American Psychological Association

Journal of Experimental Psychology: General

ISSN: 0096-3445

2021, Vol. 150, No. 1, 67–82
http://dx.doi.org/10.1037/xge0000780

Affective Forecasting: A Selective Relationship With Working Memory for Emotion

Colleen C. Frank, Alexandru D. Iordan, and Tara L. Ballouz
University of Michigan

Joseph A. Mikels
DePaul University

Patricia A. Reuter-Lorenz
University of Michigan

Affective forecasting (AF), the ability to predict one's future feelings, is important for decision making. We posit that AF entails the ability to maintain and evaluate an emotional feeling state, and thus requires affective working memory (AWM; Mikels & Reuter-Lorenz, 2019). To test this hypothesis, a series of studies investigated whether individual differences in AWM are related to AF ability. In the first study, we document that measures of AWM and AF are positively related, whereas an analogous measure of visual working memory is unrelated to AF in separate groups of participants. Two further within-group studies (1 preregistered) demonstrate that maintenance of affective information predicts AF performance, whereas maintenance of brightness information does not. Further, 2 additional measures of visual working memory (Corsi block-tapping and change detection) did not independently predict AF ability. Taken together the results demonstrate a reliable and selective relationship between AWM and AF, suggesting that AWM is a separable working memory subsystem and an elemental capacity that contributes to the type of higher-order emotional processes involved in AF.

Keywords: short-term memory, feelings, future thinking, individual differences

Supplemental materials: http://dx.doi.org/10.1037/xge0000780.supp

Our decisions are often guided by the feelings we anticipate in response to potential outcomes. Indeed, the ability to predict one's future feelings, known as *affective forecasting* (AF; Wilson & Gilbert, 2003), is thought to play an important role in decisions, preferences, and behaviors that affect health and economic outcomes and personal well-being. However, people are often biased and inaccurate when predicting their future feelings (Mathieu & Gosling, 2012; Wilson & Gilbert, 2003). Therefore, understanding the elemental processes that contribute to AF may provide valuable insight into why AF is flawed and how it can be improved. We posit that AF entails the ability to actively maintain and evaluate emotional feelings, and thus requires *affective working memory* (AWM). AWM is responsible for actively maintaining feeling states in mind to support goal-oriented behavior (Mikels, Larkin, Reuter-Lorenz, & Carstensen, 2005; Mikels, Reuter-Lorenz, Beyer, & Fredrickson, 2008; Smith & Lane, 2015; see Mikels & Reuter-Lorenz, 2019, for a review). If AWM is a core ability underlying AF, then individual differences in AWM should influence the accuracy of AF. The goal of the present report is to test this hypothesis.

Affective forecasts typically rely on self-reports in which participants predict how they will feel about future stimuli or events (Andrews & Robinson, 1991; Wilson & Gilbert, 2003). Some studies ask participants to predict their feelings to outcomes of genuine future events such as football games or presidential elections (Scheibe, Mata, & Carstensen, 2011; Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000). Others use laboratory tests that entail verbal descriptions of emotional scenes and require participants to predict how they would feel subsequently when viewing the actual scene (Hoerger, Chapman, Epstein, & Duberstein, 2012; Robinson & Clore, 2001). In each case, AF ability is assessed by measuring the discrepancy between ratings of predicted feelings and ratings of feelings experienced at a later time point in order to determine the prediction-accuracy score, also referred to as absolute accuracy (Mathieu & Gosling, 2012; Wilson & Gilbert, 2003). Participants may be asked to predict how long or how frequently they will feel a given emotion or the degree to which an event or stimulus will affect their mood. Any distinct feature of predicted emotion can be used as criteria for AF accuracy.

The present study used ratings of emotional intensity, or predictions of how strongly participants will feel in response to a particular event. The tendency to overestimate the emotional impact, including the emotional intensity, of a future outcome is

This article was published Online First July 2, 2020.

Colleen C. Frank, Alexandru D. Iordan, and Tara L. Ballouz, Department of Psychology, University of Michigan; Joseph A. Mikels, Department of Psychology, DePaul University; Patricia A. Reuter-Lorenz, Department of Psychology, University of Michigan.

Correspondence concerning this article should be addressed to Colleen C. Frank, Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109. E-mail: ccfrank@umich.edu

known as the *impact bias* (Wilson & Gilbert, 2003). For example, impact bias may be demonstrated by a student who overestimates how positive they expect to feel if they were to receive a high exam score. Some evidence indicates that emotional intensity is only moderately overestimated (Mathieu & Gosling, 2012), and other work suggests the bias may disappear depending on how emotional ratings are collected (Levine, Lench, Kaplan, & Safer, 2012). Still, others maintain that the impact bias is "alive and well" (Wilson & Gilbert, 2013). Regardless, there is substantial evidence for individual differences in AF performance (Dunn, Brackett, Ashton-James, Schneiderman, & Salovey, 2007; Hoerger et al., 2012; Hoerger, Chapman, & Duberstein, 2016) for which the mechanisms are not yet understood. To the extent that AWM plays a mechanistic role in AF, AWM may be an underlying source of interindividual variability in AF. Therefore, the primary goal of the present investigation was to determine whether AWM is related to, and predictive of, individual differences in forecasting of emotional intensity.

Working memory is a limited capacity system that temporarily maintains information actively in mind (Baddeley, 1992, 2007, 2012) in support of goal-directed behavior such as planning and problem-solving (Smith & Jonides, 1999). According to current models, working memory is composed of separable, domain-specific subsystems specialized for the short-term maintenance of different types of information (e.g., visual, verbal, spatial; Baddeley, 2012; Repovš & Baddeley, 2006). Affective working memory is posited to be an additional domain-specific subsystem specialized for maintaining emotional feelings actively in mind over a brief interval (Mikels & Reuter-Lorenz, 2019; cf. LeDoux & Brown, 2017). Previous studies have measured AWM using an affect maintenance task in which participants view an emotional image, then assess and hold their emotional reaction (i.e., the intensity of their feelings) in mind over a delay. Participants then view another emotional image and determine if the second image evoked an emotional response with higher or lower intensity than the first picture in the pair (Broome, Gard, & Mikels, 2012; Mikels et al., 2005, 2008). To determine if the ability to maintain an emotional feeling can be dissociated from working memory for affectively "neutral" information, previous investigations have used an analogous visual working memory task that measures nonaffective maintenance abilities. In this task, participants assess and hold the *brightness intensity* of a neutral image in mind over a delay before deciding if a second image has higher or lower brightness intensity (Broome et al., 2012; Mikels et al., 2005, 2008). The brightness maintenance task is particularly well-suited for comparison with the affect maintenance task; both require the subjective assessment of pictorial stimuli (International Affective Picture System [IAPS] images; Lang, Bradley, & Cuthbert, 1999) and a response decision based on assessing the relative intensity of subjective states evoked by two consecutive stimuli offset by a brief delay (Mikels et al., 2008). Furthermore, accuracy on both tasks can be scored using subjective intensity ratings. Following the maintenance phase of each task, participants rate the intensity (emotional or brightness) of each image used in the maintenance tasks. These ratings are then used to establish individualized accuracy scores for the respective maintenance task.

Several lines of research using variants of these tasks provide converging evidence that AWM is a separable subsystem that is at least partially dissociable from working memory for nonemotional, visual information (Mikels et al., 2005, 2008; see also Gard et al., 2011). Using selective interference methodology, Mikels and his colleagues (2008) found that a secondary emotion regulation task impaired performance on the affect maintenance task but not the brightness maintenance task. Conversely, a secondary cognitive task impaired performance on the brightness maintenance task but facilitated performance on the affect maintenance task, suggesting the dissociability of working memory for affective versus nonaffective information (Mikels et al., 2008). Further evidence for this conclusion derives from findings that older adults performed comparably to younger adults on the affect maintenance task, but were significantly impaired on the brightness maintenance task (Mikels et al., 2005). This pattern is consistent with the relative preservation of affective processing abilities in older adults despite cognitive decline that accompanies normal aging (Mather, 2016). Taken together, these studies suggest that AWM is dissociable from nonaffective working memory, a property that was tested further in the present investigation.

Additionally, prior research points to a potential relationship between AF and emotional intelligence (EI)—a set of skills related to the assessment, management, regulation, and expression of emotion, and the use of feelings to succeed in everyday life (Dunn et al., 2007; Hoerger et al., 2012; Mayer, Salovey, & Caruso, 2004). There are a variety of models (Petrides & Furnham, 2001; Mayer et al., 2004) and measures of the EI construct (see Hughes & Evans, 2018 for a recent review). Previous studies have identified relationships between AF and EI (albeit with some inconsistency) and these were generally hypothesized to be due to differences in basic emotion processing—such as the ability to identify, remember, and manage emotions (Dunn et al., 2007; Hoerger et al., 2012)—which could also relate to AWM. Thus, an ancillary goal of the present investigation was to reassess the relationships between trait and ability EI and AF, and to explore the potential relationships between AWM and EI. We expected that a better ability to maintain feeling states would be associated with superior EI and more successful AF.

In the first set of studies, 1a and 1b, we examined the relationship between AF and AWM, and between AF and visual working memory, respectively, in two independent samples of participants. We predicted there would be a positive relationship between AWM and AF (Study 1a), but no relationship between visual working memory and AF (Study 1b). Study 2 sought to replicate this selectivity within a new group of participants. Study 3 aimed to strengthen the evidence by replicating and extending the results using a preregistered study. Evidence for these predictions would lend credence to the idea that AWM is a fundamental capacity that contributes to the accuracy of affective forecasts and provide additional evidence that AWM is dissociable from working memory for nonaffective information.

Study 1a: Testing the Relationship Between AWM and AF Ability

Method

Participants. Seventy-nine undergraduate students participated in the study in exchange for course credit. One participant chose not to continue in the study after seeing the sample emotional images, three participants did not return for Session 2 (see

Design and Procedure) and one participant failed to follow task instructions, resulting in the exclusion of these five participants. In addition, eight participants were excluded due to poor performance on the affect maintenance task (see below for details). Thus, the reported analyses were performed on data from 66 participants (63.5% female, mean age 18.83, 71.6% self-identified White), who self-reported as right-handed and native English speakers. A power analysis based on data from Hoerger et al. (2012) Study 2 (N = 430), which used a similar AF task to measure the relationship between AF and a different variable of interest (i.e., emotional intelligence; see below), indicated we needed 46 participants to have 80% power for detecting a small–medium effect with the traditional α = .05 criterion of statistical significance (G*Power 3.1: Faul, Erdfelder, Buchner, & Lang, 2009). Our initial sample size exceeded the estimate to account for expected attrition and low performing individuals. The University of Michigan Institutional Review Board approved all procedures, and all participants provided informed consent prior to participating.

Design and procedure. Participants completed two 1-hr testing sessions, scheduled 1 week apart (see Figure 1). During the first session, participants performed 40 trials of the affect maintenance task, followed by Phase I of the AF task. Phase I of the AF task was the prediction phase, where participants were instructed to predict their future feelings based on a written description of a scene. During the second session, participants completed Phase II of the AF task, where participants were instructed to rate their experienced feelings after viewing an image of the scene. After Phase II of the AF task, participants completed another 40 trials of the affect maintenance task. This task sequence was intended to minimize the influence of the affect maintenance task on emotional ratings of the AF task between Phase I and Phase II. Finally, participants rated the emotional intensity of all images included in the maintenance task. All tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013), and are described, in turn, below.

Affect maintenance task. The affect maintenance task (see Figure 2) was similar to that employed in previous research measuring AWM (Broome et al., 2012; Mikels et al., 2005, 2008). For each trial, participants viewed one emotional picture (5 s) and were instructed to maintain the feelings elicited by this image. A retention interval ensued (3 s) before participants viewed a second emotional picture (5 s). Next, a green cross appeared, which prompted participants to report whether the second image had higher or lower emotional intensity than the first one. Emotional intensity was described to the participants as the strength or magnitude of their emotional reaction to each image, regardless of the picture's content. Participants responded to each trial by pressing either a key labeled H for higher or L for lower.

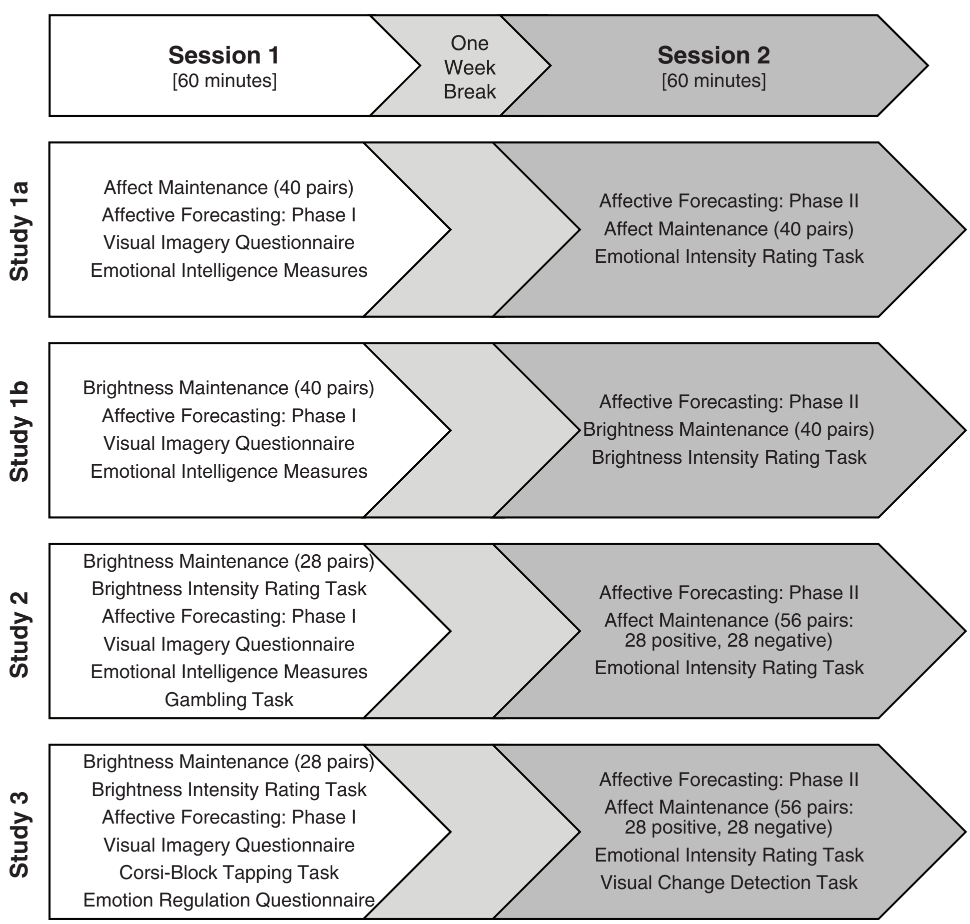

Figure 1. Protocols for Study 1a, Study 1b, Study 2, and Study 3.

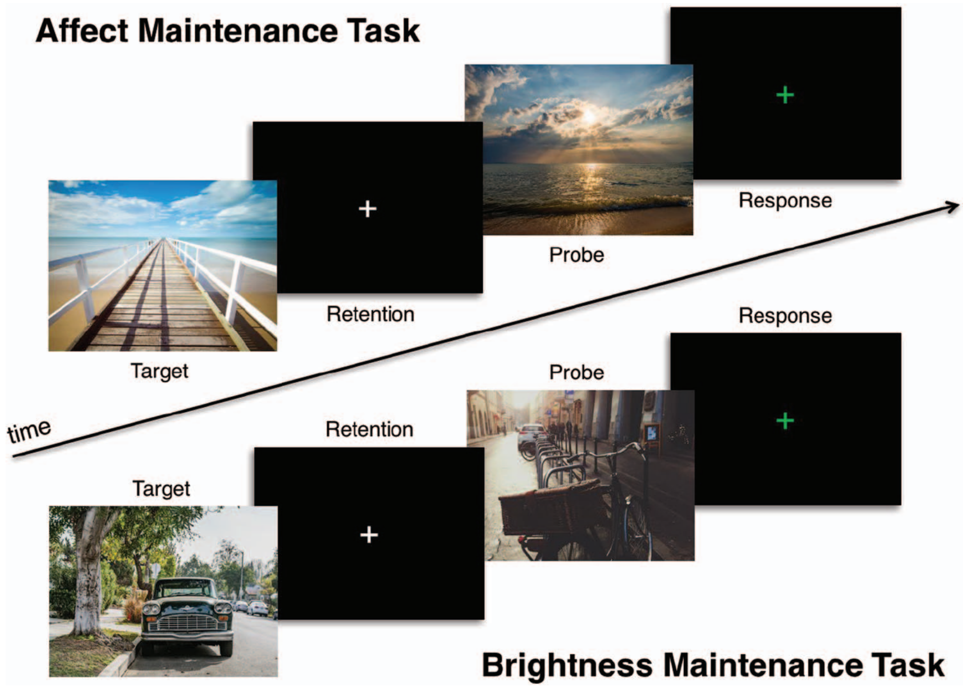

Figure 2. Schematic for the affect and brightness maintenance tasks used. Participants hold the emotional or brightness intensity of one image in mind over a delay to determine if a subsequent image has higher or lower intensity. Adapted from “Affective Working Memory: An Integrative Psychological Construct,” by J. A. Mikels and P. A. Reuter-Lorenz, 2019, *Psychological Science, 14*(4), p. 8. Copyright 2019 by SAGE Publications. See the online article for the color version of this figure.

There were 80 trials total and each trial consisted of a pair of images selected to have matching valence (40 positive and 40 negative trials). We used the stimuli selected by Broome et al. (2012), consisting of 80 matched-valence pairs created from a set of 160 images (80 positive, 80 negative). Images depicted pleasant (positive valence) or unpleasant (negative valence) scenes selected from the IAPS (Lang et al., 1999) and supplemented from an in-house database to include 17 additional high-arousal images (available upon request). Intensity ratings from all images were originally obtained from two independent samples (N_1 = 40 & N_2 = 40; Mikels et al., 2008) and were recorded using a 7-point scale that ranged from *low* (1) to *high* (7) emotional intensity, or degree of emotional reaction to each image. The difference in emotional intensity ratings between the two images in these pairs varied from 0.23 to 1.75 (M = .99, SD = .47). As in previous work with this image set, this difference, referred to as intensity distance, was used to categorize the stimulus set into near (intensity distance .88 or below) and far (intensity distance 1.03 or greater) pairs, resulting in 38 near pairs and 42 far pairs. Additionally, for each intensity distance subset, the second picture had higher intensity than the first for exactly half of the pairs, and vice versa for the other half. Following Broome et al. (2012), the 80 pairs were divided into two blocks (Image Sets A and B), such that they were equated for valence, intensity distance (near or far), and intensity order (second picture higher or lower), and their presentation at each of the two sessions was counterbalanced across participants. These restrictions were to ensure there would be no differences in performance due to the image sets, which was later confirmed by our participants, $t(65)$ = .46, p = .63. Trials within each block were presented in a randomized order for each participant.

Affective forecasting. During the first session of the AF task (i.e., Phase I; see Figure 3), participants read a description of a scene and were asked to imagine it. Participants then predicted how they would feel if they were to view the actual image and rated this feeling on a visual analog scale (actual resolution: 21 points) that ranged from endpoint anchors labeled *very unpleasant* to *very pleasant*. Participants read 10 such descriptions and rated their predicted feelings for each by clicking the computer mouse anywhere along the scale. During Session 2, one week later (i.e., Phase II; see Figure 3), participants viewed the images associated with each description and rated how each one made them feel using the same scale. Five negative (mean pleasure rating = 2.35, SD = 0.68) and five positive (mean pleasure rating = 7.42, SD = 0.69) images from the IAPS (Lang et al., 1999) were used, based on those used in Robinson and Clore (2001). Five of the images were the same as in Robinson and Clore (2001); to avoid repeating images already employed in the affect maintenance task, the other five images were replaced with alternative IAPS images matched as closely as possible in content to those used by Robinson and Clore (2001). Accordingly, five descriptions were taken from Robinson and Clore (2001) and the remaining were constructed to match the new replacement images (see Table S1 in the online supplemental material for image descriptions used in the current study). Descriptions and images were displayed on a desktop computer screen, and the order of stimuli was randomized for each participant independently for Phase I and Phase II.

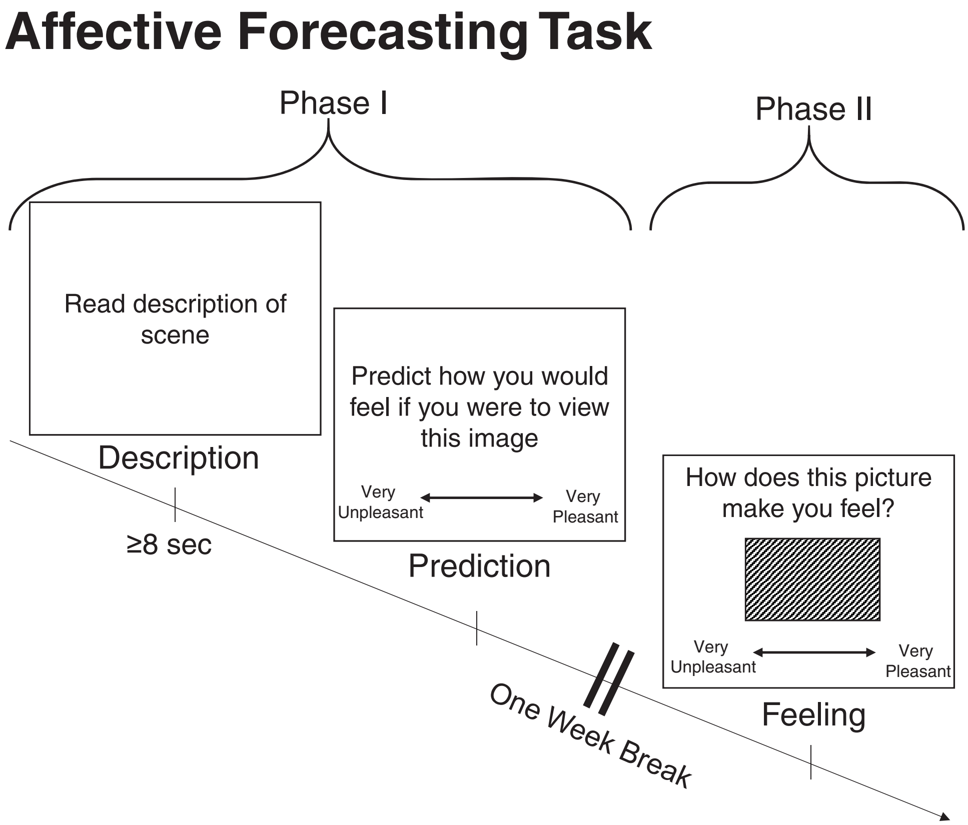

Figure 3. Schematic of the Affective Forecasting Task. During Phase I, participants read brief affectively laden descriptions of scenes and predict how they would feel if they were to view each image described. After a 1-week break, during Phase II, participants view each image and rate their experienced feelings.

Emotional intensity ratings. Accuracy on the affect maintenance task was scored on an individualized basis using each participant's subjective ratings of emotional intensity for each image (see Affect Maintenance Scoring for details). In order to obtain subjective ratings of emotional intensity, participants viewed all images from the affect maintenance task again individually, at the end of the second session. The rating task was performed at the end of Session 2 in order to minimize the possibility that the process of rating the images or that memory of prior ratings would contaminate relative intensity judgments and influence performance on the affect maintenance task. Images were randomly assigned to one of four blocks and trials within each block were presented in a randomized order. Participants were reminded that emotional intensity referred to the strength or magnitude of their emotional reaction. Then, using a visual analog scale that ranged from *not intense* to *extremely intense*, displayed below the image, participants rated the intensity of their feelings for each image. Responses were registered by a mouse click anywhere on the scale corresponding to the intensity of their feelings. The scale had a resolution of 21 units and ratings were used to calculate affect maintenance accuracy scores, as described below (see Affect Maintenance Accuracy Scores).

Additional measures. We assessed the participant's visual imagery ability to control for its potential influence on the AF task. Due to significant relationships between emotional intelligence (EI) and AF found in previous studies, we also included trait and ability measures of EI to see if the results would replicate here (Dunn et al., 2007; Hoerger et al., 2012). These additional measures were administered at the end of the first testing session (see Figure 1 for the complete order of tasks). We also assessed the participants' emotional state before and after each session to ensure they were not negatively influenced by the stimuli viewed in the study. These measures are briefly described, in turn, below.

Visual imagery. Visual imagery was assessed with the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973). The self-report measure asks participants to answer four questions about each of four imagined scenes, and to rate vividness on a 5-point scale of vividness. A response of 1 represents *perfectly clear and as vivid as normal vision* and 5 represents *no image at all, you only* know *that you are thinking of the object.* Scores for all 16 items were summed to create a composite VVIQ score where higher scores indicate poorer imaging capabilities (scores range from 16 to 80).

Emotional intelligence. Based on previous findings revealing relationships between AF and EI, we employed the Situational Test of Emotional Understanding (STEU; $\alpha = .77$; range: 0–1, MacCann & Roberts, 2008) to assess ability EI and the Assessing Emotions Scale (AES; $\alpha = .87$–$.90$; scores range from 33 to 165, Schutte et al., 1998), also known as the Schutte Emotional Intelligence Scale (Schutte, Malouff, & Bhullar, 2009), to assess trait EI.

Emotional state. To assess potential changes in emotional state due to participating in the study, the "state" version of the Positive and Negative Affect Scale (PANAS; positive items α = .89, negative items α = .85; Watson, Clark, & Tellegen, 1988) was administered at the beginning and end of each testing session. The scale consists of 20 words that describe feelings and emotions. For each word, participants are asked to rate the extent they feel that way "right now, that is, at the present moment" from 1 (*not at all*) to 5 (*extremely*). This measure was used only to ensure that the subjects' emotional states were not negatively impacted by the stimuli shown in our experiment, and will not be discussed further.

Affect maintenance scoring. Accuracy scoring was individualized based on each participant's image ratings from the emotional intensity rating task. These subjective intensity ratings were used to infer which image, when appearing with its pair, would be judged by that individual as having higher intensity. A response on the affect maintenance task was scored correct when it agreed with the participant's relative intensity ratings for the members of each pair. Trials with images receiving equal intensity ratings were excluded from the calculation of that participant's accuracy scores. On average, approximately six of each participant's 80 total trials, or 7.8% (*SD* = 3.02%) of all trials, were excluded for this reason. For each participant, accuracy scores were calculated as the sum of the affect maintenance responses that matched the intensity-based comparisons and divided by the number of trials included in the sum. Separate accuracy scores were calculated for each session and then averaged together to calculate a composite affect maintenance accuracy score. As mentioned above, eight participants were excluded from all subsequent analyses because they performed more than 2 standard deviations below the sample mean during either session. This level of performance was approximately 50% accuracy, suggesting participants were merely guessing, or otherwise not following instructions.

Affective forecasting scoring. AF prediction-accuracy scores were calculated based on the difference between the predicted rating in response to the picture description and the feeling rating in response to viewing the image for each participant. The absolute value of this difference was subtracted from the largest possible deviation score (i.e., 10), thus providing an AF score between zero (least accurate prediction) and 10 (most accurate prediction). For each participant, we averaged the scores across the 10 images to calculate a composite AF score (see also Dunn et al., 2007; Hoerger et al., 2012). Because we hypothesized that AWM would contribute to AF accuracy per se, we focused on magnitude, rather than the direction (i.e., overestimation vs. underestimation) of AF errors (see the online supplemental materials for exploratory analyses on directionality).

Results

Descriptive statistics for AF, affective maintenance, and vividness of visual imagery (VVIQ) are presented in Table 1. As can be seen from this table, in this and all subsequent studies, participants performed the affective maintenance task reasonably well, and were comfortably below ceiling. Likewise, AF performance was on average reasonably accurate but falling below a perfect 10 by approximately 2.5 points.

A multiple linear regression tested whether affective maintenance accuracy and VVIQ scores predicted AF performance. Results indicated this model was significant, $F(2, 63) = 5.47$, p = .006 with an R^2 = .15. As predicted, affect maintenance accuracy contributed significantly to the model, $\beta = 0.32$, p = .008 (see Figure 4). Additionally, VVIQ scores also contributed significantly ($\beta = -0.24$, p = .043) such that better visual imaging abilities predicted higher AF accuracy.

Ancillary Analyses

In an effort to replicate correlations found in previous research, the relationship between AF and EI was assessed using Pearson's r correlation coefficient (see Table 2 for EI scores). Neither the relationship between AF and trait EI, $r(62)$ = .06, p = .62 nor between AF and ability EI were significant, $r(64)$ = –.04, p = .75. Affect maintenance accuracy was also unrelated to either measure of EI (trait: $r(62)$ = –.13, p = .32; ability: $r(64)$ = –.07, p = .55).

Discussion

Results from the first study, while correlational, provide initial support for the hypothesis that AWM plays a role in AF. The positive relationship observed between emotion maintenance ability and AF motivates the question of specificity: Is the association between AWM and AF due to the short-term maintenance of emotional information per se, or does it reflect a more general role for working memory? Therefore, Study 1b examined the relationship between AF and visual working memory using the brightness maintenance task originally designed as the nonaffective analogue of the affect maintenance task. It requires participants to maintain in mind the brightness intensity of an emotionally neutral image over a delay to compare with the brightness of a subsequent neutral

Table 1
Mean (and Standard Deviation) Performance on Key Measures Across Studies

Measures	Study 1a (*N* = 66)	Study 1b (*N* = 68)	Study 2 (*N* = 96)	Study 3 (*N* = 85)
Affective forecasting	7.48 (0.85)	7.35 (0.90)	7.61 (0.89)	7.70 (0.78)
Affect maintenance	0.77 (0.06)	NA	0.81 (0.09)	0.81 (0.08)
Brightness maintenance	NA	0.78 (0.08)	0.85 (0.08)	0.80 (0.08)
VVIQ	38.6 (9.08)	39.4 (9.48)	37.5 (9.85)	38.5 (10.8)

Note. VVIQ = Vividness of Visual Imagery Questionnaire. Range of scores: Affective forecasting: 0–10, Affect maintenance: 0–1, Brightness maintenance: 0–1, and VVIQ: 16–80, where a score approaching 16 reflects the greatest reported imaging ability and a score approaching 80 reflects the poorest abilities.

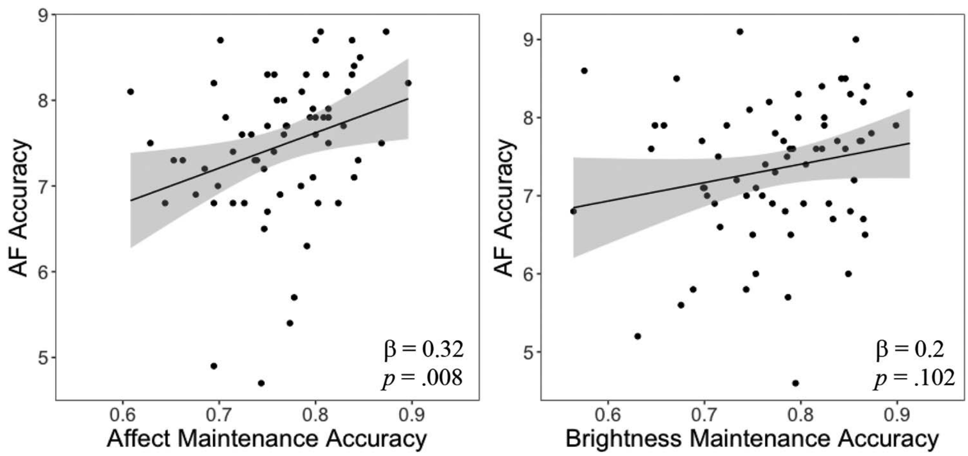

Figure 4. Scatterplots (with best fitting regression line and 95% confidence interval shaded region) showing AF accuracy as a function of maintenance task performance. The graphs show that AF accuracy is significantly predicted by affect maintenance accuracy (left; Study 1a) but not brightness maintenance accuracy (right; Study 1b) using independent samples.

image (Broome et al., 2012; Mikels et al., 2005, 2008). In contrast to affect maintenance, we predicted that the ability to maintain representations of brightness intensity would have no relationship with AF ability. Study 1a also found a correlation between AF and visual imagery as measured by the VVIQ. This measure is included in Study 1b and subsequent studies to assess the reliability of this relationship.

Study 1b: Testing the Relationship Between Visual Working Memory and AF Ability

Method

Participants. A new sample of 78 undergraduate students participated in Study 1b. Sixty-nine participated in the study in exchange for course credit, while nine participated for monetary compensation in order to continue data collection during a university break. Three participants did not return to complete Session 2 and were therefore excluded. In addition, data from five participants were excluded due to poor performance on the brightness maintenance task (see below for details), and one participant was excluded for being under the age of 18 at the time of testing, in accordance with our inclusion criteria. Thus, the present analyses were performed on data from 68 participants (66.2% female, mean age 18.86, 67.6% self-identified White), who self-reported as right-handed and native English speakers. The sample size was selected to parallel Study 1a, overrecruited to account for attrition and low-performing participants. The University of Michigan Institutional Review Board approved all procedures, and all participants provided informed consent prior to participating.

Table 2
Mean (and Standard Deviation) Performance on the Emotional Intelligence Measures

Measures	Study 1a (N = 66)	Study 1b (N = 68)	Study 2 (N = 96)
EI—Ability	.68 (.08)	.68 (.1)	.68 (.07)
EI—Trait	120.8 (8.3)	122.0 (12.3)	123.0 (11.9)

Note. EI = emotional intelligence. Range of Scores: EI–Ability: 0–1, EI–Trait: 33–165.

Design and procedure. Participants completed two 1-hr testing sessions, scheduled 1 week apart (see Figure 1). The experimental design and tasks were the same as Study 1a, except for the use of the brightness maintenance task and corresponding brightness rating task. All tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013).

Brightness maintenance task. In the brightness maintenance task, participants viewed one neutral image (5 s) and were instructed to maintain the brightness intensity actively in mind. A retention interval ensued (3 s) before participants viewed a second neutral image (5 s). Next, a green cross appeared, which prompted participants to report whether the second image had higher or lower brightness intensity than the first one. Brightness intensity was described to participants as the magnitude of overall light or illumination in the image, regardless of the picture's content. Participants responded to each trial by pressing either a key labeled *H* for higher or *L* for lower (see Figure 2). There were 80 trials total where each trial consisted of one pair of two neutral images. We used the same 80 pairs as in Broome et al. (2012). Images were neutral scenes selected from the IAPS (Lang et al., 1999) to ensure that performance would be independent from emotional processes. Intensity ratings from images were obtained previously (N = 40; Pilot Study C, Mikels et al., 2008) and recorded using a 7-point scale that ranged from *low* (1) to *high* (7) brightness intensity, or overall light or illumination of each image. The difference in brightness intensity between images in these pairs, or intensity distance, varied from 0.10 to 2.98 (M = 1.28, SD = .80) and consonant with previous work using this image set, pairs were divided into two groups: near (intensity distance 1.13 or below) and far (intensity distance 1.15 or greater), resulting in 40 near pairs and 40 far pairs. Additionally, for each intensity distance subset, the second picture would have higher intensity than the first for exactly half of the pairs. Furthermore, to establish whether image order affected performance, half of the participants viewed one intensity order and half viewed the same pairs with the image order reversed. This manipulation had no effect on performance and will not be discussed further, $t(66) = 0.59$, $p = .555$. Each order variant of the 80 pairs was divided into sets (Set A and Set B) with equal number of trials for intensity distance (near and far) and intensity order (second image higher or lower intensity) and their presentation at each of the two sessions was counterbalanced across participants. This was intended to minimize performance differences due to the image sets, and proved to be effective, $t(67) = -1.18$, $p = .244$. Within these restrictions, pair order was randomized individually for each participant.

Brightness intensity ratings. At the end of the second session, participants viewed all images from the brightness maintenance task again and rated the "intensity" or magnitude of overall light or illumination they perceived for each one. Ratings were collected after the maintenance task to avoid the potential influence of prior ratings on brightness maintenance performance. Each picture was displayed individually above a visual analog scale that ranged from *not intense* to *extremely intense*. Participants responded by clicking anywhere on the scale below the picture to indicate the intensity of their perceived brightness for each image. The scale had an actual resolution of 21 units and ratings were used to calculate the brightness maintenance accuracy scores (see Brightness Maintenance Scoring below).

Brightness maintenance scoring. Similar to the calculation of affect maintenance accuracy, brightness maintenance accuracy scoring was individualized using each participant's ratings from the brightness intensity-rating task. These subjective ratings were used to infer which image, when appearing with its pair, would be judged by that individual as having higher intensity. A response in the brightness maintenance task was considered correct when it agreed with the participant's relative intensity ratings for the members of each pair. Trials with images receiving equal intensity ratings were excluded from the calculation of that participant's accuracy scores ($M = 7.4\%$ [$SD = 4.0\%$] of all responses). For each participant, accuracy scores were calculated as the sum of the brightness maintenance responses that matched the intensity-based comparisons and divided by the number of included responses. Separate accuracy scores were calculated for each session and then averaged together to calculate a composite brightness maintenance accuracy score. As mentioned above, six participants were excluded from analyses because they performed more than two standard deviations below the mean during either session. These participants performed at approximately 50% accuracy, suggesting they were guessing or otherwise not following instructions.

Results

Descriptive statistics for affective forecasting (AF), brightness maintenance task performance, and vividness of visual imagery (VVIQ) are presented in Table 1. Performance on the brightness maintenance task was reasonable, below ceiling, and well matched with accuracy on the affective maintenance task in the prior study. AF performance was similar to that in Study 1a.

A multiple linear regression was computed to determine if brightness maintenance accuracy and VVIQ scores were significant predictors of AF performance. This model was not significant, $F(2, 65) = 1.34$, $p = .268$ with an $R^2 = .04$, and neither brightness maintenance performance ($\beta = 0.2$, $p = .108$; see Figure 4) nor visual imagery ability ($\beta = -0.007$, $p = .953$) contributed significantly to the model.

Ancillary Analyses

As in Study 1a, we tested the relationships between AF and EI and found that neither the relationship between AF and trait EI, $r(66) = .15$, $p = .21$ nor between AF and ability EI, $r(66) = .14$, $p = .26$ was significant. Brightness maintenance accuracy was not related to trait EI, $r(66) = –.09$, $p = .48$ but brightness maintenance accuracy and ability EI were related, $r(66) = .30$, $p = .01$.

Discussion

As predicted, the results from Study 1b indicate that brightness maintenance ability is not related to AF. Note also that the VVIQ was not associated with AF in this sample. Separately then, the results of Studies 1a and 1b demonstrate that AF ability is selectively related to affect maintenance. However, because these relationships were examined using two different samples, it was not possible to assess the contribution of one maintenance ability while adjusting for the effect of the other. Therefore, the hypothesis that predicting future feelings is specifically related to affective working memory ability was further tested in a second study using both maintenance tasks in a within-subjects design. Study 2 aimed to replicate the specific relationship observed in Study 1 between AF and affect maintenance and allowed us to compare the extent to which performance on each maintenance task predicted AF ability. We expected that only affect maintenance would be a significant predictor of AF ability, whereas brightness maintenance would not.

Study 2: The Contributions of Affective and Visual Working Memory to AF Ability Compared

Method

Participants. A new sample of 110 undergraduate students participated in Study 2 for course credit. One participant chose not to continue in the study after seeing the sample emotional images, two participants did not return for Session 2, and four participants failed to follow instructions, resulting in the exclusion of these seven participants. Due to experimenter error, another participant performed an incorrect version of the task and their data was excluded. Additional exclusions included five participants who did not meet performance inclusion criteria for the maintenance tasks as described in Studies 1a and 1b (see below for details) and one clearly outlying participant who performed more than 3.5 standard deviations below the mean on the AF task (Note: When this 3.5 *SD* criterion for the AF was retroactively applied to Studies 1a and 1b, no outliers were identified). Thus, the present analyses were performed on data from 96 participants (63.5% female, mean age 18.8, 67% self-identified White), who were self-reported right handers and native English speakers. A power analysis based on predictor correlations from Studies 1a and 1b indicated 81 participants would yield 80% power for detecting a small–medium effect at the traditional $\alpha = .05$ criterion of statistical significance (G*Power 3.1: Faul et al., 2009). We oversampled to account for attrition and low-performing individuals. The University of Michigan Institutional Review Board approved all procedures, and all participants provided informed consent prior to participating.

Design and procedure. All participants performed tasks used in Studies 1a and 1b to measure affective maintenance, brightness maintenance, and AF so that the relationships between these constructs could be examined. Participants completed two 1-hr testing sessions, scheduled 1 week apart. The task order used in the prior studies was adjusted in Study 2 to ensure that all assessments could be completed within the 1-hr per session design (see Figure 1). All tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013), and are described, in turn, below.

Brightness maintenance task. In order to maximize efficiency of the brightness maintenance task, we first identified and selected pairs that were most discriminative of brightness maintenance ability. While all pairs previously included in the brightness maintenance task were originally rated as low arousal with intermediate valence (Pilot Study C; Mikels et al., 2008), an additional assessment made by two independent raters excluded appetitive stimuli (e.g., food or drink) and potentially non-neutral content (e.g., lava or lightening). From the remaining 46 pairs, we used an IRT analysis of the data from Study 1b to identify the 28 pairs with the highest discrimination scores, which were presumably the most sensitive to maintenance ability (see the Supplemental Method in the online supplemental materials for full method on IRT analyses). Apart from the number of trials, the brightness maintenance task was the same as Study 1b.

Brightness intensity ratings. Immediately after the maintenance task in the first session, participants viewed all images from this task again and rated their brightness "intensity" or overall magnitude of perceived illumination for each image individually. Due to a programming error, one pair presented during the brightness maintenance task failed to be presented during the rating task, preventing us from scoring this trial. Thus, this pair was dropped when calculating the overall accuracy scores and only performance from the completed 27 pairs was used.

Affective forecasting. As in the previous studies, during the first session participants read descriptions of emotional images and were asked to predict how they would feel if they were to view each image (i.e., Phase I). Then, during the second session a week later, participants saw the images and rated how they actually felt using the same scale (i.e., Phase II). The descriptions, images, protocol, and scoring procedure were identical to those used in Studies 1a and 1b.

Affect maintenance task. In an effort to maximize efficiency of the affect maintenance task, we identified and selected the pairs that were most diagnostic of AWM ability. Using an Item-Response Theory (IRT) analysis, we selected 56 pairs from those used in Study 1a (see Supplemental Method in the online supplemental materials for details). We chose the 28 most discriminative positive pairs and the 28 most discriminative negative pairs using the data from Study 1a. This yielded the subset of pairs that were the most diagnostic of AWM ability for each valence. All pairs for the affect maintenance task were viewed during the second session. Parameters of the task, except for the number of pairs, remained identical to Study 1a.

Emotional intensity ratings. Immediately after the affect maintenance task in Session 2, participants viewed each image from the affect maintenance task again individually and rated the "intensity" or overall magnitude of emotional reaction for each. The design of this task was the same as Study 1a.

Additional measures. Similar to Studies 1a and 1b, participants completed the VVIQ (Marks, 1973), STEU (MacCann & Roberts, 2008), and AES (Schutte et al., 1998) during the first session. Participants also performed the PANAS (Watson et al., 1988) at the beginning and end of both sessions to ensure their emotional states were not negatively influenced by the images viewed in the study. For exploratory reasons we included a gambling task adapted from De Martino, Kumaran, Seymour, and Dolan (2006; see Mikels & Reed, 2009), which was administered at the very end of Session 1, and will not be discussed.

Maintenance task scoring. The accuracy for each task was determined using the individualized scoring procedure described for Studies 1a and 1b. As mentioned above, five participants were excluded from all analyses because they performed more than two standard deviations below the mean on at least one of the maintenance tasks (i.e., two participants met this criterion for the affect maintenance task and three for the brightness maintenance task).

Results

Descriptive statistics for affective forecasting (AF), affect maintenance and brightness maintenance task performance, and vividness of visual imagery (VVIQ) are presented in Table 1. As the means indicate, brightness maintenance accuracy ($M = .85$, $SD = .08$) was slightly but significantly better than affect maintenance accuracy ($M = .81$, $SD = .09$; $p < .05$), a difference that is likely due to pair selection, that we address further below. Because linear regression accounts for predictor variables with different means, we proceeded with our planned tests. Additionally, AF performance was consistent with the previous results.

A multiple linear regression was computed with AF accuracy as the outcome variable and affect maintenance accuracy, brightness maintenance accuracy, and VVIQ scores as predictors. Results indicated that this model was significant, $F(3, 92) = 5.37$, $p = .002$ with an $R^2 = .15$. As predicted, affect maintenance performance contributed significantly to the model ($\beta = 0.33$, $p = .001$) whereas brightness maintenance performance did not ($\beta = 0.16$, $p = .103$; see Figure 5). VVIQ scores did not contribute to the model, $\beta = 0.01$, $p = .89$. In order to examine the relationships between performance on each maintenance task and AF while holding the other constant, we ran partial correlations. The partial correlation between AF and affect maintenance controlling for brightness maintenance was significant ($\rho = .34$, $p < .001$), whereas the partial correlation between AF and brightness maintenance controlling for affect maintenance was not significant ($\rho = .17$, $p = .101$). Additionally, no relationship between performance on the brightness maintenance and affect maintenance tasks was found, $r(94) = .12$, $p = .252$.

Ancillary Analyses

Emotional intelligence associations. Using Pearson's r correlation coefficients, the relationships between AF and EI were assessed and no significant relationships between AF and either

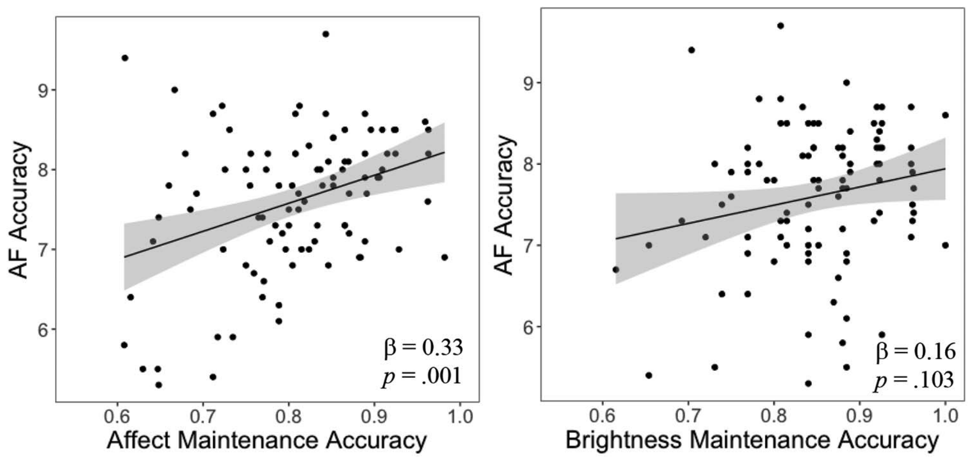

Figure 5. Scatterplots (with best-fitting regression line and shaded region indicating 95% confidence interval) showing AF accuracy as a function of maintenance task performance. The graphs show that affect maintenance accuracy (left) but not brightness maintenance accuracy (right) predicts AF ability using a within-subject design (Study 2).

type of EI (trait: $r(93) = .09$, $p = .383$; ability: $r(94) = .18$, $p = .076$) were found (see Table 2 for EI scores). Affect maintenance accuracy was not related to trait EI, $r(93) = –.07$, $p = .521$ but there was a significant relationship between affect maintenance accuracy and ability EI, $r(94) = .31$, $p = .002$. Brightness maintenance accuracy performance was not related to trait EI, $r(93) = –.08$, $p = .452$ nor ability EI, $r(94) = .02$, $p = .857$.

Combined additional measures results from Studies 1 and 2. To determine whether previously reported findings of a positive association between EI and AF replicated, we included measures of trait and ability EI in Study 1a, Study 1b, and Study 2. Although no significant relationships were found between EI and AF in each study separately ($.04 \leq rs \leq .18$; all $ps \geq .08$), we combined the data sets to create a larger sample ($N = 230$ participants) and repeated the correlations with increased power. AF was not significantly related to either type of EI (trait: $r(225) = .10$, $p = .14$; ability: $r(228) = .11$, $p = .11$). Thus, no overall relationships were found between AF and either type of EI. Additionally, there was no significant overall relationship between VVIQ and AF such that self-reported visual imagery ability was not related to forecasting accuracy, $r(228) = –.03$, $p = .692$.

Discussion

The results from Study 2 replicate and extend the findings from Studies 1a and 1b by demonstrating that emotion maintenance ability predicts AF accuracy whereas brightness maintenance does not. Further, the analyses of the combined data from the three participant samples collected thus far provide no evidence that either measure of EI or visual imagery are related to AF.

Study 3 aimed to build on these results and take the effects a step further, replicating and extending the selective association between affect maintenance and AF once again using methods and analyses that have been preregistered. This next study also included two additional, more widely used measures of visual working memory. Due to their nonaffective nature, we expected them to be unrelated to AF, which would provide further evidence for specificity and strengthen the argument that working memory for emotion plays a unique role in AF. Accordingly, we hypothesized that affect maintenance performance would be a significant predictor of AF ability, whereas performance on the brightness maintenance and other cognitive working memory tasks would not be.

Because of the difference in accuracy between affect and brightness maintenance in Study 2, in Study 3 we made several stimulus substitutions to better equate accuracy on two maintenance tasks. Of note, this unexpected accuracy difference in Study 2 is unlikely to have caused the selective association we observed between affect maintenance and AF because linear regression accounts for predictor variables with different means.

Additionally, because the relationships between AF and EI were found to be unreliable in the prior studies, we opted not to include EI measures in Study 3. Instead, for exploratory purposes, we included a measure of emotion regulation (ER; Gross, 2013). Participants self-reported the tendency to regulate their emotions using (a) cognitive reappraisal, which refers to changing how one appraises internal or external situations in an effort to modify the emotional significance, and (b) expressive suppression, which refers to the inhibition of emotional expression (Gross, 2013). As previously described, Mikels et al. (2008) found that performance was impaired on the affect maintenance task while participants were performing a concurrent ER (cognitive reappraisal) task. This interference may be indicative of a shared underlying process. Therefore, we predicted a positive relationship between AWM and ER, specifically cognitive reappraisal.

Study 3: Replicating and Extending the Selective AWM–AF Association

Method

Participants. A new sample of 96 young adults participated in exchange for course credit or payment. Five participants failed to follow instructions or complete the study and six were excluded due to poor performance on the maintenance tasks (see below for details). Thus, the reported analyses were performed on data from 85 participants (72.6% female, mean age 18.7 ($SD = 0.90$), 70.0% White), who were self-reported right handers and native English speakers. A power analysis based on Study 2 indicated that, in order to detect a medium–large effect with five predictors and 80% power at the traditional $\alpha = .05$ criterion of statistical significance, we needed at least 79 participants (G*Power 3.1). We over-recruited for anticipated attrition and low-performing individuals. All experimental procedures were approved by the University of Michigan Institutional Review Board. The methods and analyses were preregistered and can be found at https://aspredicted.org/nw3pv.pdf (Note: revised version available at https://aspredicted.org/ik28c.pdf).

Design and procedure. Similar to Study 2, all participants completed two 1-hr testing sessions scheduled 1 week apart that included the affect maintenance, brightness maintenance, and AF tasks (see Figure 1). Additionally, participants performed two visual working memory tasks, split between weeks to minimize mental fatigue (Session 1: Corsi block-tapping; Session 2: visual change detection). At the end of Session 1, participants also completed a self-report measure of emotion regulation strategies. Tasks were performed on a PC desktop computer with E-Prime (Version 2.0.10, 2013), and are described, in turn, below.

Maintenance tasks. In Study 3 we made several substitutions to the stimulus set in an effort to better equate overall performance on the two maintenance tasks. Our prior approach was to select the top most discriminative pairs for the affect and brightness tasks respectively (see Supplemental Method in online supplemental materials). Consequently, the brightness pairs were inadvertently more discriminative than the affect pairs. For Study 3, we selected again the most discriminative pairs for each domain, with the constraint that the average discrimination index across the pairs for each task would be statistically equivalent. This led to a change in 11 brightness pairs and 4 affect pairs. In all respects, the task parameters remained identical to Study 2.

Rating tasks. Participants performed the corresponding rating task after each maintenance task. For emotional intensity ratings, participants viewed all images from the affect maintenance task again individually and rated the "emotional intensity", or overall magnitude of emotional reaction for each. For the brightness ratings, participants viewed all images from the brightness maintenance task again individually and rated the perceived "brightness intensity", or overall magnitude of light or illumination of the image. The task design was identical to Studies 1 and 2.

Affective forecasting. The AF task was identical to the one previously used in Studies 1a, 1b, and 2.

Corsi block-tapping task. To assess visuospatial working memory, we employed a computerized version of forward and backward Corsi block-tapping task (Corsi, 1972). A subset of nine white blocks turn red sequentially on the computer screen in a particular sequence. In the task, participants then repeat the sequence by clicking on the squares in the same (forward) or reverse (backward) order. Set size (i.e., sequence length) increases from three to nine squares with three trials for each set size. The task is discontinued when the participant answers all three trials of a given set size incorrectly. Scores were calculated as the product of the largest set size attempted and the total number of correctly reproduced sequences, calculated separately for backward and forward versions of the task. These two scores were then standardized and averaged to obtain one composite Corsi score.

Visual change detection. A computerized delayed match-to-sample, visual change detection task (Luck & Vogel, 1997) served as an additional measure of visual working memory. In this task, participants view a sample array of two to ten variously colored squares (500 ms). The squares then change to reveal variants of a horizontal stripe pattern that includes equal parts of the six possible colors the squares can be (500 ms). A test probe then appears and participants must indicate whether the square in that position is the same color it was in the initial sample array. Cowan's capacity score (K) is calculated using the formula (hit rate + correct rejection rate – 1, for each set size) $\times N$, where N is set size (Cowan, 2001; Rouder, Morey, Morey, & Cowan, 2011). To avoid problematic averaging as advised in Rouder et al. (2011), we excluded subspan capacities (i.e., $N = 2$) and negative K scores were changed to zero before averaging across set sizes to create a composite K score for each participant.

Emotion Regulation Questionnaire (ERQ). To explore potential relationships between emotion regulation, AWM, and AF, participants completed the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003). The ERQ measures the tendency of participants to regulate their emotions using Cognitive Reappraisal and Expressive Suppression, two ER strategies. Participants read 10 statements regarding how individuals regulate and manage their emotions and rated the extent to which they agreed with each one on a scale from *strongly disagree* (1) to *strongly agree* (7). Scores for Cognitive Reappraisal (6 items) and Expressive Suppression (4 items) were calculated by following the standard procedure of averaging the items in each category.

Additional measures. Similar to Studies 1 and 2, participants completed the VVIQ (Marks, 1973). The Positive Affect and Negative Affect Scale (PANAS) was again administered before and after each session to ensure mood was not negatively affected by the study and was not analyzed further (Watson et al., 1988).

Maintenance task scoring. The accuracy for each task was determined using the individualized scoring procedure described for Studies 1 and 2. As mentioned above, six participants were excluded from all analysis due to performing 2 or more standard deviations below the mean on affective or brightness maintenance tasks. These scores were approximately 50% accuracy, indicating chance-level performance.

Results and Discussion

Descriptive statistics for AF, affective maintenance, brightness maintenance, and the VVIQ are provided in Table 1. The means make evident that performance on the two maintenance tasks was well-matched, and AF performance continued to be relatively stable in this fourth population.

We first examined the correlations among the visual working memory tasks. Brightness maintenance performance did not correlate significantly with change detection performance, $r(82) = 0.13$, $p = .222$, and the relationship with Corsi performance was only a trend, $r(83) = 0.19$, $p = .077$. In contrast, performance on the Corsi and change detection tasks were significantly related, $r(82) = 0.29$, $p = .008$. These results suggest that the brightness maintenance task measures an ability that is dissimilar from that measured by these two conventional measures of visual working memory. Affect maintenance accuracy did not correlate with either Corsi, $r(83) = .009$, $p = .932$ or change detection performance, $r(82) = -.002$, $p = .986$. As in Study 2, brightness maintenance accuracy and affect maintenance accuracy were unrelated, $r(83) = 0.12$, $p = .258$.

A multiple linear regression was then computed with AF as the outcome variable and performance on affect maintenance, brightness maintenance, Corsi, change detection, and VVIQ as predictors. The model was significant, $F(5, 78) = 2.69$, $p = .027$ with an $R^2 = .15$. Affect maintenance accuracy contributed significantly to the model, predicting AF ability ($\beta = 0.25$, $p = .022$), while brightness maintenance did not ($\beta = -.03$, $p = .802$; see Figure 6). VVIQ scores also contributed significantly, such that greater imaging abilities indicated greater AF accuracy ($\beta = -.21$, $p = .047$). Additionally, performance on the visual change detection task ($\beta = -.22$, $p = .049$) was also a significant (negative) predictor of AF performance, such that higher capacity scores were predictive of less accurate forecasts. Performance on the Corsi was not a significant predictor of AF, $\beta = .13$, $p = .226$.

The unexpected negative association between change detection performance and AF ability in the multiple linear regression led us to further examine the change detection predictor variable. While a multiple linear regression calculates the variance explained by each predictor holding the other predictors constant, this is only possible when the predictors in the model are orthogonal, or unrelated, to one another (Kutner, Nachtscheim, Neter, & Li, 2005). Therefore, correlations greater than zero between predictor variables make it impossible to determine the true unique effect of each predictor on the outcome variable due to multicollinearity, even when the simple correlations between predictors are relatively small (Cohen, Cohen, West, & Aiken, 2003). Thus, the significant correlation between Corsi and change detection performance may contribute to this unexpected effect.

To assess the relationship between AF ability and change detection without other predictors, we performed a zero-order correlation and the results were not significant, $r(82) = -.19$, $p = .088$. Because of the difference in predictive value with and without the presence of other variables, it may be that the Corsi, or another variable, acted as a moderator on the relationship between change detection and AF, increasing the effect in its presence (Cohen et al., 2003). Nevertheless, these findings suggest that the change detection variable only predicts AF ability when included in a model with other variables and may not uniquely contribute to AF ability as originally revealed in the regression.

Further support that affect maintenance is uniquely related to AF comes from partial correlations. The relationship between AF and affect maintenance controlling for brightness maintenance was significant ($\rho = .23$, $p = .036$), whereas the relationship between AF and brightness maintenance controlling for affect maintenance was not ($\rho = -.03$, $p = .789$). The relationship between AF and affect maintenance also remained significant while controlling for Corsi ($\rho = .23$, $p = .037$) and change detection ($\rho = .23$, $p = .035$) performance.

Exploratory Analyses

To determine whether use of emotion regulation strategies were related to AF and affect maintenance performance, Pearson's r correlation coefficients were computed. AF was significantly related to use of cognitive reappraisal, $r(83) = 0.29$, $p = .007$ but not expressive suppression, $r(83) = .021$, $p = .845$. Affect maintenance was not significantly related to cognitive reappraisal, $r(83) = .06$, $p = .601$ nor expressive suppression, $r(83) = .09$, $p = .422$.

General Discussion

These studies tested the hypothesis that AWM contributes to the ability to make accurate forecasts about future feelings. Across three independent samples of participants, performance on an affect maintenance task reliably and consistently predicted AF performance, supporting this hypothesis. Furthermore, three studies also demonstrated that working memory for brightness intensity was not predictive of AF, suggesting that working memory for

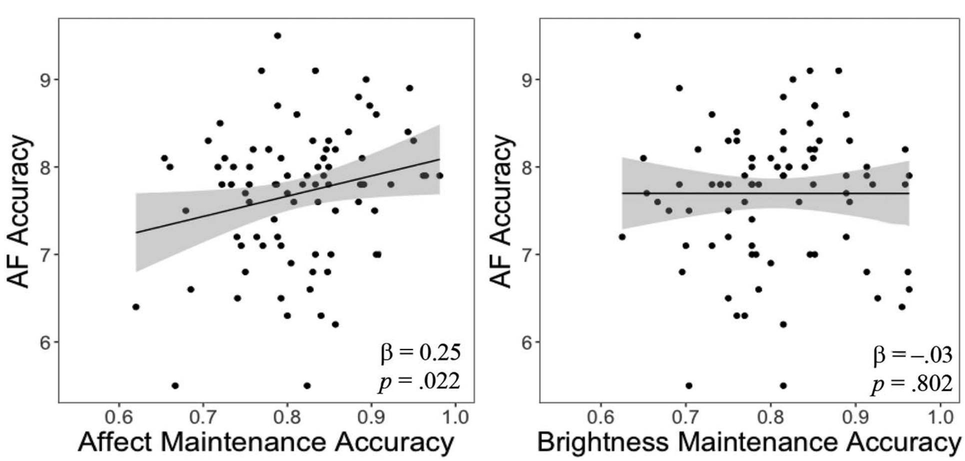

Figure 6. Scatterplots (with best-fitting regression line and shaded region indicating 95% confidence interval) showing affective forecasting (AF) accuracy as a function of maintenance task performance. Study 3 results show that affect maintenance accuracy (left) but not brightness maintenance accuracy (right) predicts AF ability, replicating the findings from Studies 1 and 2.

emotion intensity plays a unique role in AF. We consider the implications of these results and several other findings for understanding the processes that underlie AF, as well as the implications for AWM and its role in mental processes more generally. We also consider the finding that EI was not related to AF or AWM. Finally, we discuss several limitations of the present investigations and potential ways to address them.

Implications for Affective Forecasting

We posit that the selective relationship we observed between AWM and AF is due to the specific processes needed to hold in mind and reflect on emotional experiences. This interpretation is supported by the finding that the ability to maintain intensity information about a nonaffective subjective state, that is, brightness, had no bearing on AF ability. Our hypothesis that AWM plays an important role in AF is based on the idea that predicting future feeling states involves conjuring up, working with, and comparing emotional experiences—all processes hypothesized to require AWM (Mikels & Reuter-Lorenz, 2013, 2019; Smith & Lane, 2015). Our results support this hypothesis by demonstrating that individuals who are better able to hold an emotional feeling in mind for subsequent comparison, are better able to predict the intensity of their future feelings.

We included a measure of visual imagery in all studies because a potential strategy for completing the current AF task is to create a visual image of the verbally described emotional scene. It seemed plausible then that people with better and more vivid imagery would perform better on the AF task. Instead, we found that performance on the VVIQ was an inconsistent predictor of AF performance across studies. Combining the data from all four studies revealed no overall relationship between the VVIQ and AF performance, $r(313) = -.06$, $p = .223$, indicating that imagery ability does not facilitate forecasting accuracy at least using these particular measures. Alternatively, the verbal description on each AF trial may be processed directly to evoke the emotional feelings that are then rated for intensity, without the need to generate an image.

For exploratory purposes, the last study included a measure of emotion regulation (ER). Whereas the hypothesized relationship with AWM was not evident, people reporting more frequent use of cognitive reappraisal also turned out to be better forecasters. This is an interesting and novel result that requires replication. Nevertheless, the relationship suggests an interaction between forecasting and regulation processes. This is consistent with a proposal by Loewenstein (2007), who posited that participants select ER strategies based on their forecasted effectiveness and that ER success is dependent on the accuracy of these forecasts. Additionally, Ringnes, Stalsett, Hegstad, and Danbolt (2017) discovered that some individuals use prospects of future emotions to regulate their current affective states, an ER strategy they called "emotional forecasting." While the extant findings provide converging evidence for an ER–AF interaction, subsequent research should examine how individuals use their forecasts to inform regulation strategies and vice versa.

Visual working memory did not predict AF performance when measured by the brightness maintenance task across all studies, nor did Corsi block performance in Study 3. The relationship between AF and change detection is less clear. Better change detection performance was associated with worse forecasting ability in a multiple linear regression but not in an independent correlation. If change detection and AF performance are truly related, it may be that people with lower visual working memory capacities rely on different, more effective, strategies to forecast their feelings. However, given that this relationship was anomalous among the measures of visual working memory used here, and that the negative association between change detection and AF was only significant in the presence of other predictors likely reflecting either a moderating effect or collinearity, we are hesitant to conclude that the regression results reflect a true association. Future research using these measures will be needed to resolve these uncertainties.

Implications for Affective Working Memory

The present results indicating that AF is related to affective but not brightness maintenance also provide additional support for the idea that AWM is a separable, domain-specific subsystem of working memory. These results complement and extend prior evidence for separability of affect and brightness maintenance processes based on selective interference methodology, age differences, and special patient populations (Mikels et al., 2005, 2008; Gard et al., 2011). Working memory is known to be a fundamental capacity that is integral to higher-order cognition. The present evidence that AWM supports AF ability suggests AWM may play a comparable role as a crucial facet of higher-order emotion-related mentation.

The relationship we found between AWM and AF has practical implications as well. Previous reports suggest that AF may be improved by interventions. For example, by having participants think about multiple past experiences (Buehler & McFarland, 2001) or several features of the target outcome/event when making forecasts (Lam, Buehler, McFarland, Ross, & Cheung, 2005), predictions are less likely to be extreme. The present findings that AWM ability contributes to individual variability in AF suggest that improving AWM could potentially enhance forecasting abilities. Training can improve working memory performance in the cognitive domain (e.g., Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; see also Redick, 2019), and there are indications that the same may be true in the affective domain (e.g., Leone de Voogd, Wiers, Zwitser, & Salemink, 2016). Thus, future work aimed at training AWM could include measures to assess possible benefits to affective forecasting as well.

We also hypothesized AWM and emotion regulation, namely cognitive reappraisal, may share an underlying mechanism due to impaired performance on the affect maintenance task when performing a concurrent cognitive reappraisal task (Mikels et al., 2005; see also Gard et al., 2011). Contrary to our hypothesis, no significant relationship was found between performance on affect maintenance and the self-reported tendency to use cognitive reappraisal. Because the results across the three present studies demonstrate that AWM is predictive of AF ability, and AF ability was found to be related to cognitive reappraisal, it is possible the relationship between AWM and emotion regulation is not direct. Future research should investigate the possible role that AWM may play in cognitive reappraisal and how it may influence ER or other types of higher order emotional processing. Additionally, while a self-report measure for ER is convenient, it may be more informative to examine differences in performance-based cognitive reappraisal success between individuals to determine direct associations.

Relations With Emotional Intelligence

We included EI measures in the current study to reassess the relationships between AF and EI, and to explore the potential relationship between AWM and EI. Across the current studies, we found no evidence for relationships between AF and either trait or ability EI. These results are inconsistent with previous reports that AF is significantly related to ability EI (Dunn et al., 2007; Hoerger et al., 2012) and trait EI (Hoerger et al., 2012). Methodological differences among these studies may contribute to these discrepant results. For both constructs, Dunn et al. (2007) used measures different from those used in the present investigation, measuring AF by comparing predicted and actual reactions to real events (e.g., elections or college sporting events) and measuring ability EI using a composite and emotion management subscore from the Mayer-Salovey-Caruso Emotional Intelligence Test (MS-CEIT; Mayer, Salovey, & Caruso, 2002). These differed from the present verbal-description-based measure of AF and the Situational Test of Emotional Understanding, which focuses on the *understanding emotion* component, rather than the *managing emotion* component of EI. Nevertheless, even using a similar AF measure and identical EI task, we were unable to replicate the Hoerger et al. (2012) results. This discrepancy may stem from other prominent differences in study designs. In Hoerger et al. (2012) participants completed the entirety of the AF task online, rather than in-person. Moreover, in Hoerger et al. (2012, Study 2) participants performed the feelings rating portion of the AF task immediately after the prediction rating with an average delay period of only 3 minutes, and both used a 9-point rating scale. With this brief intervening interval, memory of prior numerical ratings could influence the assessments and ratings of actual feeling states. In contrast, our participants used a visual analog scale for the rating tasks and had a 1-week interval between Phase I (i.e., prediction rating) and Phase II (i.e., experienced feelings rating), design features that are likely to mitigate reliance on memory for the original predictions. These inconsistent findings suggest the need for additional research to understand the relationship between AF and EI.

Moreover, we failed to find consistent associations between EI and AWM. If AWM constitutes a fundamental emotion processing capacity, then one would expect it to be related to both superior EI and more successful AF. However, we found that AWM was unrelated to trait or ability EI. The failure to find relationships between AWM and EI despite finding a consistent relationship between AWM and AF may be due to the differences in the way the constructs were assessed. The measures for AWM and AF were both task-based and used the participants' own responses to determine accuracy. In contrast, the measures of EI are not task-based and accuracy for ability EI using the STEU is determined using Roseman's Appraisal Theory (MacCann & Roberts, 2008). Future research that tests these associations using another measure such as the MSCEIT, which calculates accuracy using consensus or expert scoring (Mayer et al., 2002), would allow us to further investigate the generality of these findings.

Limitations and Future Directions

Our study had several limitations that should be addressed in future work. We observed a strong and consistent relationship between AF and AWM. Yet, this association is based on a single measure of AF ability that uses predicted intensity ratings of text describing emotional scenes. While this measure has been used in several prior investigations of AF, an important future direction is to establish that the association we have documented between AWM and AF is generalizable to measures of AF with greater ecological validity. Future studies should investigate this relationship using forecasted feelings for real-life events, which are more likely to approximate the use of AF in everyday life. Our measure of AF was also limited to ratings of anticipated emotional intensity, which aligns well with our intensity-based measure of AWM. A recent model of AF proposed that other aspects of predictions about ones' future feelings, such as the effect of an event on ones' mood, may be more biased or error prone than intensity (Lench et al., 2019). We expect that AWM would play a role in these estimates as well; however, this remains to be explored.

Other potential limitations are due to the size and age restrictions of our sample. We acknowledge that our sample sizes in each of the four studies were moderate due to practical limitations of in-person laboratory testing. However, our sample sizes were determined by a priori power analyses and are typical compared to other studies using in-person measures of AF (see Buehler & McFarland, 2001; Dunn et al., 2007; Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998). In general, in-person testing adds experimental control, consistency, and retention advantages that are less characteristic of Internet testing, but also place practical limitations on sample size. Developing an Internet version in the future for more widespread use of the tasks would permit a larger and more diverse sample. Additionally, because of age-related effects on emotional processing (Mather, 2016; Mikels et al., 2005), our results may not generalize to other age groups, especially older adults. Given the role of AF in decision making (Loewenstein, 2007; Mellers & McGraw, 2001) and the life-altering financial and health care decisions faced by older adults, future studies should take a life span developmental approach to determine if and when the relationship between AWM and AF changes.

Finally, a potential caveat concerns the concurrent validity of the brightness maintenance task. The demands of the brightness maintenance task were well-suited to juxtapose the affect maintenance task, where the emotional intensity attribute is held in working memory. However, brightness maintenance was only weakly and nonsignificantly related to Corsi and change detection performance. The limited associations between brightness maintenance and these two canonical visual working memory tasks could reflect that the brightness maintenance task measures the ability to accurately retain an attribute of a visual image, namely intensity, rather than quantities or sequences of visual features (e.g., colors or locations). Thus, while the concurrent validity of the brightness maintenance task remains to be further clarified, affect and brightness maintenance abilities do appear separable (as evidenced by their lack of correlation and the selective relationship between AF and maintenance of emotional, but not brightness, intensity). Furthermore, the relationship between affect maintenance and AF remained unaffected when controlling for brightness maintenance and either of the two canonical measures of visual working memory.

Conclusion

In conclusion, despite considerable extant research on AF, the contributions of an elemental system such as AWM have not been previously considered. The present results indicate that unlike other measures of working memory, affect maintenance ability reliably predicts AF ability. We conclude that AWM is a core ability that supports affective prospection, and may also underlie other forms of higher-order emotional thought. For example, AWM may provide the mental work space for episodic counterfactual thinking about past emotional experiences (De Brigard & Parikh, 2019), or other forms of mental simulation where feeling states play a prominent role. The present work suggests that individual differences in AWM ability may be important to the quality and accuracy of emotion-related thought.

Context

Affective working memory (AWM) is a putative core mental ability for higher-order emotion-related thought. Because affective forecasting (AF) involves judgments about feelings that are generated, actively maintained in mind, and evaluated, processes that we and others attribute to AWM, we hypothesized that AWM ability would contribute to the accuracy of affective forecasts. That is, if someone is better able to conjure and hold emotional feeling states in mind, they would be more accurate in reflecting on and predicting their future feelings. While much of the affective forecasting literature focuses on errors due to bias, our work focuses on individual differences due to working memory abilities. The current findings support the hypothesis that AWM plays a unique role in AF ability adding credence to the proposal that AWM contributes to higher-order emotional thought. In future work, we plan to assess this relationship further using other forecasting tasks, and extend this research to older adults to determine how the AWM–AF relationship changes with age.

References

Andrews, F. M., & Robinson, J. P. (1991). Measures of subjective well-being. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 61–114). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-590241-0.50007-1

Baddeley, A. (1992). Working memory. *Science, 255,* 556–559. http://dx.doi.org/10.1126/science.1736359

Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198528012.001.0001

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

Broome, R., Gard, D. E., & Mikels, J. A. (2012). Test–retest reliability of an emotion maintenance task. *Cognition and Emotion, 26,* 737–747. http://dx.doi.org/10.1080/02699931.2011.613916

Buehler, R., & McFarland, C. (2001). Intensity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin, 27,* 1480–1493. http://dx.doi.org/10.1177/01461672012711009

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34,* 819B.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24,* 87–114.

De Brigard, F., & Parikh, N. (2019). Episodic counterfactual thinking. *Current Directions in Psychological Science, 28,* 59–66. http://dx.doi.org/10.1177/0963721418806512

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313,* 684–687. http://dx.doi.org/10.1126/science.1128356

de Voogd, E. L., Wiers, R. W., Zwitser, R. J., & Salemink, E. (2016). Emotional working memory training as an online intervention for adolescent anxiety and depression: A randomised controlled trial. *Australian Journal of Psychology, 68,* 228–238. http://dx.doi.org/10.1111/ajpy.12134

Dunn, E. W., Brackett, M. A., Ashton-James, C., Schneiderman, E., & Salovey, P. (2007). On emotionally intelligent time travel: Individual differences in affective forecasting ability. *Personality and Social Psychology Bulletin, 33,* 85–93. http://dx.doi.org/10.1177/0146167206294201

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160. http://dx.doi.org/10.3758/BRM.41.4.1149

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

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75,* 617–638. http://dx.doi.org/10.1037/0022-3514.75.3.617

Gross, J. J. (Ed.). (2013). Emotion regulation: Conceptual and empirical foundations. *Handbook of emotion regulation* (2nd ed., pp. 1–20). New York, NY: Guilford Press.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85,* 348–362. http://dx.doi.org/10.1037/0022-3514.85.2.348

Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice, 8,* 35–41. http://dx.doi.org/10.1111/j.1745-3992.1989.tb00313.x

Hoerger, M., Chapman, B., & Duberstein, P. (2016). Realistic affective forecasting: The role of personality. *Cognition and Emotion, 30,* 1304–1316. http://dx.doi.org/10.1080/02699931.2015.1061481

Hoerger, M., Chapman, B. P., Epstein, R. M., & Duberstein, P. R. (2012). Emotional intelligence: A theoretical framework for individual differences in affective forecasting. *Emotion, 12,* 716–725. http://dx.doi.org/10.1037/a0026724

Hughes, D. J., & Evans, T. R. (2018). Putting 'emotional intelligences' in their place: Introducing the integrated model of affect-related individual differences. *Frontiers in Psychology, 9,* 2155. http://dx.doi.org/10.3389/fpsyg.2018.02155

JASP Team. (2018). JASP (Version 0.9) [Computer software]. Retrieved from https://jasp-stats.org/previous-versions/

Kutner, M. H., Nachtscheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). New York, NY: McGraw-Hill.

Lam, K. C., Buehler, R., McFarland, C., Ross, M., & Cheung, I. (2005). Cultural differences in affective forecasting: The role of focalism. *Per-*

sonality and Social Psychology Bulletin, 31, 1296–1309. http://dx.doi.org/10.1177/0146167205274691

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Technical manual and affective ratings*. Gainesville, FL: University of Florida, Center for Research in Psychophysiology.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 114,* E2016–E2025. http://dx.doi.org/10.1073/pnas.1619316114

Lench, H. C., Levine, L. J., Perez, K., Carpenter, Z. K., Carlson, S. J., Bench, S. W., & Wan, Y. (2019). When and why people misestimate future feelings: Identifying strengths and weaknesses in affective forecasting. *Journal of Personality and Social Psychology, 116,* 724–742. http://dx.doi.org/10.1037/pspa0000143

Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. A. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology, 103,* 584–605. http://dx.doi.org/10.1037/a0029544

Loewenstein, G. (2007). Affect regulation and affective forecasting. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 180–203). New York, NY: Guilford Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8,* 540–551. http://dx.doi.org/10.1037/a0012746

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17–24. http://dx.doi.org/10.1111/j.2044-8295.1973.tb01322.x

Mather, M. (2016). The affective neuroscience of aging. *Annual Review of Psychology, 67,* 213–238. http://dx.doi.org/10.1146/annurev-psych-122414-033540

Mathieu, M. T., & Gosling, S. D. (2012). The accuracy or inaccuracy of affective forecasts depends on how accuracy is indexed: A meta-analysis of past studies. *Psychological Science, 23,* 161–162. http://dx.doi.org/10.1177/0956797611427044

Mayer, J. D., Salovey, P., & Caruso, D. (2002). *The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), Version 2.0*. Toronto, Canada: Multi-Health Systems.

Mayer, J., Salovey, P., & Caruso, D. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15,* 197–215. http://dx.doi.org/10.1207/s15327965pli1503_02

Mellers, B. A., & Mcgraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science, 10,* 210–214. http://dx.doi.org/10.1111/1467-8721.00151

Mikels, J. A., Larkin, G. R., Reuter-Lorenz, P. A., & Carstensen, L. L. (2005). Divergent trajectories in the aging mind: Changes in working memory for affective versus visual information with age. *Psychology and Aging, 20,* 542–553. http://dx.doi.org/10.1037/0882-7974.20.4.542

Mikels, J. A., & Reed, A. E. (2009). Monetary losses do not loom large in later life: Age differences in the framing effect. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 64B,* 457–460. http://dx.doi.org/10.1093/geronb/gbp043

Mikels, J. A., & Reuter-Lorenz, P. A. (2013). Emotion and working memory. In H. Pashler (Ed.), *Encyclopedia of the mind* (pp. 308–310). Thousand Oaks, CA: SAGE. http://dx.doi.org/10.4135/9781452257044.n113

Mikels, J. A., & Reuter-Lorenz, P. A. (2019). Affective working memory: An integrative psychological construct. *Perspectives on Psychological Science, 14,* 543–559. http://dx.doi.org/10.1177/1745691619837597

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

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15,* 425–448.

Redick, T. S. (2019). The hype cycle of working memory training. *Current Directions in Psychological Science, 28,* 423–429. http://dx.doi.org/10.1177/0963721419848668

Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139,* 5–21. http://dx.doi.org/10.1016/j.neuroscience.2005.12.061

Ringnes, H. K., Stalsett, G., Hegstad, H., & Danbolt, L. J. (2017). Emotional forecasting of happiness. *Archive for the Psychology of Religion, 39,* 312–343. http://dx.doi.org/10.1163/15736121-12341341

Robinson, M. D., & Clore, G. L. (2001). Simulation, scenarios, and emotional appraisal: Testing the convergence of real and imagined reactions to emotional stimuli. *Personality and Social Psychology Bulletin, 27,* 1520–1532. http://dx.doi.org/10.1177/01461672012711012

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review, 18,* 324–330. http://dx.doi.org/10.3758/s13423-011-0055-3

Scheibe, S., Mata, R., & Carstensen, L. L. (2011). Age differences in affective forecasting and experienced emotion surrounding the 2008 U.S. presidential election. *Cognition and Emotion, 25,* 1029–1044. http://dx.doi.org/10.1080/02699931.2010.545543

Schutte, N. S., Malouff, J. M., & Bhullar, N. (2009). The Assessing Emotions Scale. In J. Parker, D. Saklofske, & C. Stough (Eds.), *Assessing emotional intelligence* (pp. 119–134). Boston, MA: Springer. http://dx.doi.org/10.1007/978-0-387-88370-0_7

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences, 25,* 167–177. http://dx.doi.org/10.1016/S0191-8869(98)00001-4

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283,* 1657–1661. http://dx.doi.org/10.1126/science.283.5408.1657

Smith, R., & Lane, R. D. (2015). The neural basis of one's own conscious and unconscious emotional states. *Neuroscience and Biobehavioral Reviews, 57,* 1–29. http://dx.doi.org/10.1016/j.neubiorev.2015.08.003

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review, 24,* 1077–1096. http://dx.doi.org/10.3758/s13423-016-1217-0

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology, 35,* 345–411. http://dx.doi.org/10.1016/S0065-2601(03)01006-2

Wilson, T. D., & Gilbert, D. T. (2013). The impact bias is alive and well. *Journal of Personality and Social Psychology, 105,* 740–748.

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78,* 821–836. http://dx.doi.org/10.1037/0022-3514.78.5.821

Received September 14, 2019
Revision received February 13, 2020
Accepted March 20, 2020 ■