE = .142$) than those rated on a vertical scale ($M = -.268$, $SE = .122$), $t(62) = 2.74$, $p = .008$. In contrast, rating type did not matter when rating the second images ($M_h = .054$, $SE_h = .124$; $M_v = .001$, $SE_v = .136$), $t(62) = .40$, $p = .693$.

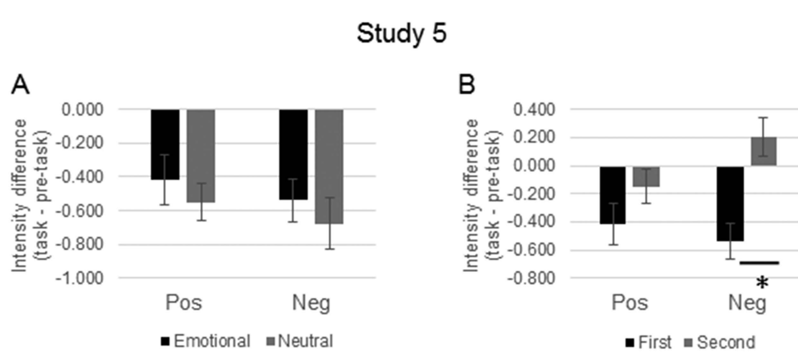

Figure 6. Study 5 emotion intensity differences (pretask ratings—maintenance task ratings) for positive (Pos) and negative (Neg) trials in which (A) the first emotional image is followed by either a similar emotional image or a neutral image and (B) the first emotional image of each pair versus the second emotional image of each pair. * $p < .05$. Error bars are standard error of the mean.

Discussion

In this study, we investigated whether the poor working memory for negative emotional states was due to decreased encoding/engagement with the initial negative image because of preparation for the second negative image. Participants reported a similar decrease (relative to pretask ratings) of the intensity of their initial negative emotional state regardless of whether it was followed by a negative or neutral image. This suggests that knowing that a neutral image was coming second did not influence participants' poor memory of their initial negative emotional state. Therefore, we do not have evidence that the poor maintenance of negative emotional states seen in the previous studies is due to poor encoding of/engagement with that initial negative emotional image. A caveat to that, however, is that participants may still be engaging poorly with the initial negative image, but that this poor engagement is due to some alternative mechanism besides the anticipation of a second negative image. Future investigations should test this possibility.

In Study 4, we found that participants' reported decrease in reported intensity for their initial emotional states was greater when that state was followed by a positive image than when it was followed by a neutral image. We thought this might be evidence that the second positive images were retroactively interfering with the maintenance of the initial positive emotional states. In the current study, however, this effect was no longer present when participants were cued in advance about the emotionality or neutrality of the second image. Notably, this difference between studies seems to be primarily due to a greater decrease in intensity for positive emotional states followed by neutral images in Study 5 (when cued) than in Study 4 (when not cued). One possible explanation is that when they were cued that the next image was to be neutral, participants did not try as hard to maintain their initial positive emotional state knowing that there would not be any interference later by another positive image. This explanation is highly speculative, however, and not parsimonious given that it posits two different mechanisms (retroactive interference, reduced effort) for the equivalent decreases in emotional intensity for positive emotional states followed by positive images and those followed by neutral images, respectively. A more parsimonious explanation is that the differences between these two studies are random and that being cued about the second image did not have a reliable effect on the maintenance of positive (or negative) emotional states.

General Discussion

In this set of studies, we provided a method of determining the degree to which the accuracy in a paired-image affective working memory paradigm (cf. Mikels et al., 2008; Waugh et al., 2014) is due to the successful maintenance of the initial emotional state. Similar to previous studies (Gard et al., 2011; Waugh et al., 2014), participants provided ratings of the images outside of the affective working memory task and these ratings, in lieu of norms, were used as references to determine whether selecting one image as more intense than the other was correct. In the current studies, we also found that these ratings proved useful as references regardless of whether participants provided them before (Study 3) or after the maintenance task (Studies 1 and 2). Findings from Studies 4 and 5 also revealed that any emotional intensity differences between these reference ratings and the affective working memory task ratings were unlikely to be due to habituation processes.

More importantly, we calculated accuracy separately for trials in which participants' ratings deemed that the initial image of the pair was supposed to be selected as inducing the more intense emotional state and for trials in which the reverse was true. This allowed us to separate aspects of accuracy that were affected by maintenance processes (1gtr2) from those aspects affected by other working memory related processes (2gtr1). This "initial idiographic accuracy" (1gtr2) metric will prove crucial when using this affective working memory paradigm to understand both basic mechanisms and individual differences in emotional maintenance.

We used the first application of this initial idiographic accuracy metric to investigate whether there are differences in people's ability to maintain negative and positive emotional states. The results from the internal meta-analysis indicate that people exhibit poorer initial idiographic accuracy when trying to maintain negative emotional states than when trying to maintain positive emotional states. If our initial idiographic accuracy metric is a valid measure of emotional maintenance, then this finding suggests that people have more trouble maintaining negative emotional states than they do positive emotional states. Importantly, this difference in negative and positive emotional working memory is for the emotional states themselves and not for emotionally charged verbal (Kensinger & Corkin, 2003) or pictoral (Jackson et al., 2014) information.

We conducted several studies to examine the mechanism that might underlie this difference in negative and positive emotional maintenance. We did not find any evidence that this difference was due to decay or to spontaneous nontask interference (Studies 2 and 3), retroactive interference of the second emotional state (Studies 4 and 5), or reduced engagement in the first emotional state due to anticipation of the second emotional state (Study 5). Therefore, the mechanism that underlies this difference in negative and positive emotional maintenance is still unclear. Future studies should explore reduced engagement of the first emotional state for some reason besides anticipation, and implicit regulation processes that could reduce negative emotion intensity even when people are tasked to maintain it.

One challenge to the formulation that individuals are not as able to maintain negative emotional states as they are to maintain positive emotional states arises when exploring the pattern of findings in Studies 4 and 5, in which participants reported relatively equivalent (in Study 4, less so in Study 5) decreases in emotional intensity from pretask ratings to maintenance task ratings for initial negative and positive emotional states. This finding could suggest that people have trouble maintaining the emotional intensity of both negative and positive states and that some other process besides maintenance underlies the initial idiographic accuracy differences between negative and positive emotional working memory. For example, it could be that when maintaining positive emotional states to compare them to a second emotional state, at the moment of comparison the initial eroded state is somehow "boosted" in an attempt to correct for these poor maintenance processes. Why this correction would only occur when comparing emotional states and not when just rating them is unclear, as is why this correction process would differentially occur for positive than for negative emotional maintenance. Although this formulation is intriguing and deserves future investigation, this finding is not consistent (only found in Study 4, not in Study 5) and there are enough differences between the emotional working memory via comparison tasks in Studies 1–3 (as well as Waugh et al., 2014) and the remembered intensity rating tasks in Studies 4 and 5 to not sway us from our more parsimonious formulation that people are just not as able to maintain negative emotional states as they are to maintain positive states.

At first glance, these findings that people had greater difficulty maintaining negative emotional states than they did positive emotional states seems to conflict with conventional wisdom about negative emotion bias (Cacioppo, Gardner, & Berntson, 1999) - the evolutionary priority of responding to negative events in which it is more important for one's fitness to respond more quickly and strongly to threats than to opportunities in the environment (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Key to this notion is that whereas negativity bias is really about primacy when responding to stimuli/events (Cacioppo et al., 1999), emotional maintenance is about continuing to respond once that stimulus/event is over. From a functional perspective, it may not be adaptive to continue to expend energy maintaining the heightened arousal and vigilance that comes from responding to these negative stimuli once those stimuli are no longer relevant. Therefore, one might only observe the effective maintenance of negative emotions beyond the relevance of the stimulus in disorders and/or traits that are characterized by poor negative emotional recovery, such as rumination, depression (Nolen-Hoeksema, 2000), and anxiety (Brosschot, Gerin, & Thayer, 2006). People who experience high levels of perseverative negative cognitions (Brosschot et al., 2006) may not show the observed deficit in the maintenance of negative emotion because to them these negative emotional states may continue to be relevant after the offset of the stimulus—a formulation that should be tested in future studies.

An important caveat to this formulation is that although these stimuli may no longer be ecologically relevant after their presentation, because in our task we tell participants to maintain the emotional states from the stimuli they still remain task-relevant throughout the maintenance period. Indeed, we found in Waugh et al. (2014) that when maintaining negative emotional states, people recruited dorsomedial prefrontal cortex and lateral prefrontal cortex; clearly, therefore, people are using top-down control systems in their attempt to maintain negative states. However, we also found that people recruited rMFC when maintaining positive emotional states, but not when maintaining negative emotional states. The rMFC tracks the natural duration of emotional states (Wager et al., 2009; Waugh et al., 2010), suggesting that although people were utilizing top-down control systems when maintaining both negative and positive emotions, they were using bottom-up emotion-generation regions only when maintaining positive emotions. This finding is consistent with the current formulation that although these stimuli were task-relevant past their presentation, they may not have remained ecologically relevant and there may be systems in place to limit the maintenance of negative emotions beyond their ecological relevance.

Alternatively, participants seem to be more successful when maintaining positive emotional states, albeit not perfectly (as is evident from Studies 4 and 5). This suggests that the systems that are limiting the maintenance of negative emotions past their ecological relevance do not do the same for the maintenance of positive emotions. It may be evolutionarily less maladaptive to have positive emotions linger than to have negative emotions persist. Consistent with this formulation is that, while people do have a bias toward reacting stronger toward negative stimuli than to positive stimuli, in the absence of strong stimuli people generally have a "positivity offset"—their natural stimulus-free mood tends to be positive (Cacioppo et al., 1999). One facet of this stimulus-free positive mood may be the lingering positive emotions from stimuli that are no longer relevant. This formulation is speculative, however, and must be tested in future investigations.

Although this positivity offset may be apparent "on average" in the population, there are reliable individual differences in people's positivity offset and default positive mood (Ito & Cacioppo, 2005), as well as cultural and situational effects on people's motivation to experience positive emotions. Compared with healthy controls, people with depression-related anhedonia tend to experience lower levels of positive mood (Clark & Watson, 1991) and are less motivated to experience positive events (Sherdell, Waugh, & Gotlib, 2012), suggesting that they may have trouble maintaining positive emotions after the offset of stimuli. In addition, there are cultural variations in the desire to experience high levels of positive emotion (Tsai, 2007), as well as situations in which letting positive emotions linger past stimulus offset and not appropriately regulating them may be detrimental (e.g., when one is about to engage in competition; Tamir et al., 2008). Taken together with the previously stated possibility that people high on perseverative cognition may show enhanced negative emotional maintenance, and with previous findings that some disorders are marked by poor emotional working memory overall (Gard et al., 2011), these formulations underscore the need to investigate how individual differences and context may affect emotional maintenance.

Although we have made broad claims about the maintenance of negative and positive emotional states, these claims are currently bound to this specific affective working memory task. For example, participants had to compare an initial emotional state to a subsequent emotional state, so it is unclear what form emotional maintenance takes in the absence of a subsequent task-relevant stimulus. Further, the stimuli we used were brief images; therefore, it is not clear what form emotional maintenance takes when individuals respond to stimuli that are longer, more ecologically relevant, and/or verbal/situational. In addition, there may have been differences in arousal among the images. Because arousal can sometimes affect memory (Libkuman, Nichols-Whitehead, Griffith, & Thomas, 1999), this construct should be examined in future studies as a possible mechanism underlying these findings. Future investigations that address these questions are needed to provide a more comprehensive understanding of emotional maintenance.

In conclusion, we showed that by tracking specific types of errors in the affective working memory paradigm, one can more accurately assess emotional maintenance success. In doing so, we discovered that people maintain positive emotional states better than they do negative states. Future investigations should focus on possible individual differences and situations that could moderate this effect. Although we did not identify specific mechanisms that explain this effect, it is clear that doing so will be important for gaining a more comprehensive understanding of emotional maintenance and the functions of positive and negative emotional states.

References

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29. http://dx.doi.org/10.1146/annurev-psych-120710-100422

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science, 21,* 413–419. http://dx.doi.org/10.1177/0963721412459513

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323–370. http://dx.doi.org/10.1037/1089-2680.5.4.323

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology, 42,* 1–15. http://dx.doi.org/10.1111/j.1469-8986.2005.00271.x

Brosschot, J. F., Gerin, W., & Thayer, J. F. (2006). The perseverative cognition hypothesis: A review of worry, prolonged stress-related physiological activation, and health. *Journal of Psychosomatic Research, 60,* 113–124. http://dx.doi.org/10.1016/j.jpsychores.2005.06.074

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76,* 839–855. http://dx.doi.org/10.1037/0022-3514.76.5.839

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100,* 316–336. http://dx.doi.org/10.1037/0021-843X.100.3.316

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177. http://dx.doi.org/10.1037/1082-989X.1.2.170

Feinstein, J. S., Duff, M. C., & Tranel, D. (2010). Sustained experience of emotion after loss of memory in patients with amnesia. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 7674–7679. http://dx.doi.org/10.1073/pnas.0914054107

Gard, D. E., Cooper, S., Fisher, M., Genevsky, A., Mikels, J. A., & Vinogradov, S. (2011). Evidence for an emotion maintenance deficit in schizophrenia. *Psychiatry Research, 187,* 24–29. http://dx.doi.org/10.1016/j.psychres.2010.12.018

Gilboa, E., & Gotlib, I. H. (1997). Cognitive biases and affect persistence in previously dysphoric and never-dysphoric individuals. *Cognition & Emotion Special Issue: The cognitive psychology of depression, 11,* 517–538.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research synthesis* (2nd ed.). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412985031

Ito, T., & Cacioppo, J. T. (2005). Variations on a human universal: Individual differences in positivity offset and negativity bias. *Cognition and Emotion, 19,* 1–26. http://dx.doi.org/10.1080/02699930441000120

Jackson, M. C., Linden, D. E. J., & Raymond, J. E. (2014). Angry expressions strengthen the encoding and maintenance of face identity representations in visual working memory. *Cognition and Emotion, 28,* 278–297. http://dx.doi.org/10.1080/02699931.2013.816655

Joormann, J., Levens, S. M., & Gotlib, I. H. (2011). Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological Science, 22,* 979–983. http://dx.doi.org/10.1177/0956797611415539

Kensinger, E. A., & Corkin, S. (2003). Effect of negative emotional content on working memory and long-term memory. *Emotion, 3,* 378–393. http://dx.doi.org/10.1037/1528-3542.3.4.378

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22,* 2693–2710. http://dx.doi.org/10.1002/sim.1482

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville, FL: NIMH Center for the Study of Emotion and Attention, University of Florida.

Levens, S. M., & Gotlib, I. H. (2010). Updating positive and negative stimuli in working memory in depression. *Journal of Experimental Psychology: General, 139,* 654–664. http://dx.doi.org/10.1037/a0020283

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1545–1551. http://dx.doi.org/10.1037/a0017010

Libkuman, T. M., Nichols-Whitehead, P., Griffith, J., & Thomas, R. (1999). Source of arousal and memory for detail. *Memory & Cognition, 27,* 166–190. http://dx.doi.org/10.3758/BF03201222

Lindquist, M. A., Waugh, C., & Wager, T. D. (2007). Modeling state-related fMRI activity using change-point theory. *NeuroImage, 35,* 1125–1141. http://dx.doi.org/10.1016/j.neuroimage.2007.01.004

McCabe, D. P., Roediger, H. L., III, McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24,* 222–243. http://dx.doi.org/10.1037/a0017619

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior Research Methods, Instruments, & Computers, 37,* 626–630. http://dx.doi.org/10.3758/BF03192732

Mikels, J. A., Reuter-Lorenz, P. A., Beyer, J. A., & Fredrickson, B. L. (2008). Emotion and working memory: Evidence for domain-specific processes for affective maintenance. *Emotion, 8,* 256–266. http://dx.doi.org/10.1037/1528-3542.8.2.256

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology, 109,* 504–511. http://dx.doi.org/10.1037/0021-843X.109.3.504

Perlstein, W. M., Elbert, T., & Stenger, V. A. (2002). Dissociation in human prefrontal cortex of affective influences on working memory-related activity. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 1736–1741. http://dx.doi.org/10.1073/pnas.241650598

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523. http://dx.doi.org/10.1037/0022-3514.45.3.513

Sherdell, L., Waugh, C. E., & Gotlib, I. H. (2012). Anticipatory pleasure predicts motivation for reward in major depression. *Journal of Abnormal Psychology, 121,* 51–60. http://dx.doi.org/10.1037/a0024945

Tamir, M., Mitchell, C., & Gross, J. J. (2008). Hedonic and instrumental motives in anger regulation. *Psychological Science, 19,* 324–328. http://dx.doi.org/10.1111/j.1467-9280.2008.02088.x

Tas, A. C., Luck, S. J., & Hollingworth, A. (2016). The relationship between visual attention and visual working memory encoding: A dissociation between covert and overt orienting. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 1121–1138. http://dx.doi.org/10.1037/xhp0000212

Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science, 2,* 242–259. http://dx.doi.org/10.1111/j.1745-6916.2007.00043.x

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion, 9,* 83–91. http://dx.doi.org/10.1037/a0014610

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Wager, T. D., Waugh, C. E., Lindquist, M., Noll, D. C., Fredrickson, B. L., & Taylor, S. F. (2009). Brain mediators of cardiovascular responses to social threat: Part I: Reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *NeuroImage, 47,* 821–835. http://dx.doi.org/10.1016/j.neuroimage.2009.05.043

Waugh, C. E., Hamilton, J. P., Chen, M. C., Joormann, J., & Gotlib, I. H. (2012). Neural temporal dynamics of stress in comorbid major depressive disorder and social anxiety disorder. *Biology of Mood & Anxiety Disorders, 2,* 11. http://dx.doi.org/10.1186/2045-5380-2-11

Waugh, C. E., Hamilton, J. P., & Gotlib, I. H. (2010). The neural temporal dynamics of the intensity of emotional experience. *NeuroImage, 49,* 1699–1707. http://dx.doi.org/10.1016/j.neuroimage.2009.10.006

Waugh, C. E., Lemus, M. G., & Gotlib, I. H. (2014). The role of the medial frontal cortex in the maintenance of emotional states. *Social Cognitive and Affective Neuroscience, 9,* 2001–2009. http://dx.doi.org/10.1093/scan/nsu011

Received June 20, 2017
Revision received November 15, 2017
Accepted January 22, 2018 ■