Behavior Research Methods
2007, 39 (2), 326-334

Multidimensional normative ratings for the International Affective Picture System

Terry M. Libkuman, Hajime Otani, Rosalie Kern, Steven G. Viger, and Nicole Novak
Central Michigan University, Mount Pleasant, Michigan

The purpose of the present investigation was to replicate and extend the International Affective Picture System norms (Ito, Cacioppo, & Lang, 1998; Lang, Bradley, & Cuthbert, 1999). These norms were developed to provide researchers with photographic slides that varied in emotional evocation, especially arousal and valence. In addition to collecting rating data on the dimensions of arousal and valence, we collected data on the dimensions of consequentiality, meaningfulness, familiarity, distinctiveness, and memorability. Furthermore, we collected ratings on the primary emotions of happiness, surprise, sadness, anger, disgust, and fear. A total of 1,302 participants were tested in small groups. The participants in each group rated a subset of 18 slides on 14 dimensions. Ratings were obtained on 703 slides. The means and standard deviations for all of the ratings are provided. We found our valence ratings to be similar to the previous norms. In contrast, our participants were more likely to rate the slides as less arousing than in the previous norms. The mean ratings on the remaining 12 dimensions were all below the midpoint of the 9-point Likert scale. However, sufficient variability in ratings across the slides indicates that selecting slides on the basis of these variables is feasible. Overall, the present ratings should allow investigators to use these norms for research purposes, especially in research dealing with the interrelationships among emotion and cognition. The means and standard deviations for emotions may be downloaded as an Excel spreadsheet from www.psychonomic.org/archive.

The purpose of the present investigation was to replicate and extend the International Affective Picture System (IAPS) norms that were developed by Lang and his colleagues (Center for the Study of Emotion and Attention, 1995; Lang, Bradley, & Cuthbert, 1995, 1997a, 1999, 2001; Lang & Greenwald, 1981; Lang, Öhman, & Vaitl, 1988). The IAPS was developed in order to provide researchers in emotion and attention with a standardized set of photographic slides that varied in emotional evocation (Lang et al., 1995). Criteria for including pictures in the IAPS were that (1) the selected pictures would sample the entire affective space, (2) all of the pictures would be in color, (3) each picture would have a clear figure–ground relationship, and (4) the affective quality of the picture could be easily identified (Lang et al., 1999). Currently, the IAPS contains over 700 color photographs. For each photograph, norms (means and standard deviations) are provided for the dimensions of arousal, pleasure (valence), and dominance. These data were collected using the Self-Assessment Manikin (SAM; Lang, 1980). SAM is an affective rating-scale system using a graphical figure that depicts the dimensions of arousal (from an excited to a relaxed figure), pleasure (from a smiling to a frowning figure), or dominance (from a large to a small figure). When collecting data for each dimension using the SAM, participants are instructed to rate how they feel about each picture by placing an X over or between the appropriate figure(s). There are five figures with four spaces between the figures, allowing for a 9-point rating scale. Lang, Bradley, and Cuthbert (1997b) consider the SAM instrument to be "largely culture-free" (p. 102).

The IAPS norms have been used in numerous investigations. A recent search using PsycINFO revealed that since 1995, more than 75 articles have used the IAPS norms. In a significant portion of these studies, the IAPS norms have been used in research investigating basic processes (e.g., learning, emotion, motivation, and attention) using psychophysiological methodology. However, more recently, studies using the IAPS norms have included different topics (aging and Alzheimer's disease—Kensinger, Brierley, Medford, Growdon, & Corkin, 2002; psychosis—Subramanian, 2003; and trait anxiety—Yiend & Mathews, 2001), as well as different methodologies (e.g., Bradley, Greenwald, Petry, & Lang, 1992; Cahill & Alkire, 2003; Kern, Libkuman, & Otani, 2002; Kern, Libkuman, Otani, & Holmes, 2005; Libkuman, Stabler, & Otani, 2004).

One important reason for collecting additional IAPS normative data is to determine whether the IAPS norms can be replicated, especially in different laboratories and at different times. Replication may be particularly important when dealing with ratings involving emotional content. For example, individuals may be sensitized or habituated to the emotional content contained in the IAPS slides simply because they have been repeatedly exposed to this type of material in the media (e.g., movies, television, and the

T. M. Libkuman, libku1tm@cmich.edu

Internet). Consistent with this view, Bradley, Lang, and Cuthbert (1993) reported that when the same emotional IAPS slides were repeated, the acoustic startle reflex habituated (although the blink response did not). Also, from an anecdotal perspective, it was not uncommon for participants in our laboratory to spontaneously report that the IAPS slides were jejune simply because they had encountered these types of stimuli in their previous experiences.

Fortunately, other data have indicated that the IAPS norms possess considerable replicability. Ito, Cacioppo, and Lang (1998) collected ratings on a subset (n = 472) of the IAPS slide set using students from Ohio State University. (The original IAPS norms were collected at the University of Florida.) The mean arousal, valence, and dominance ratings between the two schools were highly correlated, with all correlations exceeding .80, indicating reliability across time with two different student populations. Unfortunately, Ito et al. did not include the entire set of IAPS slides.

Another important reason for collecting additional IAPS normative data is that the current IAPS norms are limited to the dimensions of arousal, valence, and dominance. It is clear that the photographs contain other characteristics, and these characteristics may influence the rater's evaluative responding. In the present study, in addition to arousal and valence, we collected normative data on the dimensions of surprise, consequentiality, meaningfulness, familiarity, distinctiveness, and memorability.[1] Some of these variables have a cognitive ring to them, whereas others may be more relevant to the arousal–emotion domain. Regardless, characteristics other than the currently measured emotional content of the slides may play a role in how the slides are evaluated. For example, a major weakness in studies investigating emotion–memory relationships is the arousal–material confound (Libkuman, Nichols-Whitehead, Griffith, & Thomas, 1999); that is, it is difficult to ascertain the influence of the intended independent variable (i.e., the arousing aspect of the stimulus), because the presented independent variable includes other features in addition to the emotional aspect itself. In fact, Cahill and McGaugh (1995) have suggested that these other characteristics of the "emotional" stimulus are an integral or an inseparable part of the emotional event; that is, both are necessary for emotional arousal to occur.

Therefore, obtaining ratings on characteristics other than the emotional aspect of the slides will allow investigators to examine the relationships between valence and arousal and the additional dimensions that were included in the present study. It is also important to note that these additional norms may be useful for other purposes as well. Ratings on these dimensions may be useful in their own right, because they will allow researchers to investigate, for example, the relationship between familiarity, distinctiveness, and memory (i.e., a cognitively oriented study). Finally, knowing the norms for the additional dimensions may be useful for control purposes, allowing one, for example, to systematically vary arousal while controlling for familiarity. This is a frequent problem in the cognitive literature (Kausler, 1974) when a particular linguistic attribute (e.g., word frequency) is chosen for study and this attribute covaries with another linguistic attribute (e.g., word meaningfulness). A common methodological strategy that is used to control for the potential confound is to choose words from established norms that contain information about both of the aforementioned attributes.

We chose to collect rating data on the dimensions of surprise, consequentiality, meaningfulness, familiarity, distinctiveness, and memorability because we assumed these dimensions would be the most likely candidates for theoretical and applied research purposes, especially in research that has an emotion–cognition flavor.

Surprise (i.e., the unexpectedness of an event) and consequentiality (i.e., the personal or national relevance of an event) are well represented in the emotion–memory literature. Brown and Kulik (1977) postulated that surprise and consequentiality are critical for the formation of flashbulb memories. In a typical flashbulb study, Conway et al. (1994) measured surprise and consequentiality using a 3-point rating scale and found that moderate to high levels of surprise and consequentiality were correlated with the development of flashbulb memories. Furthermore, surprise and consequentiality played an integral role in Conway et al.'s structural equation model of flashbulb memory. Although these constructs continue to play a role in theorizing about emotion–memory relations (e.g., MacKay & Ahmetzanov, 2005; Otani et al., 2005), no studies have been conducted in which surprise and consequentiality have been independently and systematically manipulated.

Meaningfulness is a construct that has a considerable history in the human learning and cognition areas. It has been considered a critical factor in remembering ever since verbal learning psychologists (e.g., Noble, 1952) experimented with nonsense syllables using serial and paired-associate learning methodologies. In the modern era, meaningfulness became the centerpiece of memory explanations, because the levels-of-processing approach (Craik & Lockhart, 1972) emphasized the elaboration of meaning (i.e., deep level of processing) as the most critical factor in creating durable memory. Although the notion of levels of processing has been discredited, researchers have continued to use the levels-of-processing manipulation because processing verbal materials at a deep level invariably produces better recall and recognition than processing at a shallow level. In the emotion–memory literature, the importance of meaningfulness was discussed by Christianson (1992). On the basis of an extensive review of the literature, Christianson proposed two possible mechanisms by which high emotional arousal leads to enhanced memory: preattentive processing and poststimulus elaboration. He assumed the former process to be automatic and the latter process to be effortful. That is, upon encountering an emotionally arousing stimulus, one may encode its content preattentively, without using processing resources. In contrast, emotional arousal may lead one to elaborate the content of a stimulus after the exposure to the stimulus is over. Although Christianson did not clearly specify what he meant by elaboration, it is clear that meaningfulness of a stimulus would facilitate one's effort to elaborate the stimulus.

Familiarity has been defined as the frequency of experience associated with a given word, or how often the

person came into contact with the word (Kausler, 1974; Noble, 1952). Kausler noted that although familiarity is correlated with meaningfulness, the relationship is not perfect. For example, some words that are high in familiarity are not high in meaningfulness (e.g., *and*) and some words low in familiarity are high in meaningfulness (e.g., *mistletoe*). It is well established that normative frequency of words produces different effects on recall and recognition—referred to as the word-frequency effect. It has been shown that high-frequency words are easier to recall than low-frequency words, whereas low-frequency words are easier to recognize than high-frequency words (see, e.g., Shepard, 1967). It is possible that familiarity of pictures would produce a similar effect. Further, familiarity may play a possible role in processes—for example, sensitization and habituation—that mediate the relationship between cognition and emotion.

Word distinctiveness (orthography) refers to the extent to which words possess unusual or unique characteristics. For example, the word *xylem* possesses considerable orthographic distinctiveness (see Zechmeister, 1972, for ratings on 150 five-letter words). Zechmeister reported that in a recognition paradigm, distinctive words were remembered better than nondistinctive words. Schmidt (1985) examined the role of distinctiveness by placing the name of a country among a list of animal names, or vice versa. In contrast to control lists of items that were of the same category, he found that recognition and recall memory were better for the item that was different from the other items. Basically, these findings replicate an old phenomenon in the cognitive literature known as the von Restorff effect. The von Restorff effect is produced by embedding a letter within a series of digits, or vice versa. In this case, distinctiveness refers to the relationship between the target and its surround. Schmidt proposed several categories of distinctiveness in his theory of distinctiveness, including emotional distinctiveness. He assumed that emotional distinctiveness produced sympathetic activation leading to increases in memory for gist and declines in memory for detail. The role of distinctiveness in memory was also emphasized by other researchers. In particular, Hunt and colleagues (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993) proposed that memory is a joint function of organization and distinctiveness. According to these researchers, organization emphasizes similarities among to-be-remembered items, whereas distinctiveness emphasizes differences among the items. These researchers assumed that both organization and distinctiveness are necessary in creating good memory, because differences cannot be noticed unless there are similarities among to-be-remembered items. In other words, distinctiveness is derived from the difference between a given item and its surround, just like in the case of the von Restorff effect. With regard to emotional stimuli, these researchers would assume that emotionally arousing stimuli are remembered better than emotionally neutral stimuli, because emotionally arousing stimuli in an experiment are occurring against the background of emotionally neutral daily events. Accordingly, these researchers would argue that it would be possible to make emotionally neutral stimuli more memorable than emotionally arousing stimuli by presenting emotionally neutral stimuli against the background of emotionally arousing events.

Memorability was included because of the common observation that individuals exposed to highly arousing or personally significant events report that the events would not be easily forgotten. Memorability can be conceptualized as a measure of metamemory—that is, one's judgment about what one would remember. Three types of metamemory judgments have been studied in the literature (see Leonesio & Nelson, 1990; Schwartz, 1994): ease of learning (EOL), judgments of learning (JOLs), and feelings of knowing (FOKs). EOL judgments are judgments regarding how easily materials could be learned, JOLs are judgments of how well materials have been learned, and FOKs are judgments regarding whether unrecalled items are still in memory. Obviously, memorability judgments are based on EOL judgments, which have been shown to predict later recall performance (Leonesio & Nelson, 1990; Underwood, 1966). Research on metamemory has been focused primarily on verbal materials. Therefore, establishing norms for pictures, especially with emotional content, would contribute to the metamemory literature.

Research using the IAPS norms has been dominated by testing the dimensional view of emotion (e.g., Bradley, 1994; Bradley, Greenwald, & Hamm, 1993; Davis et al., 1995). However, emotion can also be characterized as a series of discrete or specific emotional states (Bradley, Greenwald & Hamm, 1993). The list of primary emotions typically includes the states of happiness, surprise, sadness, anger, disgust, and fear (Ekman & Friesen, 1974; Izard, 1972). Although considerable variance among these emotional states can be accounted for by the dimensions of valence and arousal, Bradley, Greenwald, and Hamm (1993) concluded that the IAPS slides would be a useful way to examine these theories. Therefore, we also included ratings on each slide for the primary emotions of happiness, surprise, sadness, anger, disgust, and fear.[2]

METHOD

Participants

A total of 1,302 Midwestern university students participated. The participants were 18 years old or older and included both males and females. Participants received extra course credit for taking part. The participants were tested in groups ranging in size from 2 to 20, over a period of about 3 years.

Apparatus and Materials

The International Affective Picture System (IAPS) contained 716 standardized color photographs (Lang et al., 1999).[3] The goal was to collect data on all 716 slides, but as a result of oversight and slide duplication, we collected ratings on 703 slides. The slides were randomly assigned to 1 of 40 slide sets. Each set consisted of 18 slides. There were two random orders for each set. The slides were individually projected onto a wall-mounted 5.5 ft × 5.5 ft screen. The size of the projected image was somewhat smaller than the size of the screen. Over the time course of the study, the data were collected in several different research rooms, all of which were suitable for viewing the slides in question. In each case, the room provided adequate lighting and curtains (used to occlude ambient light) and a minimum of nine chairs that could be arranged in rows so that each

participant could comfortably see the entire screen. Seating distance from the screen varied from 11 ft to 19 ft. The overhead lights were turned off during testing. However, sufficient light was still available for the participants to comfortably view and respond to the materials located on their chairs.

In order to systematize the data collection, we constructed a 75-page rating booklet. On the first page, raters were instructed not to write in the booklet, and to use the accompanying optical scanning sheets to mark their ratings. The remaining pages contained the 9-item (range 1–9) Likert-like rating scales (see Appendix). Each rating scale possessed the same format; the dimensions to be rated and how they were to be rated were described. Each participant was instructed to rate each of the 18 slides on the 14 dimensions (i.e., valence, arousal, distinctiveness, surprise, consequentiality, memorability, meaningfulness, familiarity, and the six emotions of happiness, surprise, sadness, anger, disgust, and fear; see Appendix). The dimensions were always presented in the same order. The first 6 pages were used for practice ratings. The remaining pages were used for actual ratings. It was decided to use a Likert type of rating format instead of the SAM (Lang, 1980), because we wanted to use the same format for all of the ratings, and SAM would not have been suitable for rating many of our dimensions. Additionally, and again for the purposes of uniformity, we arranged each dimension so that the anchor for 1 represented the lower end of the dimension and the anchor for 9 represented the upper end of the dimension, with the exception of the happiness dimension. In the latter case, 1 represented *completely unhappy* and 9 represented *completely happy*. In the case of the SAM instrument (Lang, 1980), the labels are reversed for happiness (from *happy* to *unhappy*) and arousal (from *arousal* to *unarousal*). In order to determine whether ratings differed as a function of rating method, we conducted a study in which the type of scale (SAM vs. Likert) was factorially combined with the labeling of the anchors (e.g., 1 = *unhappy* and 9 = *happy* vs. 1 = *happy* and 9 = *unhappy*). Eighty participants were randomly assigned to the four treatment combinations, with the restriction that 20 participants must end up in each combination. Using essentially the procedure outlined in the Procedure section of this article, we instructed the participants to evaluate a set of 18 slides on the basis of valence and arousal. The slide set was randomly chosen from one of our previously determined slide sets. Practice was provided on 2 slides before the 18 slides were presented. Separate 2 × 2 ANOVAs on the valence and arousal ratings failed to produce any significant main effects or interactions (all $ps > .18$).

Procedure

Upon their arrival, the participants were asked to be seated, and to read and sign the consent form if they were willing to participate in the study. Each participant then received a copy of the instructions and a rating booklet. The instructions were read out loud by the investigator while the participants read silently. The participants were informed that the purpose of the study was to determine how people respond to pictures that represent different events that occur in life. They were further informed that the study would take about 45 min, that during this time they would be looking at a series of different pictures individually projected onto a screen in front of them, and that they would be asked to rate each picture on a number of different dimensions. Finally, they were informed that (1) accuracy was more important than speed, (2) there were no right or wrong answers, (3) they should respond as honestly as they could, and (4) they should be quiet during the session, because making any exclamations or comments might influence other people's ratings and ruin the study. Any questions raised by the participants were answered by the investigator. The lights were then turned off, and three practice slides were presented. The practice slides were used to acquaint the participants with the rating procedure. The investigator read a description of each rating scale. For example, for the happiness (valence) scale, the participants were instructed that if they felt completely happy while viewing the picture, they would indicate their answer by circling the number 9 on the Likert scale. On the other hand, if they felt completely unhappy, they were instructed to indicate their response by circling the number 1. Furthermore, the participants were informed that the remaining numbers represented intermediate values of happiness, with 5 indicating neutral. The investigator asked the participants if they understood the procedure and if they had any questions concerning their responsibilities. This same procedure was used for the remaining dimensions, with the obvious exception that the descriptions conformed to the different characteristics. In the case of the emotion ratings, the participants were instructed to rate each emotion based on the intensity of the felt emotion that the picture elicited.

The slides were advanced via a magnetically coded audiocassette allowing each slide to be shown for 2 min, with a 3-sec interslide interval. The selection of the 2-min period was based on the pilot testing of 60 participants who were not a part of the actual study. The 60 participants were subdivided into groups of 20 participants each; each group was provided with a different set of 20 slides from the IAPS and was instructed to follow our rating procedure. Collapsed across the three subgroups, the average time to complete the ratings for each of the 20 slides was 62.27 sec ($SD = 18.00$), with a range of 34 to 110 sec. Given the extra slide in the actual procedure (3 practice and 18 countable ratings) and the maximum of 110 sec, we thought that 2 min was sufficient time, without being rushed, for the participants to complete the ratings. During the 2-min period, the participants were instructed to rate each of the 14 slide characteristics for a given slide using the rating booklet, and to mark their responses on the optical scanning sheets. During the interslide interval, the participants were shown an instructional slide with the text "Please get ready to rate the next slide." The same timing pattern was used for all 21 slides. When they had finished, the participants were asked if they had any further questions. They were then debriefed as to the purpose of the study and thanked for their participation. The entire procedure took approximately 60 min to complete. The optical scanning sheets were scored by a computer-controlled optical scanning device. Raw data were then converted to text files and placed in an SPSS database for analyses.

RESULTS

The ratings for the 703 slides are archived at the following Web sites: www.chsbs.cmich.edu/hajime_otani/research/emotion.htm; www.psychonomic.org/archive/. The archived ratings include the means and standard deviations for (1) arousal and valence for the present study, as well as the ratings from the Lang et al. (1999) and Ito et al. (1998) norms; (2) distinctiveness, consequentiality, memorability, surprise, meaningfulness, and familiarity; and (3) the six emotions (happiness, surprise, sadness, anger, disgust, and fear).

Table 1
Valence and Arousal Summary Statistics for the Lang, Ito, and Libkuman Ratings

Variable	Lang	Ito	Libkuman
Number of Slides	703	471	703
Valence			
M	5.09	5.03	4.23
SD	1.84	1.94	1.80
Median	5.22	5.19	4.34
Skewness	−0.26	−0.14	−0.02
Arousal			
M	4.84	4.91	3.69
SD	1.22	1.35	1.07
Median	4.92	4.81	3.69
Skewness	−0.17	0.10	0.07

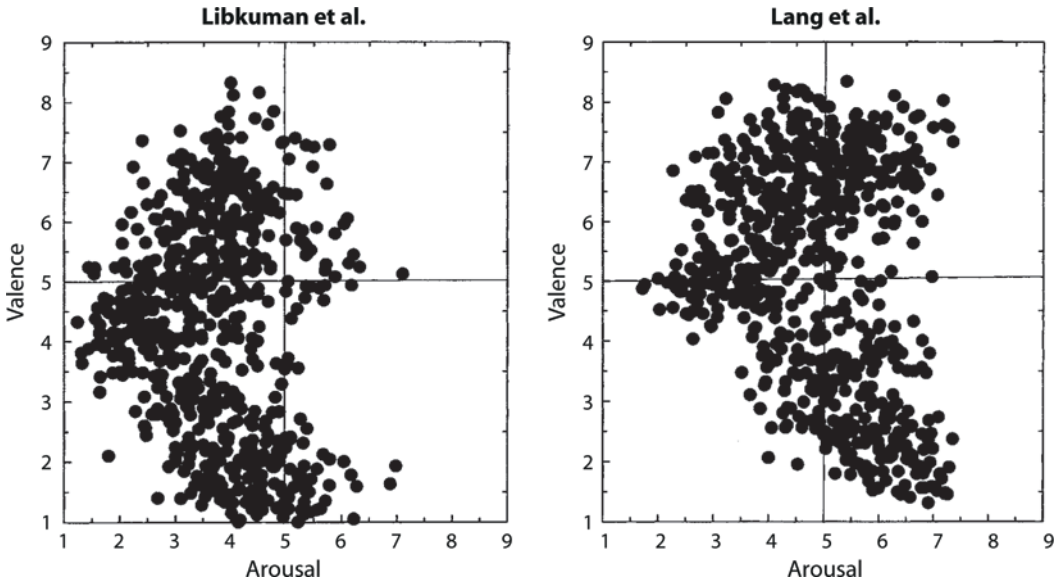

Figure 1. Scatterplots of affective space for the Libkuman et al. (2007) and Lang et al. (1999) data.

Distributions for the present arousal and valence norms and the Lang et al. (1999) and Ito et al. (1998) arousal and valence norms all approach normalcy. It is clear from Table 1 that although the overall means for valence and arousal are similar for the Lang et al. (1999) and Ito et al. ratings, these ratings differ from the Libkuman ratings. One-way ANOVAs and Tukey HSD tests confirmed this observation. In the case of valence, the main effect was significant [$F(2,1874) = 44.98$, $p = .00$, $\eta_p^2 = .05$]. The Tukey test revealed that the comparison between Lang and Ito was not significant ($p = .85$), whereas the comparisons between Lang and Libkuman ($p = .00$) and Ito and Libkuman ($p = .00$) were significant. In the case of arousal, the outcomes were the same. The main effect was significant [$F(2,1874) = 212.92$, $p = .00$, $\eta_p^2 = .18$]. The Tukey test revealed that the comparison between Lang and Ito was not significant ($p = .635$), whereas the comparisons between Lang and Libkuman ($p = .00$) and Ito and Libkuman ($p = .00$) were significant. With the exception of these two significant findings, it is important to note the similarity of the summary statistics across the three data sets. Basically, the important differences were that our participants were more likely to rate the slides as less arousing. A scatterplot of affective space for the Lang and Libkuman data (see Figure 1) shows these effects. It is clear that the pattern is similar for both data sets. However, if the quadrants are determined based on the midpoint of each scale (5), as Lang et al. (1999) did in their manual, the number of data points for each quadrant differ between the two data sets. Lang et al. (1999) stated that it was difficult obtaining slides that fell into the low arousal–unpleasant quadrant. In contrast, in our data, the low-arousal quadrants are overrepresented and the high-arousal quadrants are underrepresented. Finally, it is important to note that the arousal–valence correlation is about twice the size for the Lang data ($r = -.25$) as it is for our data ($r = -.13$).

Table 2 provides the summary statistics for distinctiveness, consequentiality, mem orability, surprise, meaningfulness, and familiarity. The distributions for all six dimensions approach normalcy, with all of the means and medians below the midpoint of the 9-point Likert scale. One-sample *t* tests on each dimension, with the midpoint of the scale as the constant (5), were all significant (all *p*s = .00), indicating that the ratings of each dimension were below the midpoint of the rating scale. Did the dimensions produce different mean ratings? A repeated measures ANOVA on the six dimensions revealed a significant main effect [$F(5,3510) = 66.72$, $p = .00$]. Paired *t* tests were conducted on all of the possible comparisons. We used the Bonferroni correction to adjust alpha to .008.

Table 2
Distinctiveness, Consequentiality, Memorability, Surprise, Meaningfulness, and Familiarity Summary Statistics

Dimension	*M*	*SD*	Median	Skewness
Distinctiveness	4.59	1.25	4.52	0.33
Consequentiality	3.89	1.10	3.88	0.19
Memorability	4.12	1.33	4.02	0.36
Surprise	3.81	1.52	3.42	0.76
Meaningfulness	4.14	1.25	4.10	0.10
Familiarity	4.67	1.53	4.69	0.00

Table 3
Summary Statistics for the Six Emotions

	Happy	Surprise	Sadness	Anger	Disgust	Fear
M	3.49	3.38	2.76	2.36	2.64	2.46
SD	1.91	1.30	1.92	1.63	1.86	1.60
Median	3.13	3.08	1.76	1.53	1.72	1.63
Skewness	0.45	0.84	1.19	1.38	1.18	1.13

Table 4
Intercorrelations Among the 18 Dimensions

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Valence—Lang																		
2	Arousal—Lang	−.25																	
3	Valence—Ito	.95**	−.33																
4	Arousal—Ito	−.41	.78**	−.41															
5	Valence—Libkuman	.88**	−.31	.87**	−.40														
6	Arousal—Libkuman	−.05	.61**	−.09	.51**	−.13													
7	Distinctiveness	−.43	.58**	−.46	.59**	−.43	.51**												
8	Consequentiality	−.14	.44**	−.15	.37**	−.08	.56**	.43											
9	Memorability	−.36	.58**	−.37	.57**	−.33	.63**	.80**	.77**										
10	Surprise	−.56	.62**	−.61	.63**	−.60	.55**	.85**	.42**	.78**									
11	Meaningfulness	−.11	.42**	−.13	.37**	−.02	.49**	.50**	.91**	.81**	.43**								
12	Familiarity	.56**	−.39	.58**	−.50	.62**	−.23	−.72	−.01	−.42	−.69	−.05							
13	Happiness	.86**	−.17	.86**	−.30	.91**	.08*	−.32	.11**	−.13	−.45	.14**	.55**						
14	Surprise	−.37	.60**	−.40	.56**	−.46	.64**	.78**	.37**	.74**	.90**	.37**	−.60	−.25					
15	Sadness	−.79	.37**	−.75	.44**	−.77	.29**	.59**	.50**	.67**	.66**	.51**	−.52	−.65	.51**				
16	Anger	−.80	.43**	−.77	.50**	−.77	.28**	.57**	.47**	.61**	.67**	.45**	−.53	−.68	.50**	.90**			
17	Disgust	−.80	.48**	−.81	.54**	−.80	.29**	.63**	.29**	.58**	.77**	.29**	−.62	−.71	.64**	.78**	.87**		
18	Fear	−.71	.53**	−.70	.62**	−.72	.31**	.53**	.39**	.54**	.61**	.37**	−.53	−.65	.46**	.72**	.80**	.72**	

$N = 471$ for the Ito correlations; $N = 703$ for the remaining correlations. $^{*}p < .05$. $^{**}p < .01$.

All of the comparisons were significant, with the exception of the consequentiality–familiarity ($p = .770$) and meaningfulness–familiarity ($p = .161$) comparisons.

Table 3 provides the summary statistics for the six emotions. All of the distributions are somewhat positively skewed. A repeated measures ANOVA on the six emotions revealed a significant main effect [$F(5,3510) = 69.72$, $p = .00$]. Paired t tests were conducted on all of the possible comparisons. Using the Bonferroni correction, we adjusted alpha to .008. All of the comparisons were significant. It is important to note that all of the means and medians are substantially below the midpoint of the 9-point Likert scale. One-sample t tests on each emotion using 5 (the midpoint of the scale) as the constant were all significant (all ps = .00), indicating that ratings of each emotion were below the midpoint of the rating scale. Basically, the participants did not perceive the IAPS slides as eliciting these basic emotions even moderately, and this was especially true for sadness, anger, disgust, and fear. The means for the latter four emotions were below the scaled value of 3, and the medians were below the scaled value of 2. Finally, it is important to note that happiness (valence) and surprise were rated twice. The means and medians for happiness (valence) in Tables 1 and 3 and surprise in Tables 2 and 3 are higher when participants initially rated these dimensions than when they later rated each emotion. Paired t tests confirmed this observation. In the case of happiness–valence, the paired t test was significant [$t(702) = 23.99$, $p = .00$], and in the case of surprise–surprise, the paired t test was significant [$t(702) = 17.35$, $p = .00$].

Table 4 provides the intercorrelations among the 18 variables. An examination of the table indicates that with a few (four) exceptions, all of the variables were significantly correlated. The mean intercorrelation among the 18 variables was high ($r = .52$, $SD = .22$). A principal axis factor analysis using varimax rotation was used to reduce the number of overlapping measured variables to a smaller number of factors. This analysis was restricted to the 14 measures in our study. The factor analysis reduced the 14 variables to four factors (see Table 5). We labeled the first factor as valence, the second factor as metacognition, the third factor as arousal, and the fourth factor as stimulus salience. In general, the assignment of the loadings to a given factor provided a reasonable fit. However, two points need to be stressed. First, the eigenvalue for the fourth factor fell below 1. Although not much of the variance is accounted for by this factor, we decided to keep it, because the factor loadings for distinctiveness were not high for the other factors. This decision leads to the second point. Although familiarity loaded more on the fourth factor than on the first factor, the difference was minimal. We decided to assign familiarity to the fourth factor simply because it is more conceptually similar to distinctiveness than to the variables associated with valence. Overall, whether a three- or four-factor solution provides the best fit is a moot point, because the analysis does suggest that some of the variance is accounted for by variables other than arousal and valence. Further investigations, especially experimental studies, will be needed to provide a more definite answer to the utility of these "nonemotional" variables.

Table 5
Factor Analysis of the Ratings

	Factor Loadings			
Dimensions	1	2	3	4
Valence	**−0.93**	0.07	−0.02	−0.11
Happiness emotion	**−0.92**	0.22	0.01	0.14
Sadness emotion	**0.79**	0.44	0.15	0.18
Anger emotion	**0.84**	0.38	0.17	0.14
Disgust emotion	**0.79**	0.17	0.35	0.22
Fear emotion	**0.73**	0.29	0.20	0.16
Consequentiality	0.10	**0.92**	0.22	−0.05
Memorability	0.26	**0.71**	0.50	0.33
Meaningfulness	0.06	**0.95**	0.15	0.12
Arousal	0.02	0.43	**0.61**	0.06
Surprise	0.48	0.26	**0.68**	0.39
Surprise emotion	0.29	0.19	**0.88**	0.28
Distinctiveness	0.30	0.49	0.35	**0.72**
Familiarity	−0.50	−0.36	0.10	**−0.58**
Eigenvalue	8.02	2.74	1.32	0.55
Cumulative percentage of variance	57.30	76.86	86.33	90.24

DISCUSSION

The purpose of the present study was to replicate and extend the IAPS norms. We found that the arousal and valence norms of the present study were roughly similar to the Lang et al. (1999) and Ito et al. (1998) norms. The major difference occurred with the arousal ratings; our ratings were consistently lower than those in the other two sets of norms. Why are our arousal ratings lower than the Lang et al. (1999) and Ito et al. arousal ratings? We can think of three reasons. First, we used Likert scales, whereas Lang et al. (1999) and Ito et al. used the SAM. Furthermore, we arranged the scales differently than Lang et al. (1999) and Ito et al. However, we found that neither the scale type nor the order had any significant impact on the ratings. Second, our participants rated more stimuli and rated them on dimensions other than affect. It is possible that simply rating more stimuli may lead to lower arousal ratings. Indirectly, we found some support for this view, because when stimuli were rated twice on different occasions (e.g., valence and then the emotion of happiness; surprise and then the emotion of surprise) we found in both instances that the means were lower for the second rating. It is also possible that a psychological set occurred in our participants, reflecting a "more objective, nonarousing" perspective when evaluating the stimuli. Third, the emotional content of the slides may not be as arousing because of repeated exposure to these types of stimuli in our culture by the media. However, the data-collection times for the three studies are not that far apart. In other words, this explanation makes sense only if we assume that increased exposure to emotional events is a recent phenomenon.

We extended the IAPS norms to include ratings on the dimensions of surprise, consequentiality, meaningfulness, similarity, distinctiveness, and memorability, as well as the six emotions of happiness, surprise, sadness, anger, disgust, and fear. Although all of the ratings for the first six dimensions were below the midpoint of the scale, there is still substantial variability in ratings across the slides, so that using these ratings as controls and/or studying the dimensions per se is feasible. In the case of the ratings for the six emotions, it was again found that all of the ratings fell below the midpoint, with the ratings of sadness, disgust, anger, and fear falling below the scaled value of 3. Further research will be needed to determine the utility of these emotion ratings. For example, is anything gained by differentiating on the basis of specific emotions in contrast to selecting slides on the basis of arousal and valence ratings? Further research will be needed to answer this question. Finally, the outcome of the factor analysis suggests that variables other than arousal and valence need to be considered when using these norms.

AUTHOR NOTE

This research was supported by a grant awarded to T.M.L. from the Faculty Research and Creative Endeavors Committee, Central Michigan University. Correspondence concerning this article should be addressed to T. M. Libkuman, Department of Psychology, Central Michigan University, Mount Pleasant, MI 48859 (e-mail: libku1tm@cmich.edu).

REFERENCES

BRADLEY, M. M. (1994). Emotional memory: A dimensional analysis. In S. VanGooze, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 61-93). Hillsdale, NJ: Erlbaum.

BRADLEY, M. M., GREENWALD, M. K., & HAMM, A. O. (1993). Affective picture processing. In N. Birbaumer & A. Öhman (Eds.), *The structure of emotion: Psychophysiological, cognitive and clinical aspects* (pp. 46-68). Toronto: Hogrefe & Huber.

BRADLEY, M. M., GREENWALD, M. K., PETRY, M., & LANG, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 379-390.

BRADLEY, M. M., LANG, P. J., & CUTHBERT, B. N. (1993). Emotion, novelty, and the startle reflex: Habituation in humans. *Behavioral Neuroscience*, **107**, 970-980.

BROWN, R., & KULIK, J. (1977). Flashbulb memories. *Cognition*, **5**, 73-99.

CAHILL, L., & ALKIRE, M. T. (2003). Epinephrine enhancement of human memory consolidation: Interaction with arousal at encoding. *Neurobiology of Learning & Memory*, **79**, 194-198.

CAHILL, L., & MCGAUGH, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness & Cognition*, **4**, 410-421.

CENTER FOR THE STUDY OF EMOTION AND ATTENTION (CSEA-NIMH). (1995). The International Affective Picture System [Photographic slides]. Gainesville: University of Florida, Center for Research in Psychophysiology.

CHRISTIANSON, S. A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, **112**, 284-309.

CONWAY, M. A., ANDERSON, S. J., LARSEN, S. F., DONNELLY, C. M., MCDANIEL, M. A., MCCLELLAND, A. G. R., ET AL. (1994). The formation of flashbulb memories. *Memory & Cognition*, **22**, 326-343.

CRAIK, F. I., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **11**, 671-684.

DAVIS, W. J., RAHMAN, M. A., SMITH, L. J., BURNS, A., SENECAL, L., MCARTHUR, D., ET AL. (1995). Properties of human affect induced by static color slides (IAPS): Dimensional, categorical, and electromyographical analysis. *Biological Psychiatry*, **41**, 229-253.

EINSTEIN, G. O., & HUNT, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 588-598.

EKMAN, P., & FRIESEN, W. V. (1974). Detection deception from body or face. *Journal of Personality & Social Psychology*, **29**, 288-298.

HUNT, R. R., & EINSTEIN, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, **20**, 497-514.

HUNT, R. R., & MCDANIEL, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language*, **32**, 421-445.

ITO, T. A., CACIOPPO, J. T., & LANG, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality & Social Psychology Bulletin*, **24**, 855-879.
IZARD, C. E. (1972). *Patterns of emotions: A new analysis of anxiety and depression*. New York: Academic Press.
KAUSLER, D. H. (1974). *Psychology of verbal learning and memory*. New York: Academic Press.
KENSINGER, E. A., BRIERLEY, B., MEDFORD, N., GROWDON, J. H., & CORKIN, S. (2002). Effects of normal aging and Alzheimer's disease on emotional memory. *Emotion*, **2**, 118-134.
KERN, R. P., LIBKUMAN, T. M., & OTANI, H. (2002). Memory for negatively arousing and neutral pictorial stimuli using a repeated testing paradigm. *Cognition & Emotion*, **16**, 749-767.
KERN, R. P., LIBKUMAN, T. M., OTANI, H., & HOLMES, K. (2005). Emotional stimuli, divided attention, and memory. *Emotion*, **5**, 408-417.
LANG, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery* (pp. 119-137). Norwood, NJ: Ablex.
LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1995). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville: University of Florida, Center for Research in Psychophysiology.
LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1997a). *International Affective Picture System (IAPS): Technical manual and affective ratings*. NIMH Center for the Study of Emotion and Attention.
LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1997b). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97-135). Mahwah, NJ: Erlbaum.
LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1999). *International Affective Picture System (IAPS): Technical manual and affective ratings*. NIMH Center for the Study of Emotion and Attention.
LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (2001). *International Affective Picture System (IAPS)*: *Technical manual and affective ratings*. NIMH Center for the Study of Emotion and Attention.
LANG, P. J., & GREENWALD, M. K. (1981). *The International Affective Picture System standardization procedure and initial group results for affective judgments* (Tech. Rep. No. 1A). Gainesville: University of Florida, Center for Research in Psychophysiology.
LANG, P. J., ÖHMAN, A., & VAITL, D. (1988). *The International Affective Picture System* [Photographic slides]. Gainesville: University of Florida, Center for Research in Psychophysiology.
LEONESIO, R. J., & NELSON, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 464-470.
LIBKUMAN, T. M., NICHOLS-WHITEHEAD, P., GRIFFITH, J., & THOMAS, R. (1999). Source of arousal and memory for detail. *Memory & Cognition*, **27**, 166-190.
LIBKUMAN, T. M., STABLER, C. L., & OTANI, H. (2004). Arousal, valence, and memory for detail. *Memory*, **12**, 237-247.
MACKAY, D. G., & AHMETZANOV, M. V. (2005). Emotion, memory, and attention in the taboo Stroop paradigm. An experimental analogue of flashbulb memories. *Psychological Science*, **16**, 25-32.
NOBLE, C. E. (1952). An analysis of meaning. *Psychological Review*, **59**, 421-430.
OTANI, H., KUSUMI, T., KATO, K., MATSUDA, K., KERN, R. P., WIDNER, R., JR., & OHTA, N. (2005). Remembering a nuclear accident in Japan: Did it trigger flashbulb memories? *Memory*, **13**, 6-20.
SCHMIDT, S. R. (1985). Can we have a distinctive theory of memory? *Memory & Cognition*, **19**, 523-542.
SCHWARTZ, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, **1**, 357-375.
SHEPARD, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, **6**, 156-163.
SUBRAMANIAN, L. (2003). *Emotional functioning of psychosis-prone individuals.* Unpublished master's thesis, Central Michigan University.
UNDERWOOD, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, **71**, 673-679.
YIEND, J., & MATHEWS, A. (2001). Anxiety and attention to threatening pictures. *Quarterly Journal of Experimental Psychology*, **54A**, 665-681.
ZECHMEISTER, E. B. (1972). Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology*, **85**, 425-430.

NOTES

1. We did not collect any normative data for dominance, nor did we do any analyses using the dominance data collected by previous authors. At the time we collected the data, we thought that this variable, in contrast to arousal and valence, had not been shown to produce any consistent effects. In hindsight, it might have been wise to include dominance in our data-collection procedure, if for nothing else than for comparison purposes. Finally, it might have been useful to include ratings on other properties of the slides—for example, complexity—but again, we only thought of this variable in retrospect.

2. Although we had already included ratings of happiness (valence) and surprise, we thought it would still be useful to collect these ratings again within the context of rating basic emotions.

3. We did not select any slides or use any data from vthe updated version published in 2001 by Lang, Bradley, and Cuthbert.

ARCHIVED MATERIALS

The following material may be accessed through the Psychonomic Society's Norms, Stimuli, and Data Archive, http://www.psychonomic.org/archive. To access this file, search the archive for this article using the journal (*Behavior Research Methods*), the first author's name (Libkuman), and the publication year (2007).

FILE: Libkuman-BRM-2007.zip

DESCRIPTION: The compressed archive file contains six files:

Libkuman(2007)APP-B.txt, containing means and standard deviations for emotions;

Libkuman(2007)APP-B.xls, containing the above information in excel spreadsheet format;

Libkuman(2007)APP-C.txt, containing means and standard deviations for emotions;

Libkuman(2007)APP-C.xls, containing the above information in excel spreadsheet format;

Libkuman(2007)APP-D.txt, containing means and standard deviations for emotions;

Libkuman(2007)APP-D.xls, containing the above information in excel spreadsheet format.

AUTHOR'S E-MAIL ADDRESS: libku1tm@cmich.edu.

(*Continued on next page*)

APPENDIX
Rating Booklet

1. Number 1 indicates you felt completely unhappy while viewing the picture; number 9 indicates you felt completely happy.

Completely unhappy								Completely happy
1	2	3	4	5	6	7	8	9

2. The number 1 indicates you felt completely relaxed or unaroused while viewing the picture, and number 9 indicates you felt completely stimulated or aroused.

Completely unaroused								Completely aroused
1	2	3	4	5	6	7	8	9

3. A rating of 1 on the scale indicates that you think the picture is completely indistinct or common, and the highest rating of 9 indicates you think the picture is completely distinct or unique.

Completely indistinct								Completely distinct
1	2	3	4	5	6	7	8	9

4. A rating of 1 indicates a picture that is completely inconsequential or unimportant to you personally. A rating of 9 indicates a picture that is completely consequential or important to you personally.

Completely inconsequential								Completely consequential
1	2	3	4	5	6	7	8	9

5. The number 1 indicates a picture that is completely forgettable or unmemorable to you. A rating of 9 indicates a picture that is completely memorable or unforgettable to you.

Completely forgettable								Completely memorable
1	2	3	4	5	6	7	8	9

6. The number 1 indicates a picture that is completely unsurprising to you, and the number 9 indicates a picture that is completely surprising to you.

Completely unsurprising								Completely surprising
1	2	3	4	5	6	7	8	9

7. For meaningfulness, the number 1 represents a picture that is completely meaningless or empty to you, and a rating of 9 indicates a picture that is completely meaningful or rich to you.

Completely meaningless								Completely meaningful
1	2	3	4	5	6	7	8	9

8. For the rating of familiarity, the scale represents **how often** in your experience you have seen a scene like the one in the picture. The number 1 represents "never" and the number 9 represents "very often."

Never								Very often
1	2	3	4	5	6	7	8	9

Now please indicate the intensity of the emotion the picture made you feel. A rating of 1 indicates you felt none of the emotion and a rating of 9 indicates you felt the emotion very strongly.

		None								Very strongly
9.	Happiness	1	2	3	4	5	6	7	8	9
10.	Surprise	1	2	3	4	5	6	7	8	9
11.	Sadness	1	2	3	4	5	6	7	8	9
12.	Anger	1	2	3	4	5	6	7	8	9
13.	Disgust	1	2	3	4	5	6	7	8	9
14.	Fear	1	2	3	4	5	6	7	8	9

(Manuscript received August 2, 2005;
revision accepted for publication February 28, 2006.)