



# HHS Public Access

Author manuscript

*Emotion*. Author manuscript; available in PMC 2020 February 01.

Published in final edited form as:

*Emotion*. 2019 February ; 19(1): 146–159. doi:10.1037/emo0000431.

## The Semantics of Emotion in False Memory

**C. J. Brainerd** and

Institute for Human Neuroscience and Department of Human Development, Cornell University

**S. H. Bookbinder**

Institute for Human Neuroscience and Department of Human Development, Cornell University

### Abstract

The emotional valence of target information has been a centerpiece of recent false memory research, but in most experiments, it has been confounded with emotional arousal. We sought to clarify the results of such research by identifying a shared mathematical relation between valence and arousal ratings in commonly administered normed materials. That relation was then used to (a) decide whether arousal as well as valence influences false memory when they are confounded and to (b) determine whether semantic properties that are known to affect false memory covary with valence and arousal ratings. In Study 1, we identified a quadratic relation between valence and arousal ratings of words and pictures that has two key properties: Arousal increases more rapidly as function of negative valence than positive valence, and hence, a given level of negative valence is more arousing than the same level of positive valence. This quadratic function predicts that if arousal as well as valence affects false memory when they are confounded, false memory data must have certain fine-grained properties. In Study 2, those properties were absent from norming data for the Cornell-Cortland Emotional Word Lists, indicating that valence but not arousal affects false memory in those norms. In Study 3, we tested fuzzy-trace theory's explanation of that pattern: that valence ratings are positively related to semantic properties that are known to increase false memory, but arousal ratings are not.

### Keywords

emotional valence; emotional arousal; false memory; fuzzy-trace theory

---

Whether and how emotion distorts episodic memory have long been topics of keen interest in the false memory literature (e.g., see Loftus, 1993; Stein, Ornstein, Tversky, & Brainerd, 1997; Storbeck, 2013). The original impetus came from the domain that first spawned interest in the study of false memory itself—namely, the reliability of legal evidence (e.g., Loftus, 1975). Two hallmarks of criminal cases are that (a) the bulk of the evidence that bears on innocence and guilt comes from memory reports that are given by witnesses during attorney and police interviews, eyewitness identifications, depositions, and courtroom testimony, and (b) the events that are being remembered are affect laden. Because criminal prosecutions rely so heavily on memory reports, false memories can have serious consequences in such cases, and it is natural to wonder whether their incidence is affected by

the fact that witnesses are retrieving emotional content. Here, two contradictory hypotheses have been proposed, prevention and distortion, both of which have been discussed in expert scientific testimony (Bookbinder & Brainerd, 2016).

The prevention hypothesis echoes Samuel Johnson's familiar quip, "When a man knows he is to be hanged in a fortnight, it concentrates his mind wonderfully." As Laney and Loftus (2010) observed in a review of jurors' perceptions of testimony, this hypothesis stipulates that emotional content inoculates memory against distortion, and that when emotion is so intense as to be traumatic, it becomes virtually impossible to develop false memories of target events. This proposal has figured prominently in the defense of individuals (e.g., police investigators, psychotherapists) who are accused of creating false memories of traumatic experiences (e.g., witnessing or committing a robbery, being sexually abused or committing abuse) in plaintiffs and defendants (for a review, see Brainerd & Reyna, 2005). It has also been used to rehabilitate evidence that has been provided by witnesses with memory limitations (e.g., young children, mentally disabled individuals) when they testify about traumatic events (Brainerd & Reyna, 2005).

In contrast, the distortion hypothesis specifies that emotional content stimulates false memories of events, relative to neutral content, and that false memories multiply as emotional content becomes more intense. This hypothesis is grounded in experiments demonstrating that people report experiencing intensely emotional events that they did not in fact experience, such as committing a major crime (Shaw & Porter, 2015), being hospitalized for serious injuries (Garry, Manning, Loftus, & Sherman, 1996), and being sexually abused (Spanos, 1996; Spanos, Cross, Dickson, & DuBreuil, 1993). In criminal cases, memory distortion pursuant to the emotional content of experience has been used to explain why some witnesses express high confidence in memories that are demonstrably false, either because the events are ruled out by physical evidence (e.g., DNA) or are too bizarre to be credible (Appelbaum, Uyehara, & Elin, 1997; Kassin, Ellsworth, & Smith, 1989; Loftus & Ketcham, 1996).

Outside the legal sphere, the question of how emotional content influences false memory has broad ramifications for a range of high stakes remembering situations, such as terrorist interrogations, patient medical reports in emergency rooms, combatant accounts of battlefield action, and client histories taken during psychotherapy (Brainerd & Reyna, 2005). The published archive of experimentation on this question was recently reviewed by Bookbinder and Brainerd (2016). In the research that they reviewed, the emotional valence of target items (negative, neutral, positive) was manipulated, and subsequent levels of true and false memory were measured for different levels of valence. In order to achieve rigorous control of valence, target items either were drawn from emotional word norms, such as the affective norms for English words (ANEW; Bradley and Lang, 1999), or they were drawn from emotional picture norms, such as the international affective picture system (IAPS; Lang, Bradley, & Cuthbert, 2008). The key feature of these norms is that groups of subjects rate large numbers of words or pictures for their levels of valence and arousal. Following Lang et al., valence has traditionally been rated by assigning individual items a number between 1 and 9 on an unhappy-to-happy scale, and arousal has traditionally been rated by

assigning individual items a number between 1 and 9 on a calm-to-excited scale. Other scales (e.g., 0–100%) have also been used.

In the modal emotion-false memory experiment, subjects first encode a series of semantically-related targets from such norms, which vary systematically in valence (see the left column of Table 1 for examples). Then, they respond to an old/new recognition test composed of three types of probes: old target items (O), new items that are similar in meaning to targets (NS; see the middle column of Table 1), and new items that differ in meaning from targets (ND; see the right column of Table 1 for examples). Notice that NS items vary systematically in valence, in accordance with the corresponding targets that subjects encoded. There is a baseline false memory effect if the false alarm rate is higher for NS items than for ND items ( $FA_{NS} > FA_{ND}$ ); that is, after reading the words *table*, *couch*, *desk*, and *sofa*, subjects are more likely to erroneously judge that they read *chair* or *seat* than to erroneously judge that they read *city* or *music*. The question of interest is whether the false memory effect is more robust for some valences than for others. The prevention hypothesis obviously predicts that it will be smallest for negatively-valenced NS items, whereas the distortion hypothesis predicts the opposite.

Bookbinder and Brainerd (2016) concluded that available data exhibit three broad effects, the first of which favors the distortion hypothesis. First, negatively-valenced NS items show higher levels of false memory than neutral or positively-valenced ones, for both words (e.g., El Sharkawy, Groth, Vetter, Beraldi, & Fast, 2008) and pictures (e.g., Bookbinder & Brainerd, 2017). Second, in the same experiments, negatively-valenced materials produce lower levels of true memory (hit rates for O items) than neutral or positively-valenced ones, so that negative valence yields across-the-board memory impairments. Third and less consistently, positively-valenced NS items produce higher levels of false memory than neutral ones.

Thus, there seems to be an overall valence effect such that regardless of its direction, valenced targets elevate false memory, a pattern that is consistent with fuzzy-trace theory's (FTT) proposal that valenced materials stimulate processing of items' semantic gist (Brainerd, Holliday, Reyna, Yang, & Toglia, 2010). However, Bookbinder and Brainerd (2016) argued that it is premature to conclude that any of these effects are due to valence per se because most valence manipulations neither controlled for correlated variability in arousal nor varied valence and arousal factorially. Instead, negatively-valenced materials were usually more arousing than neutral ones, and when both positively- and negatively-valenced materials were administered, the latter were usually more arousing than the former. This is a crucial consideration because it is possible that the deeper semantic processing that foments false memory might be due to differences in arousal rather than differences in valence (see also, Huntsinger, 2013).

In the present article, we report some studies that dealt with three questions that are precipitated by the routine confounding of valence with arousal in emotion-false memory experiments. The first is concerned with the exact mathematical relation between valence and arousal ratings in the word and picture norms that supply the target items for such experiments. The second is concerned with what a large-scale emotion word norming project

shows about the relative influence of valence and arousal on false memory. The third is concerned with how words' valence and arousal levels covary with words' other semantic properties, especially properties that are known to elevate false memory and to suppress it.

The first question figures in Study 1. Suppose that the exact mathematical function that maps arousal ratings with valence ratings were known for the word and picture norms that dominate emotion-false memory research. That function could be exploited to determine whether the aforementioned effects could possibly be due to differences in arousal as well as valence. More explicitly, the function could be analyzed to isolate properties that must be present in emotion-false memory data if arousal as well as valence is contributing to false memory (see below). The data of published experiments could then be reanalyzed to determine whether those properties were present.

Bradley and Lang (1999) remarked in connection with both the ANEW word norms and the IAPS picture norms that neutral items are less arousing than positively- or negatively-valenced items: "For items rated as neutral in valence . . . , arousal ratings do not attain the high levels associated with either pleasant or unpleasant materials" (p. 1). However, that observation neither specifies a particular valence-arousal relation nor posits one that holds for different types of materials, such as words versus pictures. With respect to emotional words, other researchers (e.g., Citron, Weekes, & Ferstyle, 2014; Kanske & Kotz, 2010; Vo, Conrad, Kuchinke, Urton, Hofmann, & Jacobs, 2009) have proposed that the valence-arousal relation is U-shaped (quadratic) and have tested that proposal with selected word pools. For instance, Citron et al. found that the regression equation  $A = .44 V^2 - .23 V + 2.66$  accounted for 64% of the variance with 300 words that comprise the Sussex Affective Word List, on which words were rated for valence and arousal on a -3 to +3 scale. However, it is unknown whether such a relation holds for the large word pools that have been central in emotion-false memory research and whether the same relation holds for the picture pools that have been central in such research. Those uncertainties were resolved in Study 1, where we report that the same quadratic function holds for both word and picture pools, and that it has three fine-grained properties that can be used to diagnose whether arousal as well as valence contributes to false memory when the two are confounded.

The second question was the focus of Study 2. Because the quadratic function's properties can be used to test for arousal effects when valence is manipulated in a false memory experiment, we analyzed a normed pool of emotional words—the Cornell-Cortland Emotional Word Lists (EWL; Brainerd, et al., 2010)—for the presence of those properties. For tasks such as those in Table 1, these norms provide levels of false recall and false recognition for a large set of lists that vary in valence.

Last, the third question was the focus of Study 3, where we examined the relation between valence and arousal ratings of words and their other semantic properties. In the mainstream memory literature, it has long been common practice to rate words for semantic properties such as concreteness and meaningfulness (e.g., Paivio, Yuille, & Madigan, 1968), and it is well established that such properties affect the accuracy of recognition and recall (Togila & Battig, 1978). More recently, in the false memory literature, the relation between these semantic properties and measured levels of false memory has been studied (Brainerd, Yang,

Howe, Reyna, & Mills, 2008; Cann, McRae, & Katz, 2011; Roediger, Watson, McDermott, & Gallo, 2001). A key result that pertains to emotion-false memory research is that false recognition and false recall load positively on a cluster of properties that Brainerd, Yang, et al. called a false memory factor, and they load negatively on a different cluster of properties that they called a true memory factor. Thus, in Study 3, we asked the obvious question about these semantic clusters: Do they map with valence or arousal ratings in the same way that they map with false memory? If they do, it is understandable that valence or arousal elevates false memory because they increase semantic properties that are known to elevate false memory or decrease semantic properties that are known to suppress it or both.

## Study 1

The aims of this study were to determine whether there is a single mathematical function that maps arousal ratings with valence ratings in emotional word and picture norms, and if there is, to identify properties of that function that can be used to interpret extant emotion-false memory data. To do that, we fit the valence and arousal ratings of the most widely used emotional word norms, the ANEW (Bradley & Lang, 1999), and the most widely used emotional picture norms, the IAPS (Lang et al., 2008). We also fit the valence and arousal ratings of another set of emotional words that resembles the ANEW, the Warriner, Kuperman, and Brysbaer (WKB; 2013) norms, and another set of emotional pictures, the EmoPics (Wessa, Kanske, Neumeister, Bode, Heissler, & Schönfelder, 2010), that resembles the IAPS. These latter norms have some desirable psychometric properties that make them especially useful for model fitting.

To preview our results, we found that a quadratic function of the form  $aX^2 - bX + c$  gave good accounts of arousal-valence relations in all of these norms. Moreover, the values of the functions' parameters were similar across the norms, indicating that the underlying quantitative tradeoff between perceptions of valence and perceptions of arousal does not depend greatly on either the specific target items or whether the items are verbal or pictorial. This function predicts that if valence-induced increases in false memory are also due to correlated increases in arousal, then (a) false memory should increase more in response to increases in negative than positive valence, (b) increases in false memory should be a positively accelerated function of increases in either positive or negative valence, and (c) false memory should increase more rapidly as a function of increases in negative valence than increases in positive valence.

## Method

In order to have a reasonable chance of isolating a single function that maps arousal ratings with valence ratings, word and picture norms must have some basic psychometric properties—specifically, that large numbers of subjects rate large numbers of items for valence and arousal. The most widely used norms, the ANEW and the IAPS, meet those criteria, and we analyzed the valence and arousal ratings from both. For generality, we also analyzed the valence and arousal ratings for an alternative set of emotional word norms, the WKB, and an alternative set of picture norms, the EmoPics. A shared methodological feature of all four

norms is that subjects used the same 1–9 scale to rate the valence of target items, and they used the same 1–9 scale to rate the arousal of target items.

In each of these norms, we determined how arousal ratings vary as function of valence ratings using a two-step procedure. First, we fit a series of familiar monotonic and nonmonotonic functions to the complete set of mean valence and arousal ratings for individual items, in order to locate the best-fitting function. In all cases, that was the general quadratic equation  $aX^2 - bX + c$ . Second, we refit this function to smoothed data, in which mean valence and arousal ratings had been averaged over blocks of items that spanned consecutive half points on the 1–9 valence rating scale. In other words, we refit the function using paired valence and arousal means for items whose mean valence was 1 to 1.49, 1.5 to 1.99, 2 to 2.49, and so on. That was to reduce the high variability in valence and arousal ratings, which arises from the fact that there are sizeable individual differences in subjects' perceptions of these attributes: In the ANEW, for instance, the *SDs* on the 9-point scales, are 2.37 for arousal and 1.65 valence, which means that fits of the ratings for individual words will be noisy. That problem is eliminated and a clearer picture emerges with the refitting procedure.

## Results

We report the fit results for the four emotion norms first, which showed that the quadratic  $aX^2 - bX + c$  always provided the best fit to valence and arousal ratings. Next, we derive some predictions about false memory that must hold if emotion-false memory effects are due to arousal as well as valence when the two are confounded.

**Valence-arousal functions**—With the ANEW, we fit the mean arousal ratings of individual words to their mean valence ratings using four familiar monotonic functions (linear, exponential, log, and power) and the two simplest nonmonotonic functions (quadratic and cubic). The quadratic (U-shaped) function  $.17X^2 - 1.71X + 8.77$  yielded by far the best fit, accounting for 29% of the variance. As a group, the monotonic functions produced poor fits, accounting for an average of 1% of the variance. A cubic function necessarily accounts for more variance than a quadratic function because it estimates one more parameter (4 rather than 3), but in this instance, the increase was not reliable, so that the mapping of arousal ratings onto valence ratings was quadratic to a statistically acceptable approximation. Finally, we refit the quadratic function following the procedure described in the Method section: Mean valence and arousal scores were summed and averaged for items that fell within each of the consecutive half-point ranges of the 9-point valence rating scale, and the best-fitting quadratic function for the paired valence and arousal means was computed. That function, which is displayed in Figure 1A and accounted for an impressive 87% of the variance, was  $A = .16V^2 - 1.65V + 8.62$ , where  $A$  denotes words' mean arousal ratings and  $V$  denotes their mean valence ratings. Notice that the numerical estimates of the three parameters were virtually the same as the corresponding estimates for the raw ANEW data, and indeed, when the quadratic function was refit to the raw data with its parameters fixed at these values, the amount of variance accounted for did not decrease significantly.



Turning to the WKB emotional word norms, we simply repeated this procedure, and the overall pattern was the same as for the ANEW. First, we fit the mean arousal ratings of individual words to their mean valence ratings using the same monotonic and nonmonotonic functions. The quadratic function  $A = .15 V^2 - 1.55 V + 9.08$  yielded by far the best fit, accounting for 14% of the variance. As before, the monotonic functions produced poor fits, and the cubic function, which estimates an additional parameter, did not account for more variance than the quadratic function. Note the instructive similarity between the numerical estimates of the quadratic function's parameters for the raw WKB data and the corresponding estimates for the raw ANEW data. Second, we refit the quadratic function to mean valence and arousal scores for items that fell within consecutive half-point ranges of the 9-point valence rating scale. The best-fitting function, which is displayed in Figure 1B and accounted for an impressive 95% of the variance, was  $A = .14 V^2 - 1.48 V + 7.87$ . As with the ANEW norms, the numerical estimates of the three parameters were similar to the corresponding estimates for the raw WKB data. However, when the quadratic function was refit to the raw data with its parameters fixed at these values, it produced a small but reliable reduction in the amount of variance accounted for.

Next, we consider the two emotional picture norms, the IAPS and the EmoPics. Taking the IAPS first, we fit the mean arousal ratings of individual pictures to their mean valence ratings using the same monotonic and nonmonotonic functions. The best-fitting function was again quadratic, the expression  $.18X^2 - 1.92X + 9.32$  accounting for 29% of the variance. Obviously, the numerical estimates of this function's parameters for the raw IAPS data are similar to the corresponding estimates for the two word norms. Second, we refit the quadratic function to mean valence and arousal scores for pictures that fell within consecutive half-point ranges of the 9-point valence rating scale. The best-fitting function, which is displayed in Figure 2A and accounted for 78% of the variance, was  $.18X^2 - 1.92X + 9.48$ . Note that the estimates of the first two parameters were the same as those for the raw IAPS data, and the estimate of the third was close to the corresponding estimate for the raw IAPS data. Hence, it is not surprising that when the quadratic function was refit to the raw data with its parameters fixed at these values, the reduction in the amount of variance accounted for was not reliable.

Moving on to the EmoPics data, we fit the mean arousal ratings of individual pictures to their mean valence ratings using the same monotonic and nonmonotonic functions. Once again, the best-fitting function was quadratic, the expression  $.40X^2 - 4.15X + 13.90$  accounting for an impressive 74% of the variance. The fact that the variance accounted for is much greater than in the corresponding analyses for the ANEW, WKB, and IAPS is due to reduced individual differences: The SDs for both arousal and valence ratings are much lower for EmoPics than for the other three norms. When we refit the quadratic function for pictures that fell within consecutive half-point ranges of the 9-point valence rating scale, the best-fitting function was  $.34X^2 - 3.59X + 12.91$ , and it accounted for 92% of the variance. That function is shown in Figure 2B.

The reason that the values of the parameters of the quadratic function are so much larger for the EmoPics norms than for the other norms is apparent from a visual comparison of Figure 2B to Figures 1A, 1B, and 2A. That comparison reveals that the average level of arousal for

negative items (valence ratings below 5) is considerably higher for the EmoPics norms than for the other norms, whereas the average level of arousal for positive pictures (valence ratings above 5) is much lower for EmoPics norms. For instance, the mean arousal rating for ANEW items with 1–3 valence ratings is 5.49 and the mean arousal rating for ANEW items with 7–9 valence ratings is 5.60, whereas the corresponding mean ratings for EmoPics items are 6.38 and 4.88, respectively. To fit the EmoPics data, then, the quadratic function must rise more steeply on the negative valence side but rise less steeply on the positive valence side, producing the observed increase in parameter values.

**Predictions about false memory**—We now present some simple techniques that researchers can use to reanalyze the data of published emotion-false memory experiments in order to determine whether differences in arousal are contributing to valence effects. As mentioned, arousal has been confounded with valence in the preponderance of those experiments. Despite that, other data indicate that emotion-false memory effects cannot be entirely due to arousal rather than valence. Specifically, Bookbinder and Brainerd (2016) identified a few experiments in which valence was manipulated with arousal controlled, and they produced higher levels of false memory when target items were valenced rather than neutral. Thus, the question of interest is whether arousal contributes *at all* to emotion-false memory effects, and if it does, whether it contributes more than valence.

To address that question, we identified three fine-grained properties of the quadratic functions in Figures 1A–2B that yield simple predictions that could be used to reanalyze the data of experiments in which valence was manipulated without controlling arousal. The first and most obvious one is apparent from the fact that in these functions, *negative valence is always more arousing than positive valence*. This is because as valence ratings increase, the negative valence arms of the functions decline more steeply than the positive arms increase, so that items of a given level of negativity (say, ratings of 2) are more arousing than items of the corresponding level of positivity (ratings of 8). Indeed, notice in all four curves that as valence ratings decline in the negative (left) arms of the curves, arousal continues to decline *beyond the objective neutral point of the valence scale* (ratings of 5), and arousal does not begin to increase until valence ratings are well into the positive range. Taking a numerical example, consider ANEW words in the negative arm of Figure 1A with a mean valence of 2 and ANEW words in the positive arm with a mean valence of 8. According to the best-fitting function, the mean arousal rating of the former would be 5.96 and of the latter would be 5.66. Second, because these functions are curves rather than straight lines, differences between the arousal levels of negative versus positive items increase as negative items become more negative and positive items become more positive. For instance, the mean arousal ratings of ANEW words with mean valence ratings of 3 and 7 would be 5.11 and 4.91, respectively, whereas we just saw that the corresponding ratings of words with mean valence ratings of 2 and 8 would be 5.96 and 5.66. The third property falls out of the second—namely, that when emotional items involve only a single valence (most extant studies involve negative valence only), the increase in false memory as positivity or negativity increases will be positively accelerated rather than constant (linear). For negative valence, for instance, the increase in false memory should be greater as valence ratings move from 2 toward 1 than as they move from 3 toward 2.



**Summary**—Study 1 produced two instructive findings about valence, arousal, and false memory for the types of materials that have been common in emotion-false memory research—namely, emotional word and emotional picture norms. First and most important, with the usual methodology for rating items' valence and arousal, there was a universal tradeoff between valence and arousal ratings that followed a quadratic rule. The rule is a type of quadratic function in which (a) increases in negative and positive valence produce positively accelerated increases in arousal, and (b) arousal accelerates more rapidly as a function of negative valence than positive valence. Second, although arousal has been routinely confounded with valence in prior experiments, there are specific features of this quadratic function that can be used to diagnose whether arousal as well as valence contributed to reported emotion-false memory effects. Those properties can be evaluated with the data of extant experiments.

## Study 2

We just saw how the influence of arousal on false memory can be studied with available data by evaluating the quadratic function's fine-grained predictions about such data. In Study 2, we provide a worked illustration by evaluating these predictions with the norming data for the EWL. As mentioned, experimentation on emotional words is by far the most common procedure in the emotion-false memory literature. The EWL is a standardized procedure of that sort. It is an emotional version of the most widely used paradigm in emotion-false memory research, the Deese/Roediger/McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) illusion. The EWL consists of 32 word lists, 16 negatively-valenced lists and 16 positively-valenced lists. Each list contains of 15 target words that are all forward associates of a missing word that serves as the false-memory item. For example, the words *hate, kiss, like, happy, heart, care, admire, adore, close, friendship, spouse, happiness, hug, kindness, and life* are all forward associates of *love*, and the words *web, insect, bug, fright, fly, arachnid, crawl, tarantula, poison, bite, creepy, Black Widow, monkey, feelers, and tiny* are all forward associates of *spider*. As is apparent from these examples, the valence and arousal levels of most target words are the same as that of the false memory item.

Valence and arousal ratings for EWL target words and false-memory items and backward associative strengths from target words to false-memory items were obtained from existing semantic word norms. The valence/arousal ratings were obtained primarily from the ANEW because most of the words had been rated in those norms. For the few words that had not been rated in the ANEW, valence and arousal ratings were available from other norms that had used the same 1–9 rating scale (e.g., the WKB). Associative strengths were obtained from the Nelson, McEvoy, and Schreiber (1999) norms. Valence ratings for false-memory items range from 1.9 to 4.8 for the 16 negative lists and from 6 to 8.7 for the 16 positive lists. Brainerd et al. (2010) normed the EWL lists for their mean levels of true and false recall and their mean levels of true and false recognition. To produce the norms, a sample of 229 adult volunteers studied these lists and responded to free recall tests and recognition tests. The 32 lists, with the mean rating scores of individual list words and false memory items, are available in Table 1 of Brainerd et al. (2010).

Because the valence and arousal levels of most target words are the same as those of the false memory item, the memory effects of these lists could be of two general sorts: item-level and list-level. The former refers to valence/arousal effects of individual list words, and the latter refers to an overall impression or gist that could be extracted from encoding several emotionally similar words. Some investigators (e.g., Brainerd, Stein, et al., 2008) have noted that list-level effects can be separated from item-level effects in recognition experiments, by administering distractors that share the valence/arousal properties of emotional lists but not the specific content of the target words. For instance, suppose that subjects study the two emotional lists in Table 1 and then respond to a recognition test that, among other things, contains the ND items in the far right column. Notice that some of these ND items share the valence/arousal properties of the emotional lists (*soft*, *warm*, *spider*, and *thief*) and others do not (*city* and *music*). If these lists produce list-level as well as item-level effects, the false alarm rate will be higher for the first group of distractors than for the second. This difference in false alarm rates has not been obtained for the EWL lists (or for other emotional word lists; Bookbinder & Brainerd, 2016), and hence, item-level effects of valence/arousal appear to predominate over list-level effects.

We analyzed the data of this norming sample in order to test each of the above predictions. Concerning the first, because negative valence is more arousing than positive valence, false memories should be more common with the negative lists, if arousal also contributes to false memory. It is apparent in Figure 3A that this prediction was confirmed. There, mean levels of false recognition and false recall are plotted for the positive and negative lists. It can be seen that false recognition was roughly 50% higher for negative than for positive lists, and false recall was roughly 25% higher for negative than for positive lists. This is only a single result, however, and negative lists might produce higher levels of false memory for reasons that have nothing to do with arousal (see Study 3, below). FTT, for example, predicts this result on the ground that negative content generates stronger semantic connections among events than positive or neutral content (Bookbinder & Brainerd, 2016). Consequently, all three predictions must be confirmed to make a solid case that correlated arousal differences are contributing to emotion-false memory effects.

Turning to the other two predictions, a single analysis will suffice for both. The second prediction specifies that on average, a unit increase in negativity (say, from 3 to 2) will elevate false memory more than a comparable increase in positivity (say, from 7 to 8). The third prediction specifies that when increases in false memory are plotted against increases in either negative or positive valence, the functions will be positively accelerated rather than linear—more simply, the increase in false memory will be greater at more extreme values of either valence than at less extreme values. Let  $f_{NF}$  be the function that maps levels of false memory with valence ratings on the negative side (1–5), and let  $f_{PF}$  be the corresponding function for valence ratings on the positive side (5–9). It follows from the third prediction that both functions will be positively accelerated rather than linear, and it follows from the second that acceleration will be greater for  $f_{NF}$  than  $f_{PF}$ .

Hence, both predictions can be tested by fitting false memory levels to valence ratings, for the 16 negative lists and the 16 positive lists, and comparing the relative fit of the general linear functions  $F = a - bN$  and  $F = a + bP$  to that of common positively-accelerated

functions, such as power functions. When we did that for the EWL norming data, the results were clear and consistent: Positively-accelerated functions never accounted for significantly more variance than linear functions; there was no evidence that false memory increases more at more extreme levels of positive or negative valence than at less extreme levels, nor that the acceleration is greater for negative lists than for positive ones.

Thus, the results for the second and third predictions do not support the hypothesis that in the EWL norms, correlated arousal differences contributed to the effects of valence on false recognition and false recall. Obviously, those results suggest a different conclusion than those in Figure 3A. Therefore, we conducted a follow-up analysis to resolve that uncertainty. Using emotional word norms such as the ANEW and the WKB, arousal ratings can be obtained for the EWL materials. With those data, the question of whether false memory increases when arousal increases can be directly addressed by comparing lists that differ in arousal *but not valence*. Using the arousal data for the EWL, the 32 lists can be split into 4 groups of 8 lists each: (a) negative/higher-arousal ( $M_V = 3.09$  and  $M_A = 5.90$ ), (b) negative/lower-arousal ( $M_V = 3.69$  and  $M_A = 4.14$ ), (c) positive/higher-arousal ( $M_V = 7.88$  and  $M_A = 6.70$ ), (d) positive/lower-arousal ( $M_V = 6.67$  and  $M_A = 4.00$ ). We already know (Figure 3A) that false recognition and false recall are higher for groups a and b than for groups c and d. The question is whether they are higher for groups a and c than for groups b and d.

The answer for recall appears in Figure 3B, and the answer for recognition appears in Figure 3C. For recall, it is obvious at a glance that increases in arousal did not elevate false memory. For positive lists, the false recall probabilities for higher- and lower-arousal lists were virtually identical (.25 and .24), whereas for negative lists, the false recall probability was slightly higher for higher- than for lower-arousal lists (.33 vs. .29). However, the latter difference was not reliable because when we computed a  $2(\text{valence: positive vs. negative}) \times 2(\text{arousal: higher vs. lower})$  analysis of variance (ANOVA) of the false recall data, there was neither a main effect for arousal nor a Valence X Arousal interaction ( $F_s < 1$ ). For recognition, the picture was somewhat different, but Figure 3C shows that false memory did not increase when arousal increased. For negative lists, the false recognition probability was slightly but not reliably higher for higher- than for lower-arousal lists (.40 and .36), whereas for positive lists, false recognition was notably *lower* for higher- than for lower-arousal lists (.14 vs. .30). The latter difference was statistically reliable because when we computed a  $2(\text{valence: positive vs. negative}) \times 2(\text{arousal: high vs. low})$  ANOVA of the false recognition data, the Valence X Arousal interaction was significant,  $F(1, 227) = 11.74, p < .001$ .

To summarize, analyses of norming data for the EWL did not support the hypothesis that arousal as well as valence contributes to false memory. Although the prediction that negative words would produce more false memory than positive words was confirmed, the other predictions were not. Further, direct comparisons of false memory rates for words that differ in arousal but not valence failed to detect differences in false memory. These data suggest that for emotional words, at least, arousal does not greatly influence false memory when valence and arousal are confounded.

### Study 3

The pattern that we just considered—that words’ valence levels have more powerful effects on false memory than their arousal levels—would be more compelling if there were a straightforward theoretical explanation of it. In Study 3, we evaluated such an explanation, which falls out of FTT. According to FTT (see Bookbinder & Brainerd, 2016), valence is a more semantic property than arousal is; indeed, rating words’ pleasantness is a classic semantic-orienting task in memory experiments (e.g., Toggia, Neuschatz, & Goodwin, 1999). On this hypothesis, words’ valence ratings should be more intimately connected to semantic properties that elevate and suppress false memory than words’ arousal ratings are. This explanation can easily be tested by relying on an existing studies in which specific semantic properties that elevate and suppress false memory have been investigated (e.g., Brainerd, Yang, et al., 2008; Cann et al., 2011). A particular set of findings that we exploited in the present study is that the properties of words that subjects falsely recall and falsely recognize in the modal false memory task (Table 1) have been identified with the aid of the Toggia and Battig (1978) semantic word norms.

The Toggia-Battig norms are perhaps the most widely used semantic word norms in the mainstream memory literature, as they include ratings of several properties that have robust effects on recall and recognition. They also incorporate properties from other influential norms (e.g., Paivio et al., 1996). Over 2,500 subjects rated a pool of 2,854 words for their levels of categorizability, concreteness, imagability, meaningfulness, familiarity, and number-of-attributes on 1 (low) to 7 (high) scales. Note that the first three dimensions involve realistic properties of words, whereas the last three involve properties that are more abstract and conceptual. When these norms are factor analyzed, two factors emerge (Brainerd, Yang, et al., 2008), one on which concreteness, categorizability, and imagability ratings all have positive loadings of  $> .8$ , and one on which meaningfulness, familiarity, and number-of-attributes all have positive loadings of  $> .8$ .

Brainerd, Yang, et al. (2008) investigated the relation between the Toggia-Battig norms and norms of false recall and false recognition for the types of tasks that have figured in most emotion-false memory experiments (Roediger, Watson, McDermott, & Gallo, 2001). A joint factor analysis of the two norms revealed two factors: (a) a *true memory* factor on which concreteness, categorizability, imagability, and true memory all loaded positively but false memory loaded *negatively* and (b) a *false memory* factor on which meaningfulness, familiarity, number of attributes, and false memory all loaded positively. Thus, increases in the concreteness/categorizability/imagability cluster of properties was associated with increases in true memory and decreases in false memory, whereas increases in the meaningfulness/familiarity/number-of-attributes cluster was simply associated with increases in false memory. The negative effect of the first cluster on false memory seems to be due to the fact that O items with these properties produce strong verbatim traces that can be used to reject NS items (“No, *infant* was not on the list because I clearly remember that it was *baby*.”) The positive effect of the second cluster seems to be due to the fact that O items with these properties stimulate formation of the meaning connections that support false memories (Brainerd, Yang, et al., 2008).

This yields an obvious working explanation of the finding in Study 2 that words' valence levels affect false memory more than their arousal levels: Valence is (a) more strongly related to the meaningfulness/familiarity/number-of-attributes cluster than arousal is or (b) is more strongly related to the categorizability/concreteness/imagability cluster than arousal is or (c) both. (Relation a would be positive and relation b would be negative, of course.) To evaluate this explanation, we expanded the Toggia-Battig norms to include valence and arousal ratings of its words, and then, we analyzed the relations between those ratings and the two clusters of semantic properties.

We also used these data to address a second question: Why does negative valence elevate false memory more than positive valence? Bookbinder and Brainerd (2016) noted this pattern in their review, and it was also present in Study 2. A candidate explanation emerged in Study 1. There, remember that the best-fitting quadratic functions for both word and picture norms showed that increases in negative valence are more arousing than increase in positive valence. If arousal contributes substantially to false memory, this would nicely explain why negative valence is more distortive than positive valence. However, the results of Study 2 did not support that explanation, and hence, another account is needed. The one that we evaluated in Study 3 assumes that positive associations with the meaningfulness/familiarity/number-of-attributes cluster and/or negative associations with the categorizability/concreteness/imagability cluster are stronger for negative than for positive valence.

## Method

As mentioned, we combined semantic and emotional word norms by adding mean valence and mean arousal ratings to the words in the Toggia-Battig norms. To do that, we searched for each word's valence and arousal rating in both the ANEW and the WKB norms and then inserted those ratings in the Toggia-Battig norms. Between the ANEW and WKB norms, we were able to locate mean valence and mean arousal ratings for 2,184 of the 2,854 Toggia-Battig words. Those are the revised semantic word norms on which the present study is based.

The valence-arousal ratings were on a 1–9 scale, whereas the scores for the original Toggia-Battig properties were on a 1–7 scale. Using the Kucera-Francis norms (Kucera & Francis, 1967), we added another property to the 2,184 words in the revised norms: word frequency in printed text, which is an objective counterpart of one of the Toggia-Battig properties (familiarity). Then, we factor analyzed the revised norms in order to generate factor scores for the Brainerd, Yang, et al. (2008) factors. Finally, we analyzed the relations between these factor scores and valence and arousal ratings, in order to test our working explanation of emotion-false memory effects.

## Results

We report the results for relations between semantic and emotional word norms first. Then, we report follow-up results that address the question of why false memory increases more as a function of negative valence than positive valence.

**Semantic and emotional word norms**—We began by conducting a principal components analysis with orthogonal rotation, which is the most common form of factor analysis in the psychological literature (Fabrigar, Wegener, MacCallum, & Strahan, 1999). It was focused on the original Toglia-Battig properties plus the added frequency scores. Two factors were extracted, using the standard eigenvalue cutoff of 1, which accounted for a total of 77% of the variance. The detailed results are shown in Table 2, where the usual convention of only treating variables with factor loadings  $\geq .40$  as reliable has been followed. It can be seen that Factor 1, which accounted for 48% of the variance, is the true memory factor identified by Brainerd, Yang, et al. (2008), whereas Factor 2, which accounted for 29% of the variance, is the false memory factor that was identified by those authors.

In order to test our working explanation of why valence influences false memory, factor scores were generated for both factors. Then, the valence and arousal ratings of the Toglia-Battig words were separately fit to their scores on each factor. We repeated the Study 1 procedure of fitting the most common monotonic (linear, exponential, log, power) and nonmonotonic (quadratic, cubic) functions to the data. Taking the valence results first, we fit words' valence ratings to the factor scores for each factor. We found that valence ratings were poor predictors of true memory factor scores but good predictors of false memory factor scores. More specifically, none of the fitted functions gave a good account of the data when valence ratings were fit to true memory factor scores, the mean variance accounted for being statistically reliable but small (3%). In contrast, when valence ratings were fit to false memory factor scores, the quadratic function  $.14X^2 - 1.18X + 2.02$  gave a good account of the data, accounting for 22% of the variance. The other monotonic and nonmonotonic functions did not perform nearly as well.

Next, we refit the quadratic function for the false memory factor, following the smoothing procedure that was used in Study 1 to reduce the influence of variability in words' valence ratings: Valence ratings and factor scores were summed and averaged for items that fell within each of the consecutive half-point ranges of the 9-point valence rating scale, and the best-fitting quadratic function for the paired valence and arousal means was computed. That function, which is displayed in Figure 4 and accounted for 98% of the variance, was  $.17X^2 - 1.49X + 2.87$ . Notice that the numerical estimates of the three parameters are virtually the same as the corresponding estimates for the raw data. When the quadratic function was refit to the raw data with its parameters fixed at the second set of values, the amount of variance accounted for did not decrease significantly. However, the key point that emerges from Figure 4 is theoretical. Consistent with our hypothesis that positive and negative valence increase false memory because they drive up semantic properties that are known to elevate false memory, scores on the Toglia-Battig meaningfulness/familiarity/number-of-attributes cluster increase as words' valence ratings either become more negative or more positive.

Turning to arousal ratings, the results were quite different and were consistent with the view that arousal's effects on false memory are small in comparison to valence's effects because arousal is not as strongly related to semantic properties that elevate false memory. As with valence ratings, we fit words' arousal ratings to their scores on the true and false memory factors, using common monotonic and nonmonotonic functions. Unlike valence ratings,



arousal ratings were not good predictors of either factor. The monotonic functions gave the best fits, but the average variance accounted for was small for the false memory factor (4%) and even smaller for the true memory factor (< 2%), although both values were reliable owing to the very large number of data points.

The strengths of the relations between the false memory factor and valence ratings versus arousal ratings can be directly compared by (a) converting the valence ratings to a monotonic numerical scale and then (b) using both valence and arousal ratings to simultaneously predict scores on the false memory factor in a multiple regression. Here, remember that although the 1–9 scale for arousal is monotonic, the 1–9 valence scale is nonmonotonic: Valence strength initially decreases as ratings increase from 1 to 5, but then, it increases as ratings increase from 5 to 9. These ratings can be converted to a monotonic scale in which valence strength always increases as ratings increase by merely subtracting the scale mid-point (5) from each rating and taking the absolute value of the signed difference in order to eliminate negative values. We generated these monotonic valence ratings for the Toggia-Battig words, and then computed a multiple regression in which arousal and valence ratings were the predictor variables and scores on the false memory factor were the dependent variable. This regression accounted for a highly reliable 16% of the variance, and the change statistic showed that arousal and valence ratings were both reliable predictors of scores on the false memory factor. Although arousal and valence ratings were both reliable predictors, valence was the better predictor by far: The partial correlation between valence ratings and false memory factor scores was .33, whereas the partial correlation for arousal ratings was only .08.

Another important finding of the multiple regression that is consistent with our working explanation concerns the observed values of the parameters of the regression equation. The best-fitting standardized regression equation was  $F_{fm} = .35V + .08A$ , where  $F_{fm}$ ,  $V$ , and  $A$  are false memory factor scores, valence ratings, and arousal ratings, respectively. The coefficients .35 and .08 are the rates at which false memory factor scores increase as valence ratings and arousal ratings, respectively, increase. Thus, the average increase in the false memory factor scores that results from a unit increase in valence scores is more than four times the corresponding increase that is produced by a unit increase in arousal. This difference is partly due to the fact that the numerical range of the arousal rating scale (1–9) is twice the numerical range of the monotonic valence scale (0–4). However, if we equate for that, either by dividing the  $V$  coefficient by 2 or multiplying the  $A$  coefficient by 2, the average increase in false memory factor scores that results from a unit increase in valence scores is still more than twice as large as the corresponding increase that results from a unit increase in arousal scores. In other words, the positive association between the meaningfulness/familiarity/number-of-attributes cluster, on the one hand, and valence and arousal, on the other, is much stronger for valence than for arousal.

**Positive versus negative valence**—A final puzzle remains about the relation between valence and false memory, which is the finding in Study 2 and in the larger emotion-false memory literature that false memory increases more as a function of increases in negative valence than increases in positive valence. As mentioned, if arousal contributes substantially to false memory, this positive-negative discrepancy could be explained by the finding in

Study 1 that increases in negative valence are inherently more arousing than increases in positive valence. However, we have failed to find evidence of substantial arousal contributions—hence, the puzzle.

We noted that an alternative explanation is that negative valence is more strongly related to semantic properties that increase or decrease false memory than positive is. That explanation can be tested by re-running some of the above regressions separately for negatively-valenced words versus positively-valenced words. Here, remember that we found (Figure 4) that the full range of valence ratings (both positive and negative valence) is a strong predictor of false memory factor scores but a poor predictor of true memory factor scores. The latter result could change if regressions were run separately for positive and negative valence; that is, one might be a better predictor of true memory factor scores than the other.

To conduct these analyses, we first split the Toggia-Batting word pool into two sub-pools: (a) negatively-valenced words (all words with mean valence ratings of 1–4 on the ANEW norms or the WKB norms) and (b) positively-valenced words (all words with mean valence ratings of 6–9 on the ANEW norms or the WKB norms). Then, we conducted a multiple regression for each sub-pool in which the criterion variable was words' valence ratings and the predictor variables were words scores on the true and false memory factors. The objective was simply to quantify how well true and false memory factors were jointly predicted negative valence versus positive valence.

Taking the results for negative valence first, the regression showed that increases in negative valence had two effects, both of which would elevate false memory: As words become more negative, scores on the false memory factor *increase*, and simultaneously, scores on the true memory factor *decrease*, both of which produce increases in false memory (Brainerd, Yang, et al., 2008). The best-fitting standardized regression equation accounted for a highly reliable 15% of the variance, and the change statistic showed that the true and false memory factors were both reliable predictors of valence scores. The best-fitting standardized regression equation was  $V = .30F_{fm} - .23F_{tm}$ , where  $V$ ,  $F_{fm}$  and  $F_{tm}$  are valence ratings, false memory factor scores, and true memory factor scores, respectively. (For this regression, we used the monotonically transformed valence scale mentioned above, so that negative valence increased from 1–4, rather than decreased.) Thus, a unit increase in negative valence was associated with a 30% increase in the cluster of false memory properties and a 23% decrease in the cluster of true memory properties. Experimentally, the implication is that increasing negative valence will drive up semantic properties that foment false memory *and simultaneously* drive down semantic properties that suppress false memory.

Turning to the results for positive valence, the regression showed that increases in positive valence had only one of the two effects that negative valence had: As words become more positive, scores on the false memory factor increase by roughly the same amount as when words become more negative, but scores on the true memory factor are not affected. The best-fitting standardized regression equation accounted for the same amount of variance (15%) as the corresponding equation for negative valence, but the change statistic showed that only the false memory factor was a reliable predictor of positive valence scores. The best-fitting standardized regression equation was  $V = .38F_{fm}$ . Thus, a unit increase in

positive valence was associated with a 38% increase in the cluster of false memory properties.

In sum, these results provide a straightforward explanation of why false memory is more strongly influenced by negative than positive valence. On the one hand, increasing negative valence has two effects, both of which elevate false memory. On the other hand, positive valence has only one of these effects.

## General Discussion

We have examined a current uncertainty about emotion-false memory effects that poses obstacles to their theoretical interpretation. Beginning with the early work of Budson et al. (2006) and Howe (2007), conventional false memory tasks have been modified so as to vary their emotional content. Most often, as in these early studies, false memory levels have been compared for negative versus neutral valence, although positively-valenced content has occasionally been studied (Dehon, Larøi, & Van der Linden, 2010; Gallo, Foster, & Johnson, 2009). The consensus finding is that negative content elevates false memory relative to neutral content, and in the few studies that have included positive content, it has usually elevated false memory relative to neutral content, too. The first result is especially instructive theoretically because it runs counter to the law's hypothesis that negative content inoculates memory against distortion.

The uncertainty about these results is whether they can reasonably be interpreted as valence effects because in most studies, the valence of target and test items has been confounded with their levels of arousal; valenced content has been more arousing than neutral content. To illustrate the scope of this problem, when Bookbinder and Brainerd (2016) reviewed the emotion-false memory literature, they found that it was present in over 70% of published experiments. Although negative valence elevated false memory in two studies that controlled arousal levels (Brainerd, Stein, et al., 2008; Dehon et al., 2010), it is unknown whether arousal accounts for some or most of the false memory effects in the large collection of studies in which it was not controlled. In the present article, we attempted to make progress on this uncertainty by pursuing two lines of investigation.

First, we generated some tools for addressing it in extant data by investigating the mathematical relation between ratings of valence and arousal for the materials that have predominated in emotion-false memory research—namely, items from emotional word and picture norms. The objective was to identify a function that delivers good accounts of the valence-arousal relation in both types of norms and then to analyze it to isolate properties that can be used to test whether arousal as well as valence contributes to emotion-false memory effects when the two are confounded. In Study 1, the quadratic function  $A = aV^2 - bV + c$  performed better than other functions and gave good accounts of the variance in the most widely used emotional word norms (ANEW), the most widely used emotional picture norms (IAPS), a second set of emotional word norms, and a second set of emotional picture norms (Figures 1 and 2). We identified three properties of this generic function that are diagnostic of arousal contributions to emotion-false memory effects. In Study 2, we illustrated how those properties can be used to test for arousal effects by analyzing norming

data for a pool of emotional word lists (EWL) that have been used in prior research. Tests of the properties showed that in those data, false recognition and false recall were affected by differences in valence but not by differences in arousal.

The second line of investigation, which was implemented in Study 3, focused on a simple semantic explanation of emotion-false memory effects for emotional words, which are the materials that have been used in the preponderance of published experiments. The explanation takes advantage of the fact that the relation between false memory and some common semantic properties of words (e.g., concreteness, familiarity) has been investigated in the mainstream memory literature. There, it has been found that false memory *increases* as a function of a cluster of properties that consists of meaningfulness, familiarity, and number-of-attributes, and that it *decreases* as a function of another cluster of properties that consists of concreteness, categorizability, and imagability. This pattern is consistent with opponent-process theories of false memory, such as FTT (Brainerd & Reyna, 2005).

According to the semantic explanation of emotion-false memory effects, valence is more strongly correlated with these clusters of semantic properties than arousal is—so that at least part of the reason why false memory increases more as a function of valence than arousal is that changes in valence produce larger correlated changes in clusters of semantic properties that are known to influence false memory. When we analyzed valence and arousal ratings of the items in the Toglia-Battig semantic word norms, the results were consistent with that explanation. On the one hand, variation in valence ratings accounted for a sizeable portion of the variance in the meaningfulness/familiarity/number-of-attributes cluster. On the other hand, variation in arousal ratings did not.

Other instructive findings emerged when the relation between valence ratings and semantic word norms was analyzed separately for positive and negative valence, findings that explain why increases in negative valence elevate false memory more than increases in positive valence. For negative valence, the analyses showed that increases in negative valence were positively correlated with semantic properties that elevate false memory and negatively correlated with properties that suppress it. Theoretically, it seems that negative valence both stimulates formation of the meaning connections that support false memory and reduces the accessibility of verbatim traces that suppress false memory (see also, Bookbinder & Brainerd, 2017). For positive valence, the analyses showed that it only has the first effect; increases in positive valence were positively correlated with semantic properties that elevate false memory but were unrelated to properties that suppress it.

Beyond these findings, a remaining question is whether valence's influences on false memory are pure semantic effects—whether valence contributes nothing beyond the influence of the semantic properties with which it is correlated. This is a fundamental theoretical question, which cannot be answered with the types of data that we reported. Answering it would require large-scale multiple-regression designs in which true recognition/recall and false recognition/recall are the dependent variables and standard semantic properties, such as those studied by Cann et al. (2011) and Brainerd, Yang, et al. (2008), are the independent variables. Because DRM lists have been the target materials in most emotion-false memory research, they are the logical place to begin: Normed levels of

true and false memory for DRM lists would supply the dependent variables, and the ratings of list words and false-memory items on semantic properties and on valence would supply the independent variables. Extant DRM norms (Roediger et al., 2001; Stadler et al., 1999) provide true recognition/recall and false recognition/recall scores for a total of 51 lists, so that studies of this sort could be conducted if the list words and false-memory items in those norms were rated for valence and for standard semantic properties. The main question would be whether valence accounts for any of the variance in false memory scores after the variance that is attributable to semantic properties has been removed. There is a much larger pool of 200 DRM lists (Atkins & Reuter-Lorenz, 2011; Brainerd & Reyna, in press) that could be tapped to conduct even more comprehensive studies of this question.

Overall, the results of our studies are promising steps toward building a consensus interpretation of how emotional content affects false memory because (a) the results converged on valence but not arousal as a source of those effects, (b) the results generated a simple explanation of why valence distorts memory more than arousal, and (c) the results generated a simple explanation of why negative valence distorts memory more than positive valence. Further, the techniques that we used can be implemented by others to evaluate previously published data for the presence of some of these same patterns. However, our results are far from being the last word on the knotty problem of separating valence and arousal effects in false memory experiments. Perhaps that most obvious aspects of the research that should be broadened in later work are the scope of the materials that generated our results and the range of valence and arousal intensity in those materials. We briefly comment on these two points in closing.

### Scope of Materials

As mentioned, the study and test materials in previous experiments have either been emotional words or emotional pictures. Study 1 encompassed both types of materials. Study 2 and Study 3 were confined to emotional words. That is because the only sets of items that have been normed for valence, arousal, and false memory are words (Study 2), and the only sets of items that have been normed for valence, arousal, and semantic properties that elevate and suppress false memory are also words. From one perspective, this is a strength of the last two studies inasmuch as the materials in 60% of emotion-false memory experiments have been words (see Bookbinder & Brainerd, 2016, Table 1).

However, emotional pictures were used in some of the other experiments (e.g., Bookbinder & Brainerd, 2017; Mirandola, Toffalini, Grassano, Cornoldi, & Melinder, 2014), and as things currently stand, it is unknown whether the key results that converge on valence as the source of emotion-false memory effects will hold for pictures because those results are confined to our last two studies. On the one hand, there is some basis for supposing that the results of the last two studies will hold for pictures as well as words: The emotion-false memory effects that have been reported for pictures are qualitatively similar to the effects that have been reported for words (see Bookbinder & Brainerd, 2017).

On the other hand, there are reasons for being cautious about whether the results of the last two studies will also hold for pictures. There is a group of experiments in the false memory literature which show that false memory behaves differently with pictures than with words

(e.g., Schacter, Israel, & Racine, 1999). For instance, it has been repeatedly found that subjects are less susceptible to false memory when the NS items are pictures rather than words (for a review, see Brainerd & Reyna, 2005). More important, pictures trigger different retrieval processes on memory tests. In particular, pictures activate a conservative process that is known as the distinctiveness heuristic (e.g., Dodson & Schacter, 2002a, 2002b). The essence of this heuristic is that subjects adopt a meta-cognitive expectation that when they remember pictorial items, their memories will be infused with recollective phenomenology (realistic details of the items' prior presentations). When this distinctiveness heuristic is in play, subjects reject items that are not accompanied by recollective phenomenology, no matter how familiar they may seem, which selectively suppresses false memory because NS items were not presented.

Such a difference in the types of memory evidence on which responses to pictures versus words are based could have consequences for valence and arousal effects. Although valence seems to influence false memory more than arousal with words, which is more influential with pictures will obviously depend on whether the distinctiveness heuristic is triggered by increases in one or the other. Suppose, for example, that (a) valence is unrelated to the distinctiveness heuristic, but (b) increases in arousal suppress the distinctiveness heuristic. Under that scenario, arousal would influence false memory (for pictures) more than valence because reliance on the distinctiveness heuristic will fade as arousal increases. Although this possibility is speculative, its implications for emotion-false memory are not. The known differences in the retrieval processes that underlie memory for pictures versus words are of such magnitude that one cannot assume without data that the effects of valence and arousal will be the same.

### Emotional Intensity

The other limiting feature of our results is that emotion-false memory effects may be different at moderate versus extreme levels of valence and arousal. The levels of these properties for items from instruments such as the ANEW, the EWL, and the IAPS are moderate in comparison to some affect-laden situations in everyday life. For instance, valence and arousal are more extreme when witnessing a birth or being threatened with a pistol during a robbery than when reading the word *baby* or viewing a picture of a pistol. Of course, extreme emotional states, especially negatively-valenced ones, are what the law's prevention hypothesis is most concerned with. From that perspective, this hypothesis might be regarded as still viable, despite the contrary findings in the emotion-false memory literature, because more extreme levels of valence and arousal may inoculate memory against distortion. Naturally, this is difficult to test experimentally owing to ethical constraints on inducing emotional states that are comparable to, say, being threatened by a weapon or witnessing a murder.

Nevertheless, there is a group of emotion-false memory experiments involving a widely-studied clinical population that provides some indirect evidence that emotion-false memory effects may be similar for extreme versus modest levels of valence and arousal. These experiments follow standard designs, in which levels of false memory are compared for valenced (usually negative) and neutral items. Their novel feature is that two groups of



subjects participate—namely, individuals with post-traumatic stress disorder (PTSD) diagnoses and matched control subjects. Another novel feature is that false memory is measured for two types of valenced items: (a) standard negative items like those in other emotion-false memory experiments (e.g., the *anger* list in Table 1) and (b) PTSD-related negative items that echo the traumatic events that PTSD subjects experienced. The range of the latter events has encompassed battlefield experiences (Brennen, Dybdahl, & Kapidzic, 2007), serious injuries caused by automobile accidents and assaults (Hauschildt, Peters, Jelinek, & Moritz, 2012), and sexual abuse (Goodman et al., 2011). As examples of standard and PTSD-related negative items when the PTSD group consists of survivors of sexual abuse, false memory might be measured for *sick* (after encoding *cough, fever, ill, flu, ...*) versus *rape* (after encoding *sex, violate, force, struggle, ...*).

The logic of this design is that PTSD-related negative items ought to induce very intense emotional states in the PTSD group but not in the control group. If so and if such states are actually protective against memory distortion, the predictions are that (a) mean false memory levels in the PTSD group will be lower for trauma-related negative items than for standard negative items or neutral items, and that (b) mean false memory levels for PTSD-related items but not for other items will be lower for PTSD subjects than for controls. However, neither of these patterns has been detected in experiments that implemented the aforementioned design, and instead, they have produced three modal results (see Bookbinder & Brainerd, 2016). First, in PTSD subjects, false memory levels have been comparable for PTSD-related items and standard negative items. Second, for PTSD subjects versus controls, false memory levels for PTSD-related items have been comparable for the two groups. Third, quantitative variation in false memory as a function of the emotional content of items has been similar in PTSD subjects versus controls.

### Concluding Comments

It has been found in several experiments that emotional valence elevates false memory, but correlated differences in arousal have typically not been controlled. The work reported here suggests that arousal does not contribute substantially to valence effects in such studies. Words supplied the memory materials in the preponderance of those studies, and for words at least, our data showed that false memory is not closely tied to differences in arousal (Study 2) and that valence is more closely connected to semantic properties that influence false memory (Study 3).

### Acknowledgments

Preparation of this article was supported by National Institutes of Health Grant 1RC1AG036915 to the second author and a grant to the second author from the Department of Agriculture.

### References

- Appelbaum, PS, Uyehara, LA, Elin, MR. Trauma and memory: Clinical and legal controversies. New York, NY: Oxford University Press; 1997.
- Atkins AS, Reuter-Lorenz PA. 2011; Neural mechanisms of semantic interference and false recognition in short-term memory. *NeuroImage*. 56:1726–1734. [PubMed: 21349335]

- Bookbinder SH, Brainerd CJ. 2017; Emotionally negative pictures enhance gist memory. *Emotion*. 17:102–119. [PubMed: 27454002]
- Bookbinder SH, Brainerd CJ. 2016; Emotion and false memory: The context-content paradox. *Psychological Bulletin*. 142:315–351.
- Bradley, MM, Lang, PJ. Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings (Tech Rep No C-1). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida; 1999.
- Brainerd CJ, Holliday RE, Reyna VF, Yang Y, Toglia MP. 2010; Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology*. 107:137–154. [PubMed: 20547393]
- Brainerd, CJ, Reyna, VF. The science of false memory. New York: Oxford University Press; 2005.
- Brainerd CJ, Reyna VF. Complementarity in false memory illusions. *Journal of Experimental Psychology: General*.
- Brainerd CJ, Stein LM, Silveira RA, Rohenkohl G, Reyna VF. 2008; How does negative emotion cause false memories? *Psychological Science*. 19:919–925. [PubMed: 18947358]
- Brainerd CJ, Yang Y, Howe ML, Reyna VF, Mills BA. 2008; Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review*. 15:1035–1053. [PubMed: 19001566]
- Brennen T, Dybdahl R, Kapidzić A. 2007; Trauma-related and neutral false memories in war-induced Posttraumatic Stress Disorder. *Consciousness and Cognition*. 16:877–885. [PubMed: 16901721]
- Budson AE, Todman RW, Chong H, Adams EH, Kensinger EA, Krangel TS, Wright CI. 2006; False recognition of emotional word lists in aging and Alzheimer disease. *Cognitive and Behavioral Neurology*. 19:71–78. [PubMed: 16783129]
- Cann DR, McRae K, Katz AN. 2011; False recall in the Deese-Roediger-McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology*. 64:1515–1542.
- Citron FM, Weekes BS, Ferstyle EC. 2014; How are affective word ratings related to lexicosemantic properties? Evidence from the Sussex Affective Word List. *Applied Psycholinguistics*. 35:313–331.
- Deese J. 1959; On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology*. 58:17–22. [PubMed: 13664879]
- Dodson CS, Schacter DL. 2002a; Aging and strategic retrieval processes: Reducing false memories with a distinctiveness heuristic. *Psychology and Aging*. 17:405–415. [PubMed: 12243382]
- Dodson CS, Schacter DS. 2002b; When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*. 46:782–803.
- Dehon H, Larøi F, Van der Linden M. 2010; Affective valence influences participant’s susceptibility to false memories and illusory recollection. *Emotion*. 10:627–639. [PubMed: 21038946]
- El Sharkawy J, Groth K, Vetter C, Beraldi A, Fast K. 2008; False memories of emotional and neutral words. *Behavioural Neurology*. 19:7–11. [PubMed: 18413909]
- Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. 1999; Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 4:272–299.
- Gallo DA, Foster KT, Johnson EL. 2009; Elevated false recollection of emotional pictures in young and older adults. *Psychology and Aging*. 24:981–988. [PubMed: 20025411]
- Garry M, Manning CG, Loftus EF, Sherman SJ. 1996; Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*. 3:208–214. [PubMed: 24213869]
- Gomes CFA, Brainerd CJ, Stein LM. 2012 Effects of emotional valence and arousal on recollective and nonrecollective recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.
- Goodman GS, Ogle CM, Block SD, Harris LS, Larson RP, Augusti EM, Urquiza A. 2011; False memory for trauma-related Deese-Roediger-McDermott lists in adolescents and adults with histories of child sexual abuse. *Development and Psychopathology*. 23:423–438. [PubMed: 23786687]
- Hauschildt M, Peters MJV, Jelinek L, Moritz S. 2012; Veridical and false memory for scenic material in posttraumatic stress disorder. *Consciousness and Cognition*. 21:80–89. [PubMed: 22154736]

- Howe ML. 2007; Children's emotional false memories. *Psychological Science*. 18:856–860. [PubMed: 17894601]
- Huntsinger JR. 2013; Does emotion directly tune the scope of attention? *Current Directions in Psychological Science*. 22:265–270.
- Kanske P, Kotz SA. 2010; Leipzig affective norms for German: A reliability study. *Behavior Research Methods*. 42:987–991. [PubMed: 21139165]
- Kassin SM, Ellsworth PC, Smith VL. 1989; The “general acceptance” of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist*. 44:1089–1098.
- Kucera, H, Francis, W. *Computational analysis of present day American English*. Providence, RI: Brown University Press; 1967.
- Laney, C, Loftus, EF. Truth in emotional memories. In: Bornstein, BH, Wiener, RL, editors *Emotion and the law*. New York, NY: Springer; 2010. 157–183.
- Lang, PJ, Bradley, MM, Cuthbert, BN. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Gainesville, FL: University of Florida; 2008. Tech Rep No A-8
- Loftus EF. 1993; The reality of repressed memories. *American Psychologist*. 48:518–537. [PubMed: 8507050]
- Loftus, E, Ketcham, K. *The myth of repressed memory: False memories and allegations of sexual abuse*. New York, NY: Macmillan; 1996.
- Nelson, DL, McEvoy, CL, Schreiber, TA. *The University of South Florida word association, rhyme, and word fragment norms*. University of South Florida; 1999. Unpublished manuscript
- McNally RJ. 2003; Recovering false memories of trauma: A view from the laboratory. *Current Directions in Psychological Science*. 12:32–35.
- Mirandola C, Toffalini E, Grassano M, Cornoldi C, Melinder A. 2014; Inferential false memories of events: Negative consequences protect from distortions when the events are free from further elaboration. *Memory*. 22:451–461. [PubMed: 23663060]
- Paivio A, Yuille JC, Madigan SA. 1968; Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*. 76:1–25.
- Roediger HL III, McDermott KB. 1995; Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 21:803–814.
- Roediger HL III, Watson JM, McDermott KB, Gallo DA. 2001; Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*. 8:385–407. [PubMed: 11700893]
- Schacter DL, Israel L, Racine C. 1999; Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*. 40:1–24.
- Shaw J, Porter S. 2015; Constructing rich false memories of committing crime. *Psychological Science*. 26:291–301. [PubMed: 25589599]
- Spanos, NP. *Multiple identities & false memories: A sociocognitive perspective*. Washington, DC: American Psychological Association; 1996.
- Spanos NP, Cross PA, Dickson K, DuBreuil SC. 1993; Close encounters: An examination of UFO experiences. *Journal of Abnormal Psychology*. 102:624–632. [PubMed: 8282933]
- Stein, NL, Ornstein, PA, Tversky, B, Brainerd, C. *Memory for everyday and emotional events*. Mahwah, NJ: Lawrence Erlbaum Associates; 1997.
- Storbeck J. 2013; Negative affect promotes encoding of and memory for details at the expense of the gist: Affect, encoding, and false memories. *Cognition and Emotion*. 27:800–819. [PubMed: 23134550]
- Toglia, MP, Battig, WF. *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum; 1978.
- Toglia MP, Neuschatz JS, Goodwin KA. 1999; Recall accuracy and illusory memories: When more is less. *Memory*. 7:233–256. [PubMed: 10645381]
- Vo MLH, Conrad M, Kuchinke L, Urton K, Hofmann MJ, Jacobs AM. 2009; The Berlin Affective Word List reloaded (BAWL-R). *Behavior Research Methods*. 41:534–538. [PubMed: 19363195]
- Warriner AB, Kuperman V, Brysbaert M. 2013; Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*. 45:1191–1207. [PubMed: 23404613]

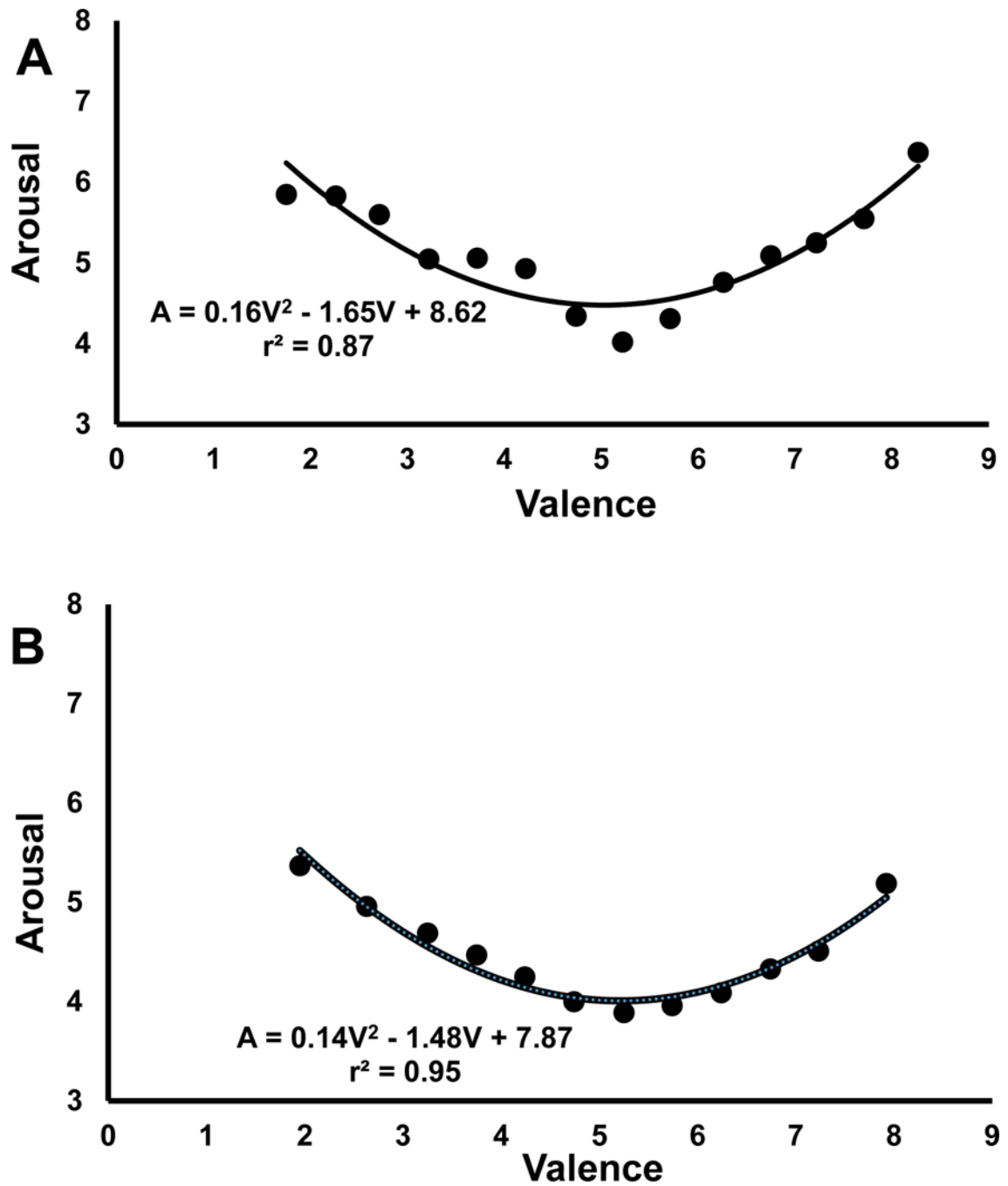
Wessa M, Kanske P, Neumeister P, Bode K, Heissler J, Schönfelder S. 2010; EmoPics: Subjektive und psychophysiologische evaluationen neuen bildmaterials für die klinisch-biopsychologische forschung. Zeitschrif für Klinische Psychologie und Psychotherapie. 39(Suppl. 1/11):77.

Author Manuscript

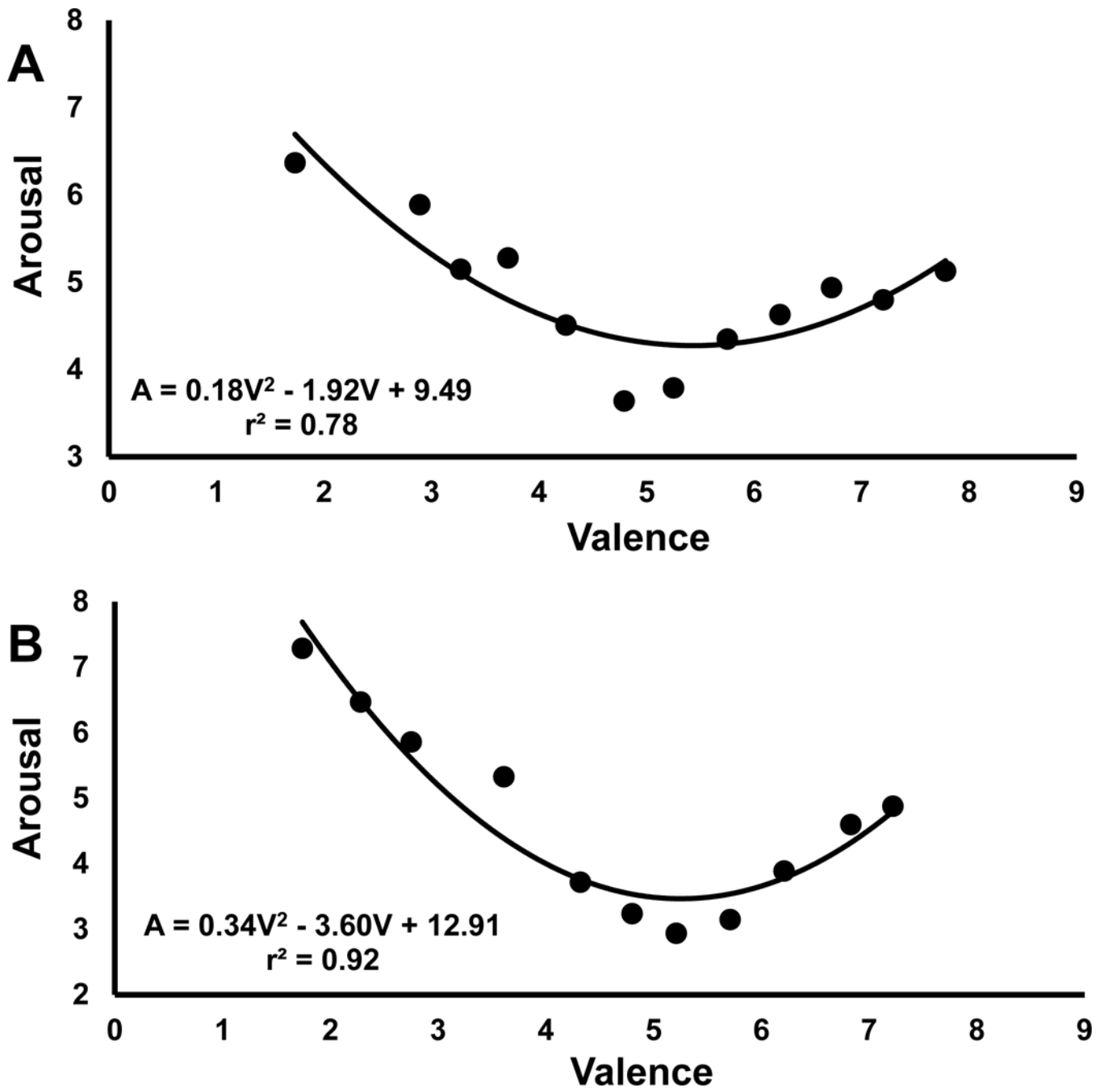
Author Manuscript

Author Manuscript

Author Manuscript

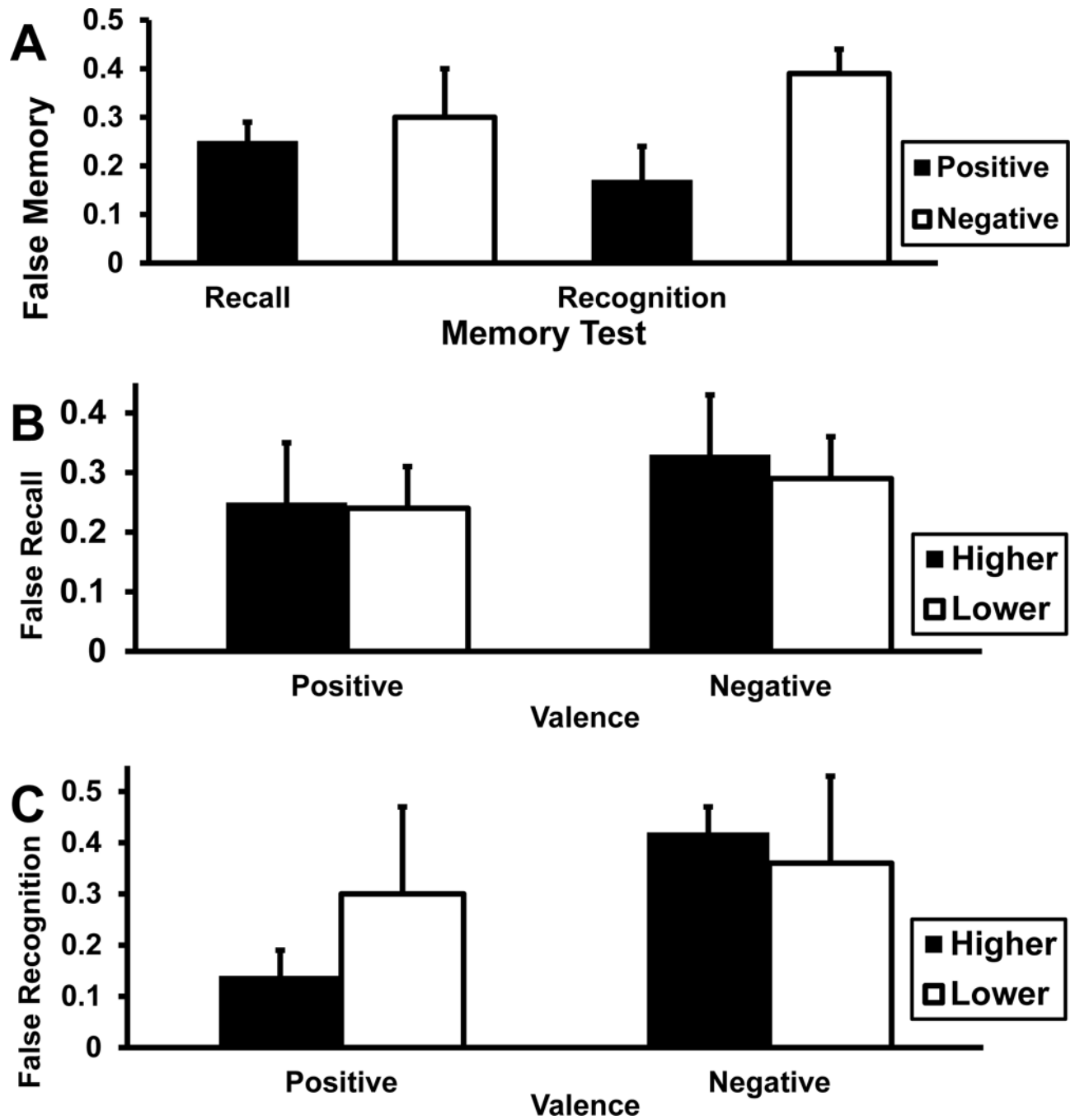


**Figure 1.** Best-fitting valence-arousal functions for emotional word norms. Panel A = ANEW norms, and Panel B = WKB norms.

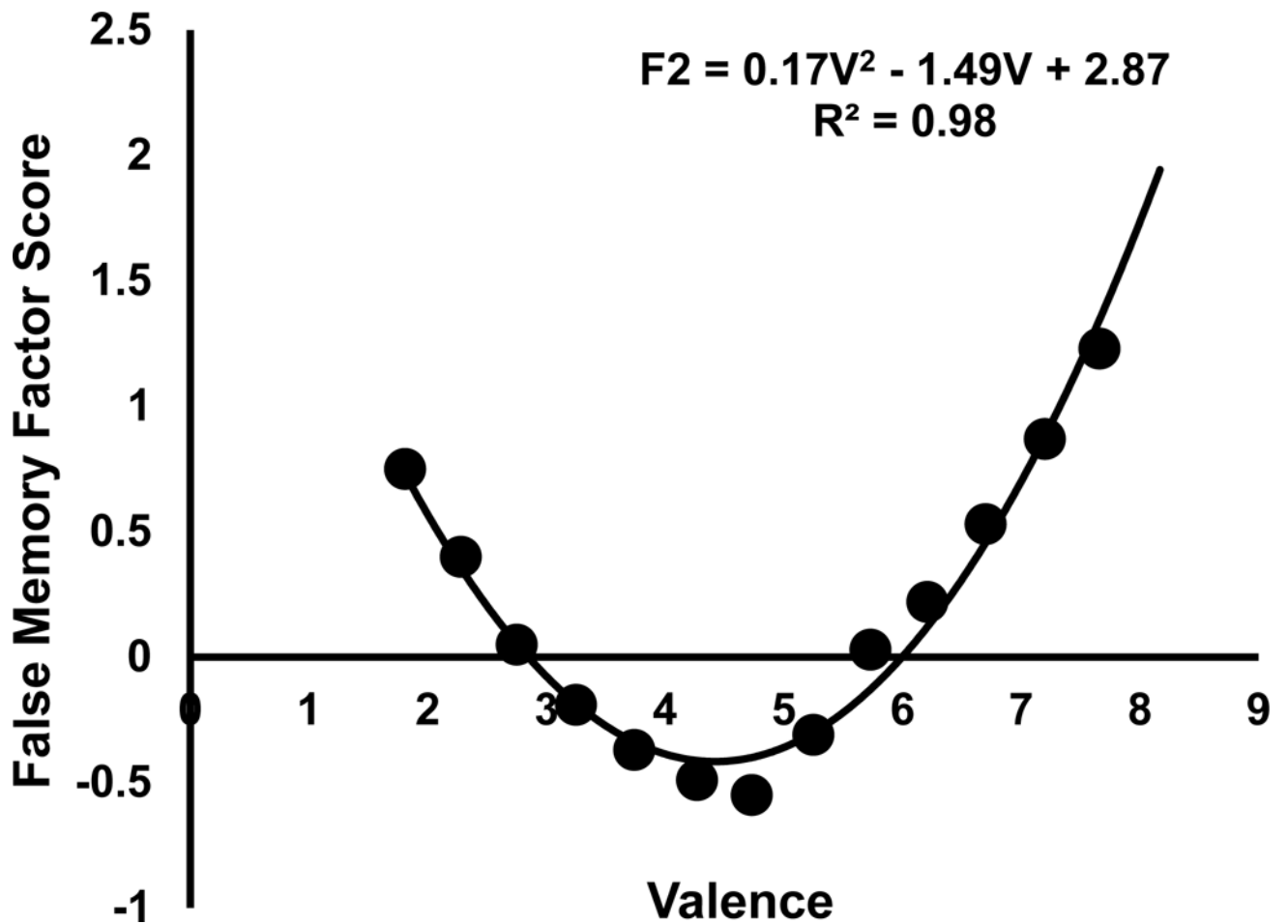


**Figure 2.** Best-fitting valence-arousal functions for emotional picture norms. Panel A = IAPS norms, and Panel B = EmoPics norms.





**Figure 3.** Panel A = Effects of valence on false recall and false recognition in the EWL emotion-false memory norms. Panel B = relation between arousal and false recall in the EWL emotion-false memory norms. Panel C = relation between arousal and false recognition in the EWL emotion-false memory norms.



**Figure 4.** Relation between valence ratings of the Toglia-Batting word pool and the words' scores on F1 = the true memory factor (categorizability, concreteness, imagability) and F2 = the false memory factor (meaningfulness, familiarity, log frequency, number-of-attributes).

**Table 1**

## Experimental Materials and Procedures for False Memory Illusions

Old words (O)	New-similar words (NS)	New-different words (ND)
Neutral: <i>table, couch, desk, sofa.</i>	<i>chair, seat</i>	<i>city, music</i>
Positive: <i>girl, beautiful, cute, nice</i>	<i>pretty, sweet</i>	<i>soft, warm</i>
Negative: <i>mad, fear, hate, temper</i>	<i>anger, mean</i>	<i>spider, thief</i>

*Note.* Subjects study word lists composed of related items, such as those in the left hand column, and they respond to a recall test or to a recognition test on which three types of test probes are presented: old (O) list words, new-similar (NS) words, and new-different (ND) words. On recall tests, the subjects' task is to recall only O words. On recognition tests, the subjects' task is to decide whether each test probe is O.

**Table 2**

Rotated Loadings of Semantic Properties on the Two Factors that were Extracted from the Toglia-Battig Norms

Semantic properties	Factor 1: true memory	Factor 2: false memory
True memory cluster:		
Concreteness	.95	
Imagability	.94	
Categorizability	.94	
False memory cluster:		
Familiarity		.84
Meaningfulness		.84
Log frequency		.78
Number-attributes		.70

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript