# Lists of Emotional Stimuli

**2 authors:**

Daniel Grühn
North Carolina State University
**61** PUBLICATIONS   **2,879** CITATIONS

SEE PROFILE

Neika Sharifian
**42** PUBLICATIONS   **858** CITATIONS

SEE PROFILE

# Lists of Emotional Stimuli

*Daniel Grühn and Neika Sharifian*
Department of Psychology, North Carolina State University, Raleigh, NC, United States

To advance our understanding of emotion or emotional processes, researchers have developed different lists of emotional stimuli. Research on emotion has made substantial strives based on the availability of these lists of emotional stimuli; in particular, standardized sets of words, images, faces, and film clips facilitated emotion research by allowing researchers to select appropriate stimuli and to compare findings across laboratory settings. In the early years, the primary purpose of constructing these lists was to get an idea of the structure of emotion. For example, Osgood and his colleagues (Osgood, 1952; Osgood, Suci, & Tanenbaum, 1957) used a list of opposing affective terms (eg, hot–cold) to arrive at the underlying emotional dimensions that they termed evaluation, activity, and potency. Nowadays, we typically label these dimensions as valence, arousal, and dominance, respectively. These initial attempts at deriving the structure of emotion were complemented by developing new assessment methods for different affective dimensions. For example, one widely used assessment method is the Self-Assessment Manikin (SAM), a picture-oriented instrument developed by Lang and colleagues (Bradley & Lang, 1994; Lang, 1980). The SAM was constructed as a nonverbal instrument to assess the dimensions of valence, arousal, and dominance based on the graphic depiction of manikins. It is probably the most widely used assessment method for evaluating emotional material.

Modern lists of emotional stimuli are, in contrast, primarily utilized for having a standardized set of stimuli with corresponding norms that are easy to use and easy to reproduce by other researchers, that is, modern lists provide a standard set of stimuli to compare and to evaluate the work of others. These lists provide a major advantage for emotion research in making findings comparable across studies. For example, if researchers use images from a published list, they can specify the image number for replication rather than having to provide all the images. This advantage can also be a major drawback: Using standardized lists of emotion stimuli assumes that the established norms of those lists are actually accurate or up-to-date. If norms were wrong or outdated, many studies would base their selection of emotional stimuli on the same wrong or outdated evaluations. It is, for example, easy to imagine that evaluations of emotional images provided by participants in the 1970s or 1980s differ from evaluations provided by participants in the more recent past. An image of the original World Trade Center in New York will probably be differently evaluated before 9/11 and after 9/11 in its emotional connotation. Most likely, if asked before 9/11, participants would probably have rather neutral evaluations of this image; whereas if asked after 9/11, participants might have more negative associations due to the terror attack, which might actually fade over time as the memory of the event fades away. Thus, the meaning of individual images may shift and past established norms might no longer accurately reflect the emotional connotation in the present. Additionally, an image of

the original World Trade Center in New York would probably be evaluated differently by participants in the United States compared to participants living in China due to different contextual knowledge and historical associations with the image. That said, norms of emotional stimuli are embedded in a cultural and historical context and may not generalize to other cultural and historical contexts. Despite this drawback, modern lists of emotional stimuli provide a major advantage for researchers in many different contexts. It even gives researchers the tool—a standard—to evaluate whether norms changed over historic times (however, we are not aware of research that has done that) or whether norms systematically differ between groups.

In this chapter, we will review some major lists of emotional stimuli for words, images, faces, and film clips. In this context, we use the term "list" rather broadly to indicate a collection of stimuli. We wrote this chapter to provide researchers who are interested in utilizing emotional stimuli for experimental settings with a quick overview of different lists. Thus, our comments focus on aspects that might be relevant for researchers in making decisions about the type of emotional stimuli.

# 1    Comparing different types of emotional stimuli

A crucial decision for researchers in designing experiments is the choice of a specific type of emotional stimuli, that is, whether words, images, faces, or film clips will be used as emotional material. In some cases, the decision is obvious depending on the research question. If a researcher is interested in emotion recognition in facial expressions, the type of stimuli is already determined. In other situations, the decision is less obvious. If a researcher is interested in doing research on emotional memory, there is flexibility in the utilized material. Many emotional memory tasks use images or words as to-be-remembered material. But what criteria could be used to make an informed decision about the type of emotional stimuli, that is, whether it should be words or images? To make this and similar decisions easier, we propose to consider the Emotion Matrix of the emotional stimuli. The Emotion Matrix is a graphical representation of five crucial characteristics of emotional stimuli: (1) ecological validity, (2) temporal resolution, (3) controllability, (4) complexity, and (5) emotional intensity. The Emotion Matrix provides a representation of the expected range of stimuli in a stimuli category, such as words or images, on these five characteristics. We have found in our own work (Grühn & Scheibe, 2008; Grühn, Smith, & Baltes, 2005; Kunzmann & Grühn, 2005) that we frequently considered these five characteristics in the selection and decision process for emotional stimuli.

Ecological validity refers to how similar the emotional reaction is to real emotional experiences in everyday life. This characteristic assumes that most emotional experiences are embedded in a larger context of triggering the emotional episode. For example, imagine that when you first moved in with your long-term partner, your partner agreed to be responsible for washing the dishes. However, over the next few months, you continually have to remind your partner to wash the dishes. The next day you come home to find a sink full of dirty dishes, and your partner sitting

in the living room playing video games. You would appropriately become very angry about your partner's lack of participation in cleaning the house. This whole episode is embedded in the larger context of how you and your partner divided responsibilities and what expectations for the situation are. For example, you would probably not be angry if your partner would do the dishes all the time, or if you knew that your partner had a miserable day at work and just needed a few minutes to relax before doing the dishes. The contextual information that makes emotions meaningful to us is typically lost when using highly standardized emotional stimuli. Thus, most types of stimuli are rather at the low end of ecological validity than on the high end. *Temporal resolution* indicates the time it takes to process (or to use) the emotional stimuli. High temporal resolution refers to stimuli that can be processed very fast; low temporal resolution refers to stimuli that require more processing time. Typically, more processing time indicates that more context information is provided and is required to be processed in making the stimulus meaningful. *Controllability* refers to how easy or difficult it is to control stimuli in experimental settings. This includes the control of different potential confounding variables. For example, different word characteristics (eg, word frequency in common language) are relatively easily assessed and can be used to match different groups of emotional words (eg, matching positive and negative words on word frequency). Film clips, in contrast, provide an abundance of information that is difficult to control. Thus, controllability incorporates the idea of how easy it is for a researcher to ensure that findings are due to the desired effects rather than some arbitrary variable. In that sense, controllability encloses the idea of internal validity. *Complexity* is nearly the opposite of controllability. Complexity refers to how complex this type of emotional stimuli could be. Low complexity refers to stimuli with few aspects; high complexity refers to stimuli with many aspects. Given that very complex stimuli are also difficult to control in experimental settings, complexity and controllability tend to go in opposite directions. Finally, *emotional intensity* refers to how intense the emotional response could be that is elicited by the stimulus. Some stimuli, such as words, elicit rather low emotional reactions. It seems rather difficult to design an experiment with emotional words to elicit strong emotional reactions, such as crying during the experiment. On the other hand, film clips provide contextual information to allow participants to experience rather strong emotional reactions. As such, it is much more likely to observe emotional expressions of experiencing an emotion in participants viewing film clips than in participants viewing words.

For each of the following types of emotional stimuli—words, images, faces, and film stimuli—an emotion matrix is presented. Please note that the depicted emotion matrices provide only a rough portrait of these five crucial characteristics. Unfortunately, little research has been conducted in comparing different types of stimuli directly. Thus, the emotion matrices are based on our experience of working with these types of stimuli rather than based on empirical research. We hope, however, that the emotion matrix might be helpful in guiding researchers in their decision process and it might elicit more research on comparing different types of emotional stimuli in the future.

## 2   Emotional words

Emotional words are frequently used in emotional research settings, especially for research on memory, attention, priming, impression formation, and product evaluation. Fig. 7.1 shows the Emotion Matrix for words. Dark areas represent the expected range of word stimuli on the five characteristics. Words have major advantages for experimental designs. Words have a high temporal resolution; thus, persons can process the emotional content quickly. This either means that experiments—all other things equal—can be designed to be relatively brief in duration or that experiments can include many word stimuli to obtain improved estimates of the expected effects. Words are also relatively simple stimuli in terms of the contextual information provided to elicit an emotional response. Given that other word characteristics can be readily assessed, it is easy to control words for potentially confounding variables. This property makes words ideal for controlled experimental settings. There are, however, two drawbacks for emotion research: Emotional words provide low ecological validity and low emotional intensity. Words are stripped of context creating some ambiguity in the meaning of the word and words typically do not provide enough information to elicit strong emotional reactions. If they do, the context for the emotion is typically not provided by the word itself but by the participant. For example, a word recall task may include "baby" as a to-be-remembered word but unknown to the researcher, a participant had a spontaneous abortion and shows strong emotional reactions to this particular word. The participant brings the context for the emotional episode and the word triggers the person-specific context in creating an emotional reaction. Thus, the emotional aspect of words is often in the eye of the beholder rather than in the word itself. In normal experiment settings, such situations should occur at random and should cancel each other out over a larger pool of words and participants. One could, however, also take advantage of this property of words that they could mean different things to different persons: For example, by extending the example above, one could design a study with childless women consisting of
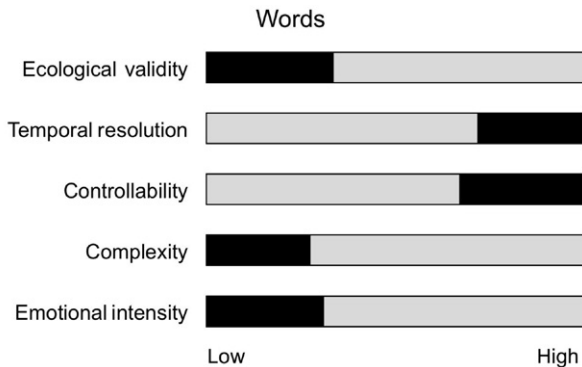


**Figure 7.1** Emotion Matrix of Emotional Words. Dark areas represent the expected range for words on the five characteristics.

one group who strongly desires an own child but were unsuccessfully so far and one group who do not want to have a child. These two groups of women probably have different associations to baby-related words. A task (eg, memory or attention task) could be designed to take advantage of this difference between groups in associating different aspects with the same words. Similarly, to manipulate the meaning of the same word for different people, researchers could create a context for the words by using specific instructions or priming conditions in an experiment. That said, words are rather versatile stimuli in many different experimental settings.

There are some general findings on the usage of emotional word material that need to be considered for designing experiments for emotion research. In particular, there are three interrelated patterns that need to be considered: There are more negative words than positive words in a language (and in corresponding lists of emotional word stimuli); in contrast, positive words occur more frequently in common speech (eg, "happy" is used more often than "fearful") and tend to be shorter in length (in number of letters or number of syllables) than negative words (Grühn, 2015; Ortony, Clore, & Foss, 1987). This pattern signals that we more often use positive than negative words in our language but also that negative words are more differentiated to fit different situations than positive words (eg, Mehl & Pennebaker, 2003). When designing cognitive tasks using emotional word stimuli, these objective characteristics of word length and word frequency are typically major factors in affective cognitive processing. For example, words with a high frequency are better remembered in free recall tasks than words with a low frequency and the reverse is true for recognition tasks, in which low-frequency words are better recognized than high-frequency words (Gorman, 1961; Schulman, 1967). Given the link between objective word characteristics, such as word frequency and word length, and the memorability for these words, research on emotional memory for words as to-be-remembered material needs to control for these potentially confounding variables. Thus, a researcher interested in emotional memory for positive, neutral, and negative words should match the emotional word categories on word length and word frequency to avoid spurious effects due to these objective word characteristics rather than to the emotional connotation of the words. Similar associations between the emotional connotation of the words and objective word characteristics can be found for arousal: Typically, high-arousing words are longer words and less-frequent words than low-arousing words. Thus, lists of emotional words should ideally include subjective evaluations relevant for research on emotion as well as word characteristics relevant for cognitive processing.

Most word norms for cognitive experiments, however, do not contain emotional aspects at all (eg, Paivio, Yuille, & Madigan, 1968) or contain only ratings of emotionality, that is, from neutral to very emotional, how emotional the stimulus is (eg, Strauss & Allen, 2008). Emotionality is, however, not the same as valence or arousal. For example, "love" and "rage" are highly emotional word; however, both are on opposite ends of the valence dimension and "rage" is typically a more arousing word than "love." There are few word norms that contain valence ratings, also called goodness or pleasantness (Brown & Ure, 1969; Clark & Paivio, 2004; Rubin & Friendly, 1986). There are two widely used lists for emotion research. To assess the emotional undertone of natural language, Whissell (1989, 2009) developed a list of 8734 words

called the Dictionary of Affect containing ratings for pleasantness, activation, and imagery. The goal was to use the list to measure the emotionality of a larger segment of text or speech. Therefore, the list of words had to be large. Probably as a consequence of the large number of words, the ratings were simplified and only conducted on 3-point scales (eg, pleasantness: unpleasant—in between—pleasant). The employed 3-point-scale (compared to 7-point or 9-point scales used in this field) limits the differentiation of the words on the subjective word characteristics. That said, Whissell's Dictionary of Affect is probably a good tool for evaluating the emotional content of natural language; but it might be less ideal for selecting word material for experimental settings. One systematic attempt to assess word stimuli for emotion research is the list of Affective Norms for English Words (ANEW; Bradley & Lang, 1999). In the ANEW, valence, arousal, and dominance ratings were assessed for 1034 words on 9-point SAMs. The ANEW has been widely used in hundreds of research settings on emotion and cognition.

A drawback of the ANEW and the Dictionary of Affect is that hardly any other relevant word characteristic is included in the lists. Two recent attempts have been made to provide a more generalized framework for subjective and objective word characteristics that could advance the usage of words in research focusing on emotional, social, and personality aspects. First, Warriner, Kuperman, and Brysbaert (2013) provided valence, arousal, and dominance ratings for 13,915 English words that are directly tied to previous work assessing word frequency (Brysbaert & New, 2009) and age-of-acquisition data (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Given the tremendous number of emotional words in addition to other sources of word characteristics, this list provides a unique contribution to advance the field of emotion research.

A slightly different approach has been used to develop the list of English EMOtional TErms (EMOTE, Grühn, 2015). Whereas Warriner et al. (2013) obtained few ratings for a large number of words, EMOTE was intentionally designed to provide a large number of word characteristics for a medium-size list of words. In particular, EMOTE provides subjective ratings for 1287 nouns and 985 adjectives. Both, nouns and adjectives, were rated on valence, arousal, emotionality, concreteness, imagery, familiarity, and clarity of meaning. Only adjectives were also rated on control, desirability, and likeableness. Moreover, memorability scores (ie, how easy it was to recognize a word) were obtained for the nouns from a large memory task. Ratings were complemented by word length and word frequency data. The EMOTE norms provide an easily accessible word pool for research in the socio-emotional domain that provides norms for different experimental designs. For example, the selection of stimuli for emotional memory experiments could benefit from incorporating imagery ratings or memorability ratings to improve the matching of positive and negative words. Another suggestion could be the use of the desirability or likeableness ratings for impression formation experiments. Likeableness refers to how much one likes an attribute in another person with the characteristic; desirability refers to how much one likes an attribute for oneself.

In addition to these large-scale word lists for general emotion research, some word databases have taken into account potential interindividual differences in the

perception of the emotional words. This is an issue given that most ratings were obtained from young college students rather than from a diverse background. For example, one important aspect to be considered is whether different age groups perceive emotional words—and emotional stimuli more generally—similarly. That is, if an emotional stimulus is presented to an older person, can I expect to see the same emotional reactions that I got from a younger person? If one is interested in age differences of emotional memory, that is, whether younger and older adults remember positive, negative, and neutral material differently, a crucial question is whether the to-be-remembered material provides similar emotional connotations to younger and older adults. If there are age differences in the emotional perception of the material, age differences in memory for that material would be confounded with age differences in the salience of the emotional material. To address this concern, the Age-dependent Evaluations of German Adjectives (AGE; Grühn & Smith, 2008) as well as the French Emotional Evaluation List (FEEL; Gilet, Grühn, Studer, & Labouvie-Vief, 2012) were developed. For AGE, younger and older adults were asked to evaluate the valence, arousal, control, imagery, self-relevance, and age-relevance for 200 German adjectives. For FEEL, younger, middle-aged, and older adults were asked to evaluate 835 French adjectives on valence, arousal, and imagery. Both, AGE and FEEL show substantial age differences in the evaluation of the emotional material. The pattern of findings clearly suggests that younger and older adults evaluate words differently in their emotional meaning. There are many possible reasons for these observed age differences, such as age-related or cohort-related differences in the perception and associated meaning of the word material. These differences may occur as a function of life experience, amount of exposure to certain situations or objects, or age-related biological changes. Similar differences may occur for other group comparisons.

Another interesting route in extending large-scale word lists for emotion research is the development of word lists describing different aspects of an emotional state. For example, the feeling of disgust has not only the subjective affective component of feeling "disgusted" but also corresponding bodily reactions in different forms, such as "vomiting." There are corresponding common phrases in everyday speech on these associations, such as being disgusted by something or someone that it "makes you wanna throw up." However, these and similar aspects of the emotional state are typically not assessed when asking participants for their feelings. To provide researchers with an organized list of representing different aspects of the emotional state, Ortony and colleagues (Clore, Ortony, & Foss, 1987; Ortony et al., 1987) developed a list of 585 words representing four aspects of emotional states: (1) words that represented external conditions such as subjective evaluations (eg, disagreeable, wonderful) and objective descriptions (eg, abandoned, safe), (2) words that represented nonmental internal conditions such as physical and bodily states (eg, rested, weary), (3) words that represented mental internal conditions such as cognitive states (eg, amazed, baffled) or cognitive-behavioral frames of mind (eg, stubborn, friendly), and (4) words with an affect focus, such as commonly used emotion words (eg, afraid, fear, proud). Although other lists contain similar words, the contribution of this work is to organize the words according to the associated affective meaning.

# 3   Emotional images

Emotional images are frequently used for emotion research. There are probably two major reasons for the frequent use of images. One reason is the availability of large-scale databases, such as the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008). These databases make it easy to use images for experimental settings. The second reason is the flexibility of using images for different purposes. This flexibility is visible in the rather wide range of all five characteristics in the emotion matrix (see Fig. 7.2). Images have low to medium ecological validity depending on the depicted content. On the one hand, some images probably require more contextual information that is not given in the image to understand fully the implications of the depicted content. For example, an image showing a group of men with guns could induce some form of fear but without further contextual information, it is unclear what the men are doing. On the other hand, some images provide enough visual information to elicit an emotional response. In this category are primarily disgusting images, such as a dirty toilet or a mutilated body. A feeling of disgust can easily be elicited with images. Emotional responses to images range from low (neutral) to medium intensity. It is rare to elicit a clear and strong emotional response just by viewing images, but images clearly have the potential to provide stronger emotional reactions than reading a corresponding emotional word. For example, the word "mountain" may elicit—if any—rather neutral emotional reactions. Images of mountains, in contrast, may vary widely in the elicited reactions from inspiring awe to neutral feelings to fear of heights.

The temporal resolution of images varies from medium to high; typically, content in images can be processed quickly; however, some images contain a lot of detail and the emotion depends on the accurate processing of all information. For example, one image in the IAPS shows a person sitting on a bench; some people in our research failed to recognize the snake approaching the person's feet in the grass. Obviously, the snake is an important detail in the image for the emotional response but it
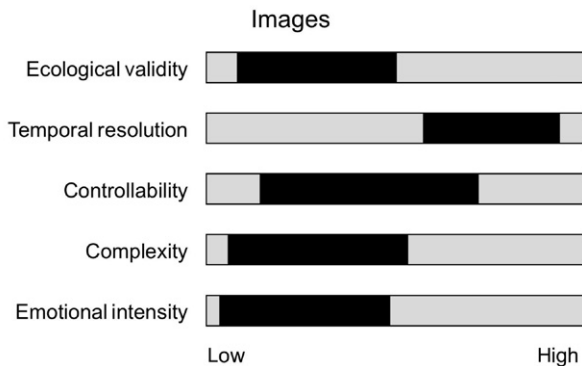


**Figure 7.2** Emotion Matrix of Emotional Images. Dark areas represent the expected range for images on the five characteristics.

requires some active scanning of the image not to miss it. That said, complexity of images varies similarly from low to medium. The depicted content in images ranges from a simple pattern to household objects to people and to complex social scenes. Controllability is generally difficult for images. Images vary in many dimensions, including size, brightness, color scheme, resolution, and type of depicted content. Past research suggests that these variables influence the affective response (eg, Bradley, Hamby, Löw, & Lang, 2007; Codispoti & De Cesarei, 2007) but it is rather difficult to match images on these characteristics.

The IAPS (Lang, Bradley, & Cuthbert, 2001) is a well-known and widely used stimulus set of image material. Using Self-Assessment Manikins, images were rated on valence, arousal, and dominance. The IAPS include a wide range of scenes, including animals, people, landscapes, objects, war scenes, sexual interactions, illness, and others. Hundreds (if not thousands) of studies have selected their pictorial material from the IAPS. However, the IAPS was developed at a time where image quality was not consistently high and the image content clearly shows a large number of images from North American contexts. That said, a drawback of the IAPS is that the images appear to be outdated compared to modern standards of image quality and modern cultural contexts as well as being biased in the culturally depicted content.

To address these issues, three recent image databases have been developed: The Geneva Affective Picture Database (GAPED; Dan-Glauser, & Scherer, 2011), the Nencki Affective Picture System (NAPS, Marchewka, Żurawski, Jednoróg, & Grabowska, 2013); and the Affective Image Database (AID, Grühn, Kapkin, & Sharifian, 2015). The GAPED collected 730 images intended to cover specific themes more systematically. In particular, for negative images, there was an emphasis on including images of spiders, snakes, and scenes of moral or legal violations; for neutral images, primarily images of unanimated objected were chosen; and for positive images, mainly human and animal babies as well as landscapes were selected. Images were evaluated on valence and their social relevance. The NAPS consisted of 1256 higher-quality images covering five content categories: People, faces, animals, objects, and landscapes. Images were evaluated on valence, arousal, and approach-avoidance behavior. For studies requiring a large number of images, the AID collected 7232 images, which were categorized into a two-level hierarchy. The main level consists of five groups of images: *Animals, Humans, Objects, Indoors*, and *Outdoors*. Each main category is divided into subcategories. All images in the AID were rated on valence and arousal using 7-point versions of the Self-Assessment Manikins. These recent efforts in creating new lists of images for emotion research should further our understanding of images as emotional stimuli and should spark new endeavors into emotion research.

## 4 Emotional faces

Faces are an interesting subsection of images. All of the major image databases also contain portrait-like shots of faces. However, the number of these face stimuli in general image lists is rather low, which prevents a systematic utilization of face stimuli. Thus, systematic lists of emotional faces have been developed for experimental

research. The specific subgroup of faces is slightly different from images in general in the emotion matrix. Compared to images in general, emotional face stimuli have the highest ecological validity, that is, faces provide better information about the emotional state than an image of a gun. Although the ecological validity of faces is high compared to other images, it is still only in the low to medium range. We evaluate facial expressions constantly in our social partners; however, this is typically done in a dynamic fashion based on contextual information. For example, if your best friend talks about having a divorce, you know—even without looking at their face—that your best friend is not in the best mood. Thus, although we are good at decoding emotions from static facial expressions, it is rather uncommon in everyday life as a single source of information. Faces have a rather high temporal resolution; again, humans are very good at decoding facial expressions and can recognize and differentiate an angry face from a happy face quickly. Face stimuli also have the advantage that they are easier to control than other images given that the depicted object—the face—is constant. In addition, many face databases contain different emotional expressions posed by the same actor. This helps to design experiments where it is clear that the effects are due to the facial expression rather than the actor. Facial images tend to have low to medium emotional complexity, and low to medium emotional intensity. Faces typically elicit some kind of affective response—even faces with neutral expressions have some affective associations. For example, some faces might elicit more positive associations because they are rather attractive and some might elicit more negative associations because they are less attractive. We typically don't have that affective evaluation for images of neutral objects, such as an image of a table. On the other side, there are probably nonfacial stimuli that can elicit stronger emotional reactions than just viewing an emotional face. Again, seeing an image of a dirty toilet or of a mutilated body probably elicits a stronger feeling of disgust than seeing a face with a disgusting expression (Table 7.1).

Early attempts to create a normed set of facial images were made by Ekman and Friesen (1976) in the Picture of Facial Affect database (POFA). Participants were instructed with a highly controlled protocol to manipulate specific muscles that were characteristic of six discrete emotional expressions: happiness, sadness, surprise, fear, disgust, and anger. This database consists of 110 black-and-white images with participants facing forward. The Productive Aging Lab (PAL) Face database (Minear & Park, 2004) developed a facial image database that consisted of an adult lifespan sample of actors posing the expressions (19–93 years). The list consisted of 1142 black-and-white images from 576 individuals. Images consisted of neutral images where participants were forward-facing or in profile. Additional photos were taken of participants expressing a wider range of emotions, such as happiness, sadness, anger, disgust, and surprise. However, these emotional expressions do not exist for all individuals within the PAL Face database. Ebner, Riediger, and Lindenberger (2010) created the FACES database, in which face stimuli were varied systematically across age, gender, and emotional expression. The FACES database consists of 2052 facial images of 171 individuals stemming from three age groups (young, middle-aged, or older adult) and posing six emotional expressions (neutral, sadness, disgust, fear, anger, and happiness) twice.

**Table 7.1** **Sample lists of emotional stimuli by type of stimulus**

| Authors | List | Type and number | Ratings |
|---|---|---|---|
| **Words** | | | |
| Bradley and Lang (1999) | ANEW | | Dimensional on 9-point SAMs<br>• Valence, arousal, dominance |
| Whissell (2009) | DoA | 8742 English words | Dimensional on 3-point bipolar scales<br>• Pleasantness, activation, and imagery |
| Warriner et al. (2013) | | 13,915 English lemmas | Dimensional on 9-point bipolar scales<br>• Valence, arousal, and dominance |
| Grühn (2015) | EMOTE-A | 985 English adjectives | Dimensional on 7-point bipolar scales<br>• Valence, arousal, emotionality, concreteness, imagery, familiarity, clarity of meaning, control, desirability, and likeableness |
| Grühn (2015) | EMOTE-N | 1287 English nouns | Dimensional on 7-point bipolar scales<br>• Valence, arousal, emotionality, concreteness, imagery, familiarity, and clarity of meaning<br>Objective Word Characteristic:<br>• Memorability |
| **Images** | | | |
| Lang et al. (2008) | IAPS | 956 images | Dimensional on 9-point SAMs<br>• Valence, arousal, dominance |
| Marchewska et al. (2013) | NAPS | 1356 images | Dimensional on 9-point bipolar scales<br>• Valence, arousal, and dominance |
| Grühn et al. (2015) | AID | 7232 images | Dimensional on 7-point SAMs<br>• Valence and arousal |

**Table 7.1** **(Continued)**

| Authors | List | Type and number | Ratings |
|---|---|---|---|
| **Faces** | | | |
| Ekman and Friesen (1976) | POFA | 110 images | Discrete Facial Expressions<br>• Happiness, sadness, surprise, fear, disgust, anger |
| Minear and Park (2004)<br>Ebner et al. (2010) | PAL Faces<br><br>FACES | 1142 images by 576 persons<br>2052 images by 171 persons | Dimensional<br><br>• Discrete—neutral, anger, disgust, happiness, fear, and sadness |
| **Film clips** | | | |
| Philippot (1993) | | 12 film clips | Discrete—6 emotions<br>• Amusement, anger, disgust, fear, neutral, and sadness |
| Gross and Levenson (1995)<br>Schaefer, Nils, Sanchez, and Philippot (2010) | | 16 film clips<br><br>70 film clips | Discrete—16 emotion terms<br>Dimensional<br>• Arousal on 7-point bipolar scale<br>• Positive and negative affect on subscales<br>Discrete—7 emotions<br>• Anger, fear, disgust, sadness, amusement, tenderness, neutral |
| Sharifian and Grühn (2015) | SEEMS | 21 film clips | Dimensional on 7-point SAMs<br>• Valence and arousal<br>Discrete—20 emotion terms |

Several other emotional face databases have been developed to address specific issues: In particular, a major criticism of highly cited face databases is that they tend to use only white Caucasian individuals. Dozens of smaller facial databases exist that aimed to fill this gap, such as databases of emotional expressions of Japanese women (Lyons, Akamatsu, Kamachi, & Gyoba, 1998) or Iranian persons (Bastanfard, Nik, & Dehshibi, 2007). These databases try to address the issue that facial recognition might be influenced by cultural experiences and might correspondingly differ between different ethnic groups (Fig. 7.3).
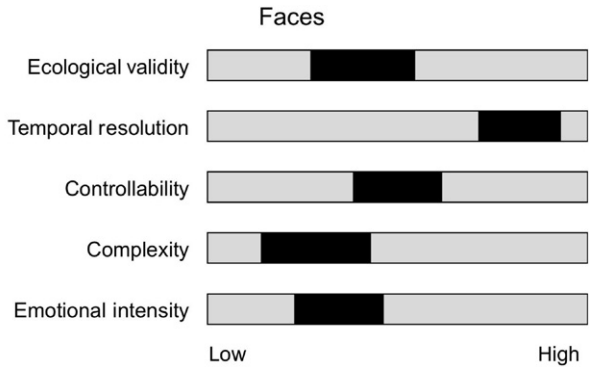
## Faces

| | |
|---|---|
| Ecological validity | |
| Temporal resolution | |
| Controllability | |
| Complexity | |
| Emotional intensity | |

Low                                    High

**Figure 7.3** Emotion Matrix of Emotional Faces. Dark areas represent the expected range for faces on the five characteristics.

## Film clips

| | |
|---|---|
| Ecological validity | |
| Temporal resolution | |
| Controllability | |
| Complexity | |
| Emotional intensity | |

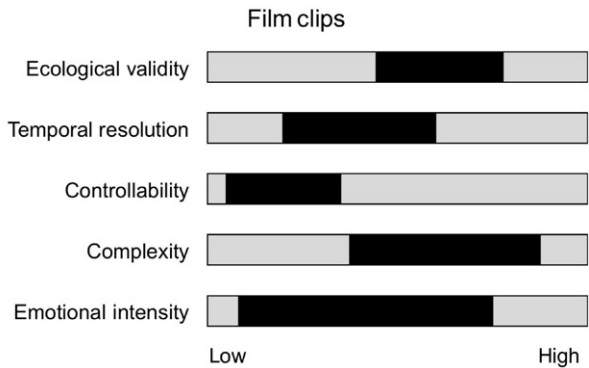Low                                    High

**Figure 7.4** Emotion Matrix of Emotional Film Clips. Dark areas represent the expected range for film clips on the five characteristics.

## 5   Emotional film clips

Film clips provide rich contextual information eliciting emotional responses that are closer to real-life emotional episodes. Similar to real-life emotional episodes, the emotion is elicited dynamically rather than statically. As such, film clips have relatively high ecological validity compared to other types of emotional stimuli but are also relatively high in complexity (see emotion matrix in Fig. 7.4). The temporal resolution is low to medium; the duration of film clips typically varies from one to several minutes. Due to the high complexity and also the low to medium temporal resolution, experimental control of film clips is difficult to attain. Film clips typically vary in too many aspects to hinder effective control of the content. Film clips can, however, elicit a wide range of emotional intensities. Even neutral film clips elicit some kind of affective response; however, film clips can elicit strong emotional reactions.

Film clips are probably not as strong in eliciting emotional reactions as real-life events, such as losing a loved one, but film clips are probably the strongest elicitor of emotion available to researchers in the lab. Thus, film clips are ideal for research settings to elicit strong emotional reactions or to manipulate mood for other tasks.

In contrast to words and images, film clips are—similarly to faces—typically derived to elicit discrete emotions. Thus, ratings are typically done on discrete emotion terms rather than on dimensional ratings, such as valence and arousal. As such, a comparison of different film clips eliciting different discrete emotions is difficult. For example, is a sad film clip eliciting a 5 on a 7-point sadness item as emotional as a fearful film clip eliciting a 5 on a 7-point fearful item? There is no available clear answer to that question. Moreover, the emotion terms used to assess the subjective emotional responses vary widely in the number of emotion terms and the exact wording.

Studies using film clips typically require only a small set of stimuli to elicit an intense emotional reaction (eg, Gross, 1998; Richards & Gross, 2000). This is in contrast to studies using images and words as emotional stimuli, in which researchers typically use many stimuli (eg, Driscoll, Tranel, & Anderson, 2009; Gavazzeni, Wiens, & Fischer, 2008). Thus, there are several lists of a small number (in comparison to the large number of words or images in available lists) of film clips. We just mention here a few: In an early attempt, Philippot (1993) assessed 12 French film clips for six discrete target emotions (two film clips per emotion): amusement, anger, disgust, fear, neutral, and sadness. Findings suggested that film clips were successfully eliciting discrete feelings of amusement, sadness, and a neutral state, but film clips were less consistent in eliciting a discrete feeling of anger, disgust, and fear. Gross and Levenson (1995) selected 16 English film clips for eight discrete emotion categories (two film clips per emotion): amusement, anger, contentment, disgust, fear, neutral, sadness, and surprise. Film clips were rated on a 16-item emotion inventory. Overall, film clips intended to elicit amusement, disgust, sadness, and surprise elicited these emotions clearly. Film clips intended to elicit anger and contentment demonstrated a less clearly distinct emotion profile; and fear film clips were least successful in generating a distinct emotion profile of fear. Thus, the effectiveness of eliciting a target emotion for film clips seems to differ systematically by the emotion.

Recently, two attempts have been made to provide a systematic list of film clips. Schaefer et al. (2010) compiled a set of 70 French film clips from commercially available films across seven emotion categories: anger, fear, disgust, sadness, amusement, tenderness, and neutral. Film clips were rated based on their discreteness, intensity, and level of positive and negative affect. The results showed that most film clips were successful at eliciting discreteness and arousal. Sharifian & Grühn (2015) developed a standardized Set of Emotion-Elicitation Movie Scenes (SEEMS) consisting of 21 English film clips for eight discrete target emotions: neutral (2), sad (3), amusement (3), happiness (2), anger (3), tenderness (2), fear (3), and disgust (3) film clips. Film clips were evaluated retrospectively with a dimensional approach on valence and arousal as well as with a discrete emotion on a 20-item emotion inventory. In addition, valence was rated continuously throughout the presentation of each film clip. There are several lists that have been developed to address specific issues

or cultural aspects. For example, Jenkins and Andrewes (2012) compared the use of nonverbal and verbal contemporary film clips for emotion induction. Hewig et al. (2005) extended previous film clip sets by developing a small subset of neutral film clips. Liang, Hsieh, Weng, & Sun (2013) developed a set of 27 film clips applicable for Taiwanese culture.

In sum, film clips—due to their high similarity to real emotional experiences—are ideal stimuli to elicit strong emotional reactions in the laboratory. Film clips are, thus, often used as stimuli for mood-induction techniques or emotional reactivity studies. The high complexity and low controllability of film clips make them, however, less attractive for highly controlled experimental settings. For example, it is less easy to create a well-controlled memory task based on content presented in different film clips.

## 6   Future avenues

Lists of emotional stimuli have been used in many contexts for furthering research on emotion and emotional processes. Despite the widespread use of these lists, there are three major unresolved issues: First, what is the underlying structure of emotion? Second, how is the emotional structure assessed? And finally, are there systematic interindividual differences in the emotional evaluation of the stimuli?

The first question about the structure of emotion entails the broad question of whether one assumes a dimensional or a discrete approach to emotion. Most lists of words and images use a dimensional approach incorporating—among others—a valence (ie, pleasantness, goodness) and arousal dimension, whereas most lists of faces and film clips use a discrete approach of categorizing stimuli into discrete emotional categories. These two approaches obviously influence the response to the second issue of how to assess the emotional responses to the stimuli. But even without considering this broad question of whether the structure of emotion is dimensional or discrete, there are open questions within each approach. For example, there are conflicting findings about the association of valence and arousal within the dimensional approach. On the one hand, there are several lists of emotional stimuli that report a U-shaped function between valence and arousal, that is, positive and negative stimuli were evaluated as high-arousing, whereas neutral stimuli were evaluated as low-arousing. The U-shaped pattern has been found for words (Bradley & Lang, 1999) and for images (Lang, Bradley, & Cuthbert, 2001). On the other hand, there are several studies finding basically a linear association between valence and arousal that negative stimuli are high-arousing and neutral and positive stimuli are rather low-arousing. Again, this linear pattern has been found for emotional words (Gilet et al., 2012; Grühn, 2015) and emotional images (Keil & Freund, 2009). Sometimes, this conflicting pattern has been found for the same list of stimuli. For example, the U-shaped pattern between valence and arousal of the IAPS based on US college students (Bradley & Lang, 1999) has been found to be linear in a German (Grühn & Scheibe, 2008) and a Brazilian sample (Ribeiro, Sabine, & Bueno, 2005) using similar assessment methods. Obviously, the underlying association between valence and arousal is a major issue for matching stimulus material and for the interpretation of

observed effects. Unfortunately, to our knowledge, no clear explanation has yet been offered to account for the different pattern. It seems that future research on lists for emotional stimuli has to address this issue to determine under which conditions the association would be U-shaped or linear.

The second major issue—how to assess the emotional response to the stimuli—has important ramifications for interpreting the lists. For the discrete account, the exact emotion terms of describing the emotion category matters. For example, is "sad" the same as "unhappy"? Some lists using a discrete account use only a single item such as "sad" for the sadness category, whereas other lists use subscales, such as a sadness scale composed of "sad," "depressed," "blue," and "dejected." These decisions clearly affect how the stimuli are evaluated. For the dimensional account, many lists employed the Self-Assessment Manikin scales (Bradley & Lang, 1994; Lang, 1980) as a nonverbal method of assessing emotional responses. Other lists use bipolar scales (eg, positive–negative, tensed–relaxed) to assess valence and arousal. Again, it is very possible that these assessment decisions have consequences for the selection of stimuli and for the interpretation of findings. Thus, future research would benefit from a coherent framework for comparing and evaluating lists of emotional stimuli.

The assessment issue not only entails the question of how to assess the emotional response but also what to assess. In this chapter, we focused solely on self-report data—given that most lists of emotion stimuli are based on self-report data averaged across persons for the stimuli. There are, however, few attempts at complementing the subjective evaluations with more objective or performance-based responses, such as memorability data or physiological responses. Providing lists with these additional pieces of information could initiate a new wave of advancements in the field of emotion research. For example, instead of using a subjective evaluation of valence and arousal for selecting images for an experiment, it would be great to have information about the facial expression made by persons viewing an image. That could make it very easy to select images eliciting clear facial expressions, such as disgust, for an experiment. Similarly, it is very possible that some emotional images or film clips show similar subjective reactions but different physiological reactions or similar physiological reactions but different subjective reactions (for an age comparison example, see Kunzmann & Grühn, 2005). If there are such stimuli available, it may allow researchers to disentangle the subjective and physiological influences on a specific process.

Finally, more research has to be conducted on understanding systematic interindividual differences in evaluating emotional stimuli. On a practical level, if there are substantial group differences in the evaluation of emotional material, these group differences have to be taken into account when using the material and designing a study. On a theoretical level, systematic group differences may provide insights into the structure, function, and development of emotion. Thus, we would like to encourage more empirical work on generating norms for different groups that might vary systematically in the emotional evaluation of the stimuli. For example, in our own work on age differences in emotional functioning, we obtained norms for younger and older adults for German words (Grühn & Smith, 2008; see also Keil & Freund, 2009), French words (Gilet et al., 2012), and images (Grühn & Scheibe, 2008). We found substantial age differences in all of these studies suggesting to us that the

affective perception of emotional content shifts with age. It seems that the emotional structure gets simpler in old age; but more research is necessary to evaluate this claim. Nevertheless, similar findings might be obtained looking at other groupings, such as culture, occupation, or education.

## 7 Conclusion

Lists of emotional stimuli are advancing systematic research on emotion and emotional processes. Lists are available for every modality of emotional stimuli. In this chapter, we highlighted some of the lists for emotional words, images, faces, and film clips. The chapter was intended as a gateway into exploring lists of emotional stimuli rather than an exhaustive overview of lists for emotional stimuli. We clearly missed mentioning dozens of emotion lists and we would like to apologize for not including all lists.

We would like to emphasize how important the collection of lists of emotional stimuli is for making progress in our field. Not only for using emotional stimuli but also for comparing lists. Even if a list of emotional stimuli is already in existence for your area, it might still be important collecting data on a new list to examine potential differences. For example, it might well be that students evaluate emotional stimuli differently depending on their major; it might be that customs officials evaluate face stimuli differently than other occupations would do; it might be that French-speaking adults evaluate film clips differently than German-speaking adults; or it might be that product designers evaluate products differently than customers who buy the products. All these examples are highly speculative; but without assessing systematic lists of emotional stimuli for these groups, we will never know whether it might not be true. That said, if possible, collect ratings for your emotional material. Maybe it is the beginning of a new major list of emotional stimuli.

## References

Bastanfard, A., Nik, M. A., & Dehshibi, M. M. (2007). Iranian face database with age, pose, and expression. In: *International conference on machine vision* (pp. 50–55). http://dx.doi.org/10.1109/ICMV.2007.4469272.

Bradley, M. M., Hamby, S., Löw, A., & Lang, P. J. (2007). Brain potentials in perception: Picture complexity and emotional arousal. *Psychophysiology*, *44*, 364–373. http://dx.doi.org/10.1111/j.1469-8986.2007.00520.x.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavioral Therapy and Experimental Psychiatry*, 25, 49–59.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. *Technical report C-1, The Center for Research in Psychophysiology*. USA: University of Florida.

Brown, W. P., & Ure, D. M. J. (1969). Five rated characteristics of 650 word association stimuli. *British Journal of Psychology*, *60*, 233–249. http://dx.doi.org/10.1111/j.2044-8295.1969.tb01196.x.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. http://dx.doi.org/10.3758/BRM.41.4.977.

Clark, J. M., & Paivio, A. (2004). Extensions of the paivio, yuille, and madigan (1968) norms. *Behavior Research Methods, Instruments & Computers*, *36*, 371–383. http://dx.doi.org/10.3758/BF03195584.

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, *53*, 751–766. http://dx.doi.org/10.1037/0022-3514.53.4.751.

Codispoti, M., & De Cesarei, A. (2007). Arousal and attention: Picture size and emotional reactions. *Psychophysiology*, *44*, 680–686.

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*, 468–477.

Driscoll, D., Tranel, D., & Anderson, S. (2009). The effects of voluntary regulation of positive and negative emotion on psychophysiological responsiveness. *International Journal of Psychophysiology*, *72*, 61–66. http://dx.doi.org/10.1016/j.ijpsycho.2008.03.012.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES – a database of facial expressions in young, middle-aged, and older women and men: Developmental and validation. *Behavior Research Methods*, *42*, 351–362. http://dx.doi.org/10.3758/BRM.42.1.351.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Gavazzeni, J., Wiens, S., & Fischer, H. (2008). Age effects to negative arousal differ for self-report and electrodermal activity. *Psychophysiology*, *45*, 148–151. http://dx.doi.org/10.1111/j.1469-8986.2007.00596.x.

Gilet, A. -L., Grühn, D., Studer, J., & Labouvie-Vief, G. (2012). Valence, arousal, and imagery ratings for 835 French attributes by young, middle-aged, and older adults: The French Emotional Evaluation List (FEEL). *European Review of Applied Psychology*, *62*, 173–1181.

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, *61*, 23–29. http://dx.doi.org/10.1037/h0040561.

Gross, J. J. (1998). Antecedent- and response focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*, 224–237. http://dx.doi.org.prox.lib.ncsu.edu/10.1037/0022-3514.74.1.224.

Gross, J., & Levenson, R. (1995). Emotion elicitation using films. *Cognition and Emotion*, *9*, 87–108. http://dx.doi.org/10.1080/02699939508408966a.

Grühn, D. (2015). *An English Word Database of EMOtional TErms (EMOTE)*. Under Review.

Grühn, D., Kapkin, E., & Sharifian, N. (2015). *Affective Image Database (AID). Technical report A*. Raleigh, NC: North Carolina State University.

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Does age make it more extreme? *Behavior Research Methods*, *40*, 512–5521.

Grühn, D., & Smith, J. (2008). Characteristics for 200 words rated by young and older adults: Age-dependent evaluations of German adjectives (AGE). *Behavior Research Methods*, *40*, 1088–11097.

Grühn, D., Smith, J., & Baltes, P. B. (2005). No aging bias favoring memory for positive material: Evidence from a heterogeneity-homogeneity list paradigm using emotionally toned words. *Psychology and Aging*, *20*, 579–588.

Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., & Bartussek, D. (2005). A revised film set for the induction of basic emotions. *Cognition and Emotion*, *19*, 1095–1109. http://dx.doi.org/10.1080/02699930541000084.

Jenkins, L., & Andrewes, D. (2012). A new set of standardised verbal and nonverbal contemporary film stimuli for the elicitation of emotions. *Brain Impairment*, *13*, 212–227. http://dx.doi.org/10.1017/BrImp.2012.18.

Keil, A., & Freund, A. M. (2009). Changes in the sensitivity to appetitive and aversive arousal across adulthood. *Psychology and Aging*, *24*, 668–680.

Kunzmann, U., & Grühn, D. (2005). Age differences in emotional reactivity: The sample case of sadness. *Psychology and Aging*, *20*, 47–459.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990. http://dx.doi.org/10.3758/s13428-012-0210-4.

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 systems* (pp. 119–137). Norwood, NJ: Ablex.

Lang, P., Bradley, M., & Cuthbert, B. (2001). International affective picture system (IAPS): Instruction manual and affective ratings. *Technical report A-5, The Center for Research in Psychophysiology*. University of Florida.

Lang, P., Bradley, M., & Cuthbert, B. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical report A-8*. Gainesville, FL: University of Florida.

Liang, Y., Hsieh, S., Weng, C., & Sun, C. (2013). Taiwan corpora of chinese emotions and relevant psychophysiological data – Standard chinese emotional film clips database and subjective valuation normative data. *Chinese Journal of Psychology*, *55*, 601–621.

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In: *Proceedings, third IEEE international conference on automatic face and gesture recognition* (pp. 200–205). Japan: Nara.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2013). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*, 596–610.

Mehl, M. M., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*, 857–870. http://dx.doi.org/10.1037/0022-3514.84.4.857.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments & Computers*, *36*, 630–633. http://dx.doi.org/10.3758/BF03206543.

Ortony, A., Clore, G. L., & Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive Science*, *11*, 341–364. http://dx.doi.org/10.1207/s15516709cog1103_4.

Osgood, C. (1952). The nature and measurement of meaning. *Psychological Bulletin*, *49*, 172–237.

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

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*, 1–25. http://dx.doi.org/10.1037/h0025327.

Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and Emotion*, *7*, 171–193.

Ribeiro, R. L., Sabine, P., & Bueno, O. F. A. (2005). Comparison of Brazilian and American norms for the international affective picture system (IAPS). *Revista Brasileira de Psiquiatria*, *27*, 208–215. http://dx.doi.org/10.1590/S1516-44462005000300009.

Richards, J., & Gross, J. (2000). Emotion regulation and memory: The cognitive costs of keep-ing one's cool. *Journal of Personality and Social Psychology*, *79*, 410–424. http://dx.doi.org/10.1037/0022-3514.79.3.410.

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition*, *14*, 79–94.

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, *24*, 1153–1172. http://dx.doi.org/10.1080/02699930903274322.

Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science*, *9*, 211–212.

Sharifian, N., & Grühn, D. (2015). A standardized Set of Emotion-Elicitation Movie Scenes (SEEMS): Development and validation. *Technical report B*. Raleigh, NC: North Carolina State University.

Strauss, G. P., & Allen, D. N. (2008). Emotional intensity and categorization ratings for emotional and nonemotional words. *Cognition and Emotion*, *22*, 114–133. http://dx.doi.org/10.1080/02699930701319154.

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

Whissell, C. (1989). The dictionary of affect in language. In R. Plutchik & H. Kellerman (Eds.), *Emotion: theory, research, and experience* (Vol. 4., pp. 113–131). New York: Academic Press.

Whissell, C. (2009). Using the revised Dictionary of Affect in language to quantify the emo-tional undertones of samples of natural language. *Psychological Reports*, *105*, 509–521.