Research Article

Psychological Science
2018, Vol. 29(10) 1706–1715

DOI: 10.1177/0956797618780353
www.psychologicalscience.org/PS

The Emotional-Ambiguity Hypothesis: A Large-Scale Test

C. J. Brainerd
Department of Human Development and Human Neuroscience Institute, Cornell University

Abstract

Valence and arousal are core dimensions of emotion, but the relation between them has eluded scientific consensus. The emotional-ambiguity hypothesis is the first new model of this relation to appear in some years. It introduces the novel principle that the relation between valence and arousal is controlled by a variable that is not traditionally measured: the uncertainty of perceived valence. A comprehensive evaluation of this principle was conducted using publicly available emotional word and emotional picture databases. There was compelling support for the hypothesis in both types of databases and for both positive and negative valence: The strength of the relation between perceived arousal and perceived positivity or negativity decreased linearly as valence perceptions became more ambiguous. These results explain some puzzling facts about the valence–arousal relation that figure prominently in literature reviews, and they provide a solution to the problem of how to remove arousal confounds from valence effects.

Keywords

emotional control, emotions

Received 3/19/18; Revision accepted 5/11/18

Since the beginnings of scientific psychology, there has been widespread interest in how human emotion influences such fundamental processes as memory, reasoning, attitude formation, and decision making. To take but one example from this long lineage of research, how the emotional content of experience affects episodic memory has been studied since the work of Flügel (1925) and Wohlgemuth (1923). Extensive literature has accumulated on emotional content effects in true memory for actual events (for reviews, see Buchanan, 2007; Kensinger, 2004, 2009) and false memory for fictitious events (for reviews, see Bookbinder & Brainerd, 2016; Stein, Ornstein, Tversky, & Brainerd, 1997). Brain regions that support some of these emotion-memory effects have been identified (e.g., Adolphs et al., 2005; Payne & Kensinger, 2011), and emotion-memory effects have been found to vary developmentally, with young adults exhibiting heightened sensitivity to negative content (for a review, see Vaish, Grossmann, & Woodward, 2008) but older adults exhibiting heightened sensitivity to positive content (for a review, see Murphy & Isaacowitz, 2008).

Regardless of one's field of study, determining how emotion affects basic psychological processes requires reliable methods of inducing affective states, which in turn requires consensus as to the dimensions of emotion. A dominant approach has been to rely on materials that have been normed for the subjective levels of valence (positivity or negativity) and arousal (calming or exciting) that they provoke (e.g., Bradley & Lang, 1999; Lang, Bradley, & Cuthbert, 2008). This approach is often traced to Wundt (1912), who proposed that although emotion is multifaceted, two invariable properties are that it is valenced and arousing. That proposal derives support from data showing that when subjects judge items for perceived emotionality, the bulk of the variance is explained by valence and arousal judgments about those same items (Kuppens, Tuerlinckx, Russell, & Barrett, 2013; Mattek, Wolford, & Whalen, 2017). Consequently, multiple sets of materials have been created in which thousands of words and pictures have been rated on bipolar scales of valence (negative to positive)

Corresponding Author:
C. J. Brainerd, Cornell University, Department of Human Development and Human Neuroscience Institute, G331 Martha Van Rensselaer Hall, Ithaca, NY 14853
E-mail: cb299@cornell.edu

Table 1. Six Models of the Relation Between Subjective Impressions of Valence and Arousal

Model	Description
Bipolar	Valence and arousal are independent, but higher levels of positivity mean lower levels of negativity and vice versa.
Bivariate	Positivity and negativity are independent, but higher levels of positivity or negativity mean higher levels of arousal.
Independence	Valence and arousal are independent dimensions.
Positive linear	Arousal increases linearly as valence moves from high negativity through neutrality to high positivity.
Negative linear	Arousal decreases linearly as valence moves from high negativity through neutrality to high positivity.
V shaped	Arousal increases linearly as negativity increases from a neutral point, and it also increases linearly as positivity increases from a neutral point.

and arousal (calming to exciting; e.g., Bradley & Lang, 1999; Lang et al., 2008; Vo et al., 2009; Warringer et al., 2013; Wessa et al., 2010). Items from these databases are widely used both to manipulate the emotional content of information that subjects encode while performing various tasks (e.g., recalling word lists, making decisions) and to manipulate mood (see Bookbinder & Brainerd, 2016).

Owing to the prevalence of this approach, it is essential to understand the relation between the dimensions of valence and arousal. In their review of models of this relation, Kuppens et al. (2013) noted that six models have been proposed during the past century; they are summarized in Table 1. In another review, Mattek et al. (2017) focused on the only two models that have generated significant empirical support: the bipolar and bivariate models. The bipolar model posits, first, that perceived negativity increases as perceived positivity decreases (and vice versa) and, second, that changes in arousal are distinct from changes in valence. In contrast, the bivariate model posits, first, that changes in perceived positivity and negativity are distinct from each other and, second, that changes in arousal are a function of changes in valence. Although these models have been the most influential accounts of the valence–arousal relation for some time, Kuppens et al. and Mattek et al. concluded that there are key findings that run counter to each. For instance, the bipolar model predicts that unipolar (low to high) judgments of items' perceived positivity will be correlated with unipolar judgments of perceived negativity, and the bivariate model predicts that bipolar judgments of perceived valence (negative to positive) and arousal (calming to exciting) will be correlated on both the positive and negative sides of the valence scale. Although both types of correlations have been reported, there are several studies in which (a) correlations of the first type were not reliable (Mattek et al., 2017) or (b) correlations of the second type explained only modest amounts of variance (Kuppens et al., 2013).

In response, Mattek et al. (2017) formulated a new interpretation, which I will refer to as the *emotional-ambiguity hypothesis*. It treats the valence–arousal relation as an example of a general statistical situation, in which an unmeasured variable Z is correlated with the relation between two measured variables, X and Y, as opposed to each variable being separately correlated with Z. This contrasts with the more familiar third-variable-artifact situation of spurious correlations between X (e.g., sales of hockey sticks) and Y (e.g., sales of Caribbean vacations) being claimed because each is separately correlated with Z (e.g., season of the year). Instead, it is analogous to Simpson's (1951) paradox, wherein authentic correlations between X and Y appear and disappear, or even reverse direction, depending on levels of Z. Although examples of third-variable artifacts are far more numerous in the psychological literature, there are also instances of Simpson's paradox—for example, studies in which it has been used to explain judgment biases (Fiedler, 2000), pseudocontingencies (Kutzner & Fiedler, 2017), illusory dissociations between memory and reasoning (Howe, Rabinowitz, & Grant, 1993), and reversals in the direction of activity correlations between brain regions (Roberts, Hach, Tippett, & Addis, 2016).

Returning to the emotional-ambiguity hypothesis, it specifies that the relation between perceived valence and perceived arousal is controlled by an unmeasured variable—explicitly, *valence ambiguity*, which is the degree of indefiniteness or uncertainty in people's subjective impressions of the valence of an item or event. This principle captures the notion that some items consistently provoke specific feelings of positivity or negativity, but other items provoke feelings that are more variable over measurement occasions, subjects, or both. As examples of the latter, consider the words *mountain* and *religion*. When judging their valence, subjects might recall the majestic beauty of mountains and the selflessness of acts of religious charity, stimulating very positive feelings. At other times, they might recall deaths from falls and acts of religious violence, stimulating very negative feelings. Across subjects and measurement occasions, the result is items whose perceived valence is hazy.

The central tenet of the emotional-ambiguity hypothesis is that the unmeasured ambiguity variable controls the valence–arousal relation as follows. Valence and arousal judgments are strongly correlated when perceived positivity or negativity is unambiguous—more intense valence in either direction means higher arousal—but the relation wanes as ambiguity increases, eventually approaching zero. (Note that ambiguity is different from the central tendency of valence judgments; items with similar mean valence could differ dramatically in how variable their ratings are.) Because these predictions are quite novel, I conducted a comprehensive assessment of them using large-scale emotional word and emotional picture norms.

In order to conduct this assessment, I required a measure that could serve as a valid proxy for the theoretical concept of valence ambiguity, one that is distinct from items' mean levels of judged valence. Previously, Mattek et al. (2017) discussed a tripartite measure, in which subjects make categorical judgments about individual items by classifying each as positive, negative, or neutral, in addition to the standard method of judging them on bipolar numerical scales of valence and arousal. Valence ambiguity is said to increase as the frequencies of positive and negative classifications of an item become more equal. The prediction is that correlations between numerical judgments of valence and arousal will decrease as these frequencies become more equal.

Although the tripartite measure has high face validity, it cannot be used to test this prediction with large-scale emotional word and picture norms. Bipolar numerical ratings are the only valence measure in such norms. Fortunately, they provide another ambiguity measure that also has high face validity: variability of valence ratings. Conceptually, such variability captures the sense of differences in the uncertainty of perceived valence over subjects and measurement occasions. With this index, the predicted instability in the valence–arousal relation was evaluated by (a) arraying the items in emotional word and picture databases according to the degree of variability in their valence ratings and (b) computing best-fitting regression equations for valence and arousal ratings separately for different levels of those arrays. There were two key findings. First, there was remarkably strong confirmation that the valence–arousal relation is conditional on valence ambiguity in all databases and for both positive and negative valence. Second, that confirmation took the form of a linear decrease in the strength of the correlation between perceived valence and arousal as valence ambiguity increased.

Method

Truly comprehensive tests of the emotional-ambiguity hypothesis necessitate that valence and arousal judgments be available for massive numbers of items. Although such judgments have most often been gathered for small sets of items as a prelude to particular experiments (e.g., Brainerd, Holliday, Reyna, Yang, & Toglia, 2010), there are some publicly available databases in which they were gathered for hundreds or thousands of items. Those norms focus on two types of items, words and pictures, and for the sake of generality, I examined multiple examples of both types of norms in the present study. The six databases that were included and the valence ambiguity measure that was used for all of them are summarized below.

Emotional word databases

Three word databases were used: the Affective Norms for English Words (ANEW; Bradley & Lang, 1999); the Warriner, Kuperman, and Brysbaert (WKD; 2013) norms; and the emotional version of Toglia and Battig's (1978) semantic word norms (Toglia-Battig emotional, or TBE; Brainerd & Bookbinder, 2018). The ANEW database evolved from Osgood, Suci, and Tannenbaum's (1957) landmark work on the semantic differential and is the most widely used set of emotional word norms. Students in introductory psychology classes judged sets of 56 words for how each word made them feel. Individual words were rated on bipolar scales from 1 to 9 of perceived valence and perceived arousal, using the Self-Assessment Manikin. Over subjects, these judgments were gathered for 1,034 words, with many being drawn from earlier emotional word pools that contained smaller numbers of items.

The WKD is a similar but vastly larger database. In these norms, valence and arousal judgments were gathered for 13,915 English words. The subjects were 1,827 adults recruited via the Amazon Mechanical Turk crowdsourcing website. Individual subjects rated 346 to 350 words. Subjects judged how individual words made them feel, using bipolar scales from 1 to 9 for valence and for arousal, similar to those used in the ANEW. A key difference between the WKD and ANEW procedures was that individual WKD subjects made either valence or arousal judgments about words, whereas ANEW subjects made both types of judgments.

The TBE is the most commonly used collection of semantic norms in the mainstream memory literature. In the original database, which was generated by Toglia and Battig (1978), words were not rated for perceived valence and arousal. Instead, more than 2,500 subjects judged 2,854 words for levels of seven semantic properties that are known to have robust effects on recognition and recall: categorizability, concreteness, imagability, meaningfulness, familiarity, number of attributes, and pleasantness. Judgments were made on 7-point unipolar scales. To study the relation between meaning and emotion, Brainerd and Bookbinder (2018) revised these norms to include bipolar valence and arousal ratings for 2,184 of the original 2,854 words. The words were rated for how they made subjects feel, using the same scales of valence and arousal that were used to generate the ANEW and WKD norms. The revised database is available from the author on request.

Emotional picture databases

The emotional-ambiguity hypothesis was also investigated in three emotional picture databases: the International Affective Picture System (IAPS; Lang et al., 2008), the Nencki Affective Picture System (NAPS; Marchewka, Łukasz, Jednoróg, & Grabowska, 2014), and the EmoPics (Wessa et al., 2010). Of available emotional picture norms, the IAPS is by far the most widely used. It is a picture follow-up to the ANEW, conducted by the same authors and relying on very similar procedures. Undergraduate students who were enrolled in introductory psychology courses judged how individual color photographs of positive (e.g., babies, puppies), negative (e.g., cemeteries, guns), and neutral (e.g., boxes, light bulbs) objects made them feel. Judgments were performed with the same Self-Assessment Manikin and bipolar valence and arousal scales as in the ANEW norms. Each subject judged 60 pictures, each picture was presented to approximately 100 subjects, and 956 pictures in total were presented.

The NAPS and the EmoPics are additional picture norms in which subjects made valence and arousal judgments about positive, negative, and neutral color photographs on the same bipolar scales used in the IAPS. The distinguishing features of the NAPS are that substantially more pictures (1,356) were rated than in the IAPS, and each picture belonged to one of five broad categories (people, faces, animals, miscellaneous objects, and landscapes). The distinguishing features of the EmoPics are that a substantially smaller number of pictures (390) were rated than in the IAPS, and judgment tasks were administered in the German language.

Measuring valence ambiguity

All of these databases contain variability information in the form of standard deviations of the valence and arousal judgments that subjects made. Valence standard deviations were used as an intuitive index of the ambiguity of the perceived valence of individual items; the higher the standard deviation, the greater the presumptive ambiguity. A noteworthy methodological point is that the range of valence standard deviations was substantial in all of the databases—in the neighborhood of ±2.5 points, on average, on the 9-point scale—making it a statistically attractive measure for determining whether the valence–arousal relation was conditional on the level of valence ambiguity.

Identifying ambiguity with variability is natural, of course, and it is distinct from the central tendency of valence judgments about an item. An item's perceived valence may vary considerably, even though the central tendency of its perceived valence is clearly positive or negative. For instance, consider the words *heaven* and *destruct*. Not surprisingly, according to the ANEW and the WKD, the mean valence rating of *heaven* is well into the positive region of the bipolar scale, whereas that of *destruct* is well into the negative region. However, the valence standard deviation for both is in the upper 2% of standard deviations for these norms. That level of variability is so high that the lower half of *heaven*'s confidence interval extends into the negative region of the scale, and the upper half of *destruct*'s confidence interval extends into the positive region.

In addition to being conceptually distinct from the central tendency of valence judgments, valence standard deviations are not strongly associated with mean valence ratings. I computed the correlations between valence standard deviations and valence means separately for items with positive mean valences (> 5) and for items with negative mean valences (< 5), in all of the word and picture databases. These correlations were reliable because hundreds or thousands of items were involved in each, but they were only slightly above zero (mean $r < .03$).

Results

Valence–arousal relations can differ for positive versus negative items (Kuppens et al., 2013). For instance, Brainerd and Bookbinder (2018) and Citron, Weekes, and Ferstyle (2014) reported that the best-fitting quadratic functions for the relation between bipolar valence and arousal judgments about selected pools of words were asymmetrical. Specifically, unit changes in valence on the negative sides of valence scales produced larger changes in arousal than did unit changes on the positive sides, and hence, the average level of arousal associated with a given valence value on the negative side was higher than for the corresponding value on the positive side. Therefore, in the present research, each database was split into a positive subfile (mean valence ratings > 5) and a negative subfile (mean valence ratings < 5), and predictions were tested separately for each subfile. I report the findings for emotional words first, followed by the findings for emotional pictures, followed by an account of the global pattern of conditionalization.

Emotional words

The positive and negative subfiles of the word databases were analyzed in three steps. First, the items in each subfile were arrayed according to the ambiguity index, from the lowest to highest valence standard deviation. Second, the stratified subfiles were sliced into Vincentized quantiles, with the number of quantiles depending on the number of items whose valence and arousal had been rated. Vincentized deciles were used for the WKD norms because more than 5,500 words were rated in both the positive and negative subfiles. For the ANEW, Vincentized quintiles were used because fewer than 1,000 words were rated in the positive and negative subfiles. For the TBE, quintiles were used for the negative subfile (755 words), but deciles were used for the much larger positive subfile (1,429 words). Third, within each quantile of each subfile, the best-fitting regression equation $A = f(V)$ was determined, where A stands for mean arousal and V for mean valence, by fitting several possibilities to the data: the linear regression equation, the most commonly used two-parameter nonlinear equations (e.g., exponential, hyperbolic, log, power), and the quadratic and cubic equations. The linear equation always supplied the best fit, and thus, all of the results that follow are for linear regression.

The emotional-ambiguity hypothesis assumes that the valence–arousal relation is conditional on valence ambiguity, with both positive and negative valence becoming progressively weaker predictors of arousal as ambiguity increases. The corresponding predictions about the Vincentized subfiles are that (a) valence will correlate most strongly with arousal in quantiles where the mean valence standard deviation is lowest, and (b) the correlation will decrease steadily as one moves to quantiles with larger mean valence standard deviations. Both predictions were confirmed in all six word subfiles.

Of all the word and picture databases, the WKD supplies the most sensitive tests of these predictions because the pools of rated items are vast. The results are displayed in Figure 1a (negative subfile) and Figure 1b (positive subfile), where it can be seen that the data fell out as predicted. The coordinates of the individual data points in these figures are the correlation coefficient between items' mean valence and arousal ratings (ordinate) and the mean valence standard deviation (abscissa), within each Vincentized decile. As predicted, the correlations were highest in the deciles with the lowest mean valence standard deviation. Thereafter, it can be seen that correlations decreased linearly as a

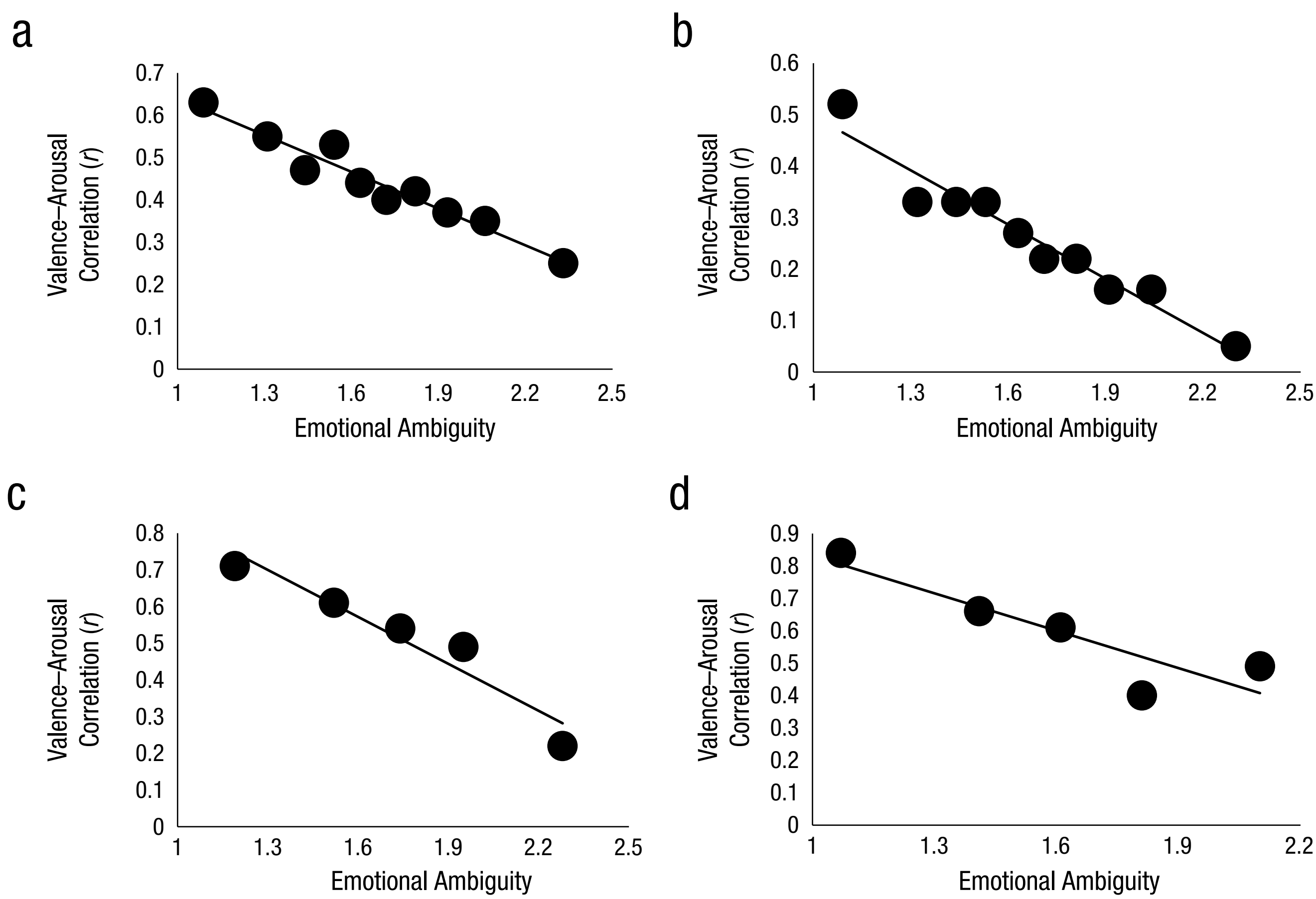

Fig. 1. Best-fitting regression equations for the relation between emotional ambiguity and the valence–arousal correlation for the (a) negative and (b) positive subfiles of the Warriner, Kuperman, and Brysbaert (2013) norms and for the (c) negative and (d) positive subfiles of the Affective Norms for English Words (Bradley & Lang, 1999).

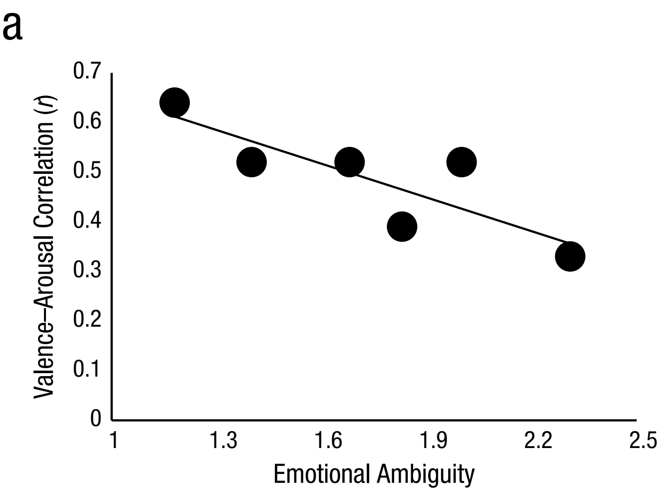

Fig. 2. Best-fitting regression equations for the relation between emotional ambiguity and the valence–arousal correlation for the (a) negative and (b) positive subfiles of the Toglia-Battig-emotional database (Brainerd & Bookbinder, 2018).

function of increasing ambiguity. In addition to the valence–arousal relation being linearly conditional on valence ambiguity, it was weaker for positive than for negative valence in all deciles, with mean correlations over deciles of .42 (negative) and .27 (positive).

Turning to the ANEW and the TBE, the outcome was similar. The data for the ANEW appear in Figure 1c (negative valence) and Figure 1d (positive valence), and the data for the TBE appear in Figure 2a (negative valence) and Figure 2b (positive valence). Two findings should be noted. First, the quantile with the lowest mean valence standard deviation always yielded the highest valence–arousal correlation, and second, correlations always declined linearly as mean valence standard deviation increased. Negative valence was a better predictor of arousal than positive valence in all quantiles with the ANEW, similar to results for the WKD, but the opposite was true with the TBE.

Clearly, valence ambiguity was implicated in the word databases because the valence–arousal relation was linearly conditional on the ambiguity index in all instances, for both positive and negative valence. In addition, overall, the strength of the conditionalization was greater for negative than for positive valence. In Figures 1 and 2, the strength of the conditionalization is indexed by how much variance was accounted for when valence–arousal correlations were fit to mean valence standard deviation. In these fits, the average percentage of variance accounted for was greater for the negative subfiles than for the positive subfiles (85% vs. 70%).

Emotional pictures

The analysis for emotional picture databases was the same as for the emotional word databases, and so were the principal results. The subfiles of the IAPS, NAPS, and EmoPics were split into Vincentized quantiles, and the best-fitting regression equation was located for each quantile. (It was always the linear equation.) Because the numbers of items in the picture databases were smaller than the numbers of items in the word databases, I established a Vincentizing criterion that required that individual quantiles contain valence and arousal ratings for a minimum of 100 pictures. On that basis, octiles were used for the positive subfile of the NAPS, quintiles were used for the negative subfile of the NAPS and the positive subfile of the IAPS, quartiles were used for the negative subfile of the IAPS, and halves were used for the positive subfile of the EmoPics. The negative subfile of the EmoPics was not Vincentized because it did not contain a sufficient number of pictures.

The results for the IAPS and the NAPS appear in Figure 3. For the IAPS, valence–arousal correlations for negative valence (Fig. 3a) and positive valence (Fig. 3b) were strongest in the quantile with the lowest valence ambiguity, and the strengths of the correlations declined linearly as ambiguity increased. For the NAPS, the correlations for negative valence (Fig. 3c) and positive valence (Fig. 3d) were also strongest in the quantile with the lowest valence ambiguity. The correlations in the negative subfile declined linearly as ambiguity increased, but the relation was different in the positive subfile. As in the negative subfile, correlations declined over the first seven quantiles. However, the correlation rose in the final quantile, so quadratic regression provided a better fit to the full set of data points than did linear regression. Finally, for the EmoPics, the valence–arousal correlation declined from .87 in the first half to .67 in the second half.

As with the emotional word databases, the strength of the conditionalization of the valence–arousal relation on valence ambiguity was more pronounced for negative than for positive valence: The mean percentage of

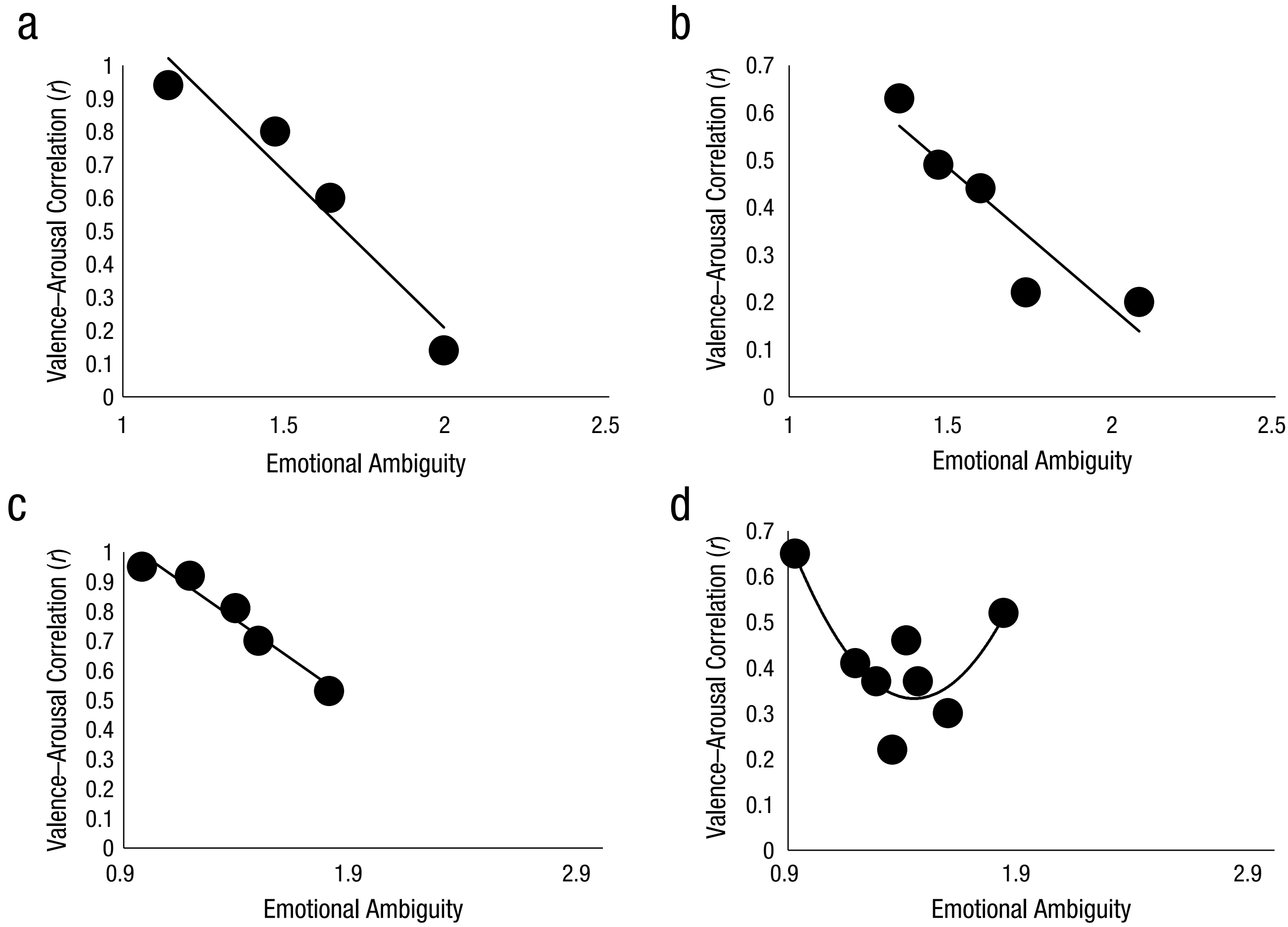

Fig. 3. Best-fitting regression equations for the relation between emotional ambiguity and the valence–arousal correlation for the (a) negative and (b) positive subfiles of the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008) and for the (c) negative and (d) positive subfiles of the Nencki Affective Picture System (Marchewka, Łukasz, Jednoróg, & Grabowska, 2014).

variance that was accounted for by the regressions was 95% for the negative subfiles and 77% for the positive subfiles.

The global picture

Table 2 supplies a simple, qualitative portrait of what this study reveals about whether the relation between valence and arousal was conditional on the ambiguity of perceived valence. For each of the subfiles that generated the results in Figures 1 to 3, Table 2 displays three valence–arousal correlations: the full-sample correlation for all items, the correlation for items in the quantile with the highest mean valence standard deviation, and the correlation for items in the quantile with the lowest mean valence standard deviation. The ordering predicted by the emotional-ambiguity hypothesis was that correlations for the lowest quantile would be greater than correlations for the full sample, which in turn would be greater than correlations for the highest quantile. It can be seen that this ordering was present in all subfiles, except the positive subfile of the NAPS (where correlations for the full sample were less than those for the highest quantile). Based on the grand means for all subfiles, analyses showed that valence–arousal correlations accounted for more than half the variance (53%) in quantiles with the lowest mean valence standard deviation, compared with 10% in quantiles with the highest mean valence standard deviation. Importantly for the general hypothesis of conditionalization, r-to-z transformations showed that the correlation for the lowest quantile was reliably larger than the correlation for the highest quantile in all rows of Table 2, except for the last one. That row contains the data for the positive subfile of the NAPS, which was the only data set in which the valence–arousal correlation increased between the next-to-last and the last quantile.

Discussion

According to a new model of the valence–arousal relation (Mattek et al., 2017), the relation will wax and

Table 2. Valence–Arousal Correlations for Items in the Quantiles With the Lowest and Highest Mean Valence Standard Deviation and for Items in the Full Sample, Separately for Each Database

Database and subfile	Lowest quantile	Highest quantile	Full sample
Emotional words			
Affective Norms for English Words			
Negative	.71**	.22**	.46**
Positive	.84**	.49**	.64**
Toglia-Battig emotional			
Negative	.64**	.33**	.45**
Positive	.38**	.20**	.28**
Warriner, Kuperman, and Brysbaert norms			
Negative	.63**	.25**	.42**
Positive	.62**	.05*	.27**
Emotional pictures			
EmoPics			
Negative			
Positive	.87**	.67**	.79**
International Affective Picture System			
Negative	.94**	.14**	.70**
Positive	.63**	.20**	.37**
Nencki Affective Picture System			
Negative	.95**	.52**	.78**
Positive	.65**	.52*	.48**

*$p < .002$. **$p < .0001$.

wane as a function of an unmeasured variable: the ambiguity of perceived valence. The basic prediction was that the tendency for more intensely positive or negative feelings to be associated with higher states of arousal would weaken as perceived positivity or negativity became more uncertain. That idea received compelling support. Another key result, which was not specifically predicted by the hypothesis, was linear conditionalization. In all but one of the analyses, changes in valence–arousal correlations were linear functions of valence ambiguity.

These results have two unexpected dividends. The first is that they explain a pair of puzzling findings about the valence–arousal relation that have been noted in literature reviews. On the one hand, although valence and arousal judgments were correlated on both the positive and negative sides of bipolar scales, regression equations accounted for only modest amounts of variance when they were fitted to the data of individual items (Kuppens et al., 2013). On the other hand, impressive amounts of variance were accounted for when (a) items were arrayed in order of their valence ratings and (b) the same equations were fit to mean valence and arousal ratings of blocks of consecutive items (Brainerd & Bookbinder, 2018).

Both findings can now be seen as consequences of the hidden ambiguity variable. Concerning the first finding, valence ambiguity was left uncontrolled when regressions were computed for judgments about individual items, and that will inflate the error terms of regression equations: $A_i = \beta_0 + \beta_1 V_i + \varepsilon$ for linear regression and $A_i = \beta_0 + \beta_1 V_i^2 - \beta_2 V_i + \varepsilon$ for quadratic regression. The stronger the conditionalization of the valence–arousal relation on valence ambiguity, the larger ε will be when ambiguity is free to vary. Concerning the other finding, the second method of computing regressions reduced the freedom of valence ambiguity to vary by averaging valence ratings over blocks of items. Reducing ambiguity's freedom to vary decreased ε and enhanced fit.

The other dividend is a new solution to a conundrum that bedevils research on emotion effects. Valence is often viewed as more psychologically meaningful than arousal (e.g., Rivers, Reyna, & Mills, 2008), and that has stimulated interest in identifying emotion effects that are due to valence per se (e.g., Kensinger, 2004; Murphy & Isaacowitz, 2008). The conundrum is how to ensure that the effects of valence manipulations (e.g., viewing pictures that vary in degree of positivity or negativity) are not due to the influence of arousal, given that the two are at least moderately correlated when ambiguity is left uncontrolled. This problem has cropped up repeatedly in literature that deals with how emotion affects specific processes—for instance, in the literature on emotion effects in true memory (see Kensinger, 2004), false memory (see Bookbinder & Brainerd, 2016), and cognitive aging (see Murphy & Isaacowitz, 2008). The conventional approaches to this problem are (a) to manipulate valence with arousal fixed at some intermediate level for all levels of a valence manipulation or (b) to manipulate both in Valence Level × Arousal Level factorial designs (see Brainerd & Bookbinder, 2018).

Valence–arousal conditionalization provides a more direct control method, which derives from the fact that valence ambiguity determines the strength of valence–arousal correlations in valence manipulations. Thus, valence standard deviation can be used to control arousal confounds in both of the preceding approaches. In the first, valence would be manipulated with valence standard deviation held fixed at high levels, where arousal would be uncorrelated with valence. In the second, valence intensity and valence standard deviation would be manipulated factorially. Valence main effects would be of central interest in the first approach, whereas Valence Intensity × Valence Standard Deviation interactions would be of central interest in the second approach.

Action Editor

Eddie Harmon-Jones served as action editor for this article.

Author Contribution

C. J. Brainerd is the sole author of this article and is responsible for its content.

Acknowledgments

I am grateful to A. M. Mattek, M. P. Toglia, and G. L. Wolford for their comments on a draft of this article.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The research reported in this article was supported by National Institutes of Health (National Institute on Aging) Grant 1RC1AG036915.

Open Practices

Data and materials for this study have not been made publicly available, and the design and analysis plans were not preregistered.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*, 68–72.

Bookbinder, S. H., & Brainerd, C. J. (2016). Emotion and false memory: The context-content paradox. *Psychological Bulletin*, *142*, 315–351.

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Stimuli, instruction manual and affective ratings* (Technical Report No. C-1). Gainesville: The Center for Research in Psychophysiology, University of Florida.

Brainerd, C. J., & Bookbinder, S. H. (2018). The semantics of emotion in false memory. *Emotion*. Advance online publication. doi:10.1037/emo0000431

Brainerd, C. J., Holliday, R. E., Reyna, V. F., Yang, Y., & Toglia, M. P. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology*, *107*, 137–154.

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychological Bulletin*, *133*, 761–779.

Citron, F. M., Weekes, B. S., & Ferstyle, E. C. (2014). How are affective word ratings related to lexicosemantic properties? Evidence from the Sussex Affective Word List. *Applied Psycholinguistics*, *35*, 313–331.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*, 659–676.

Flügel, J. C. (1925). A quantitative study of feeling and emotion in everyday life. *British Journal of Psychology*, *15*, 318–355.

Howe, M. L., Rabinowitz, F. M., & Grant, M. J. (1993). On measuring (in)dependence of cognitive processes. *Psychological Review*, *100*, 737–747.

Kensinger, E. A. (2004). Remembering emotional experiences: The contributions of valence and arousal. *Reviews in the Neurosciences*, *15*, 241–251.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*, 99–113.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*, 917–940.

Kutzner, F., & Fiedler, K. (2017). Stereotypes as pseudocontingencies. *European Journal of Social Psychology*, *28*, 1–49.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (Technical Report No. A-8). Gainesville: The Center for Research in Psychophysiology, University of Florida.

Marchewka, A., Łukasz, Z., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavioral Research*, *46*, 596–610.

Mattek, A. M., Wolford, G. L., & Whalen, P. J. (2017). A mathematical model captures the structure of subjective affect. *Perspectives on Psychological Science*, *12*, 508–526.

Murphy, N. A., & Isaacowitz, D. M. (2008). Preferences for emotional information in older and younger adults: A meta-analysis of memory and attention tasks. *Psychology and Aging*, *23*, 263–286.

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Payne, J. D., & Kensinger, E. (2011). Sleep leads to changes in the emotional memory trace: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *23*, 1285–1297.

Rivers, S. E., Reyna, V. F., & Mills, B. (2008). Risk taking under the influence: A fuzzy-trace theory of emotion in adolescence. *Developmental Review*, *28*, 107–144.

Roberts, R. P., Hach, S., Tippett, L. J., & Addis, D. R. (2016). The Simpson's paradox and fMRI: Similarities and differences between functional connectivity measures derived from within-subject and across-subject correlations. *NeuroImage*, *135*, 1–15.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society B: Statistical Methodology*, *13*, 238–241.

Stein, N. L., Ornstein, P. A., Tversky, B., & Brainerd, C. (1997). *Memory for everyday and emotional events*. Mahwah, NJ: Erlbaum.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403.

Vo, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*, 534–538.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191–1207.

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 [EmoPics: Subjective and psychophysiological evaluations of new visual material for clinical biopsychological research]. *Zeitschrif für Klinische Psychologie und Psychotherapie*, *39*(Suppl. 1/11), 1–30.

Wohlgemuth, A. (1923). The influence of feeling on memory. *British Journal of Psychology*, *13*, 405–416.

Wundt, W. M. (1912). *An introduction to psychology* (R. Pintner, Trans. & M. D'Hooghe, Ed.). London, England: Allen & Unwin.