Emotion

# Semantic and Affective Representations of Valence: Prediction of Autonomic and Facial Responses From Feelings-Focused and Knowledge-Focused Self-Reports

Oz Hamzani, Tamar Mazar, Oksana Itkes, Rotem Petranker, and Assaf Kron

Online First Publication, February 4, 2019. http://dx.doi.org/10.1037/emo0000567

American Psychological Association

Emotion

© 2019 American Psychological Association
1528-3542/19/$12.00

http://dx.doi.org/10.1037/emo0000567

# Semantic and Affective Representations of Valence: Prediction of Autonomic and Facial Responses From Feelings-Focused and Knowledge-Focused Self-Reports

Oz Hamzani, Tamar Mazar, and Oksana Itkes
University of Haifa

Rotem Petranker
York University

Assaf Kron
University of Haifa

The term *valence* can refer to either the affective response (e.g., "I feel bad") or the semantic knowledge about a stimulus (e.g., "car accidents are bad"). Accordingly, the content of self-reports can be more "experience-near" and proxy to the mental state of affective feelings, or, alternatively, involve nonexperiential semantic knowledge. In this work we compared three experimental protocol instructions: feelings-focused self-reports that encourage participants to report their feelings (but not knowledge); knowledge-focused self-reports that encourage participants to report about semantic knowledge (and not feelings); and "feelings-naïve", in which participants were asked to report their feelings but are not explicitly presented with the distinction between feelings and knowledge. We compared the ability of the three types of self-report data to predict facial electromyography, heart rate, and electrodermal changes in response to affective stimuli. The relationship between self-reports and both physiological signal intensity and signal discriminability were examined. The results showed a consistent advantage for feelings-focused over knowledge-focused instructions in prediction of physiological response with feelings-naïve instructions falling in between. The results support the theoretical distinction between affective and semantic representations of valence and the validity of feelings-focused and knowledge-focused self-report instructions.

*Keywords:* semantic valence, affective valence, self-reports, feelings-focused, knowledge-focused

*Supplemental materials:* http://dx.doi.org/10.1037/emo0000567.supp



In the affective science literature, the term *valence* refers to the distinction between attractive/aversive, pleasant/unpleasant, good/bad, and positive/negative (Barrett, 2006). Adding to the confusion, valence can refer to both the affective response (e.g., "I feel bad") or to knowledge about a stimulus (e.g., "car accidents are bad"; Itkes, Kimchi, Haj-Ali, Shapiro, & Kron, 2017; Robinson & Clore, 2002a, 2002b). Accordingly, people can respond affectively to a noxious event, such as negative feelings while watching news about a car accident, or, in other cases, may know that an event is harmful but still develop no affective response to it (e.g., watching the news, knowing that car accidents are negative but experiencing no affective response). We have recently described this distinction in terms of affective and semantic valence; affective valence refers to valence of the affective response (e.g., feeling is positive or negative) while semantic valence refers to knowledge about the valence of a stimulus (e.g., knowing that X is positive or negative; Itkes et al., 2017).

The distinction between affective and semantic valence is highly relevant to self-reported data. When participants report about valence, the content of self-reports can be more "experience-near" and a proxy to the mental state of pleasant and unpleasant feelings, or, alternatively, involve some nonexperiential semantic knowledge about the stimuli being positive or negative (Itkes et al., 2017; Robinson & Clore, 2002a, 2002b). Despite self-reports being the principal data in emotion research, experimental routines often overlook the explicit distinction between affective and semantic valence. To fill this empirical gap, in this work we compared self-report data from three experimental protocol instructions: feelings-focused self-reports, in which participants were instructed to report the degree to which they feel pleasure and displeasure, but not knowledge; knowledge-focused self-reports, in which participants were instructed to report the degree to which they know the stimuli to be positive or negative, but not about their feelings;

Oz Hamzani, Tamar Mazar, and Oksana Itkes, Department of Psychology, University of Haifa; Rotem Petranker, Department of Psychology, York University; Assaf Kron, Department of Psychology, University of Haifa.

Oz Hamzani, Tamar Mazar, and Oksana Itkes Contributed equally to this work. We thank Leigh Hall for proofreading this paper.

Correspondence concerning this article should be addressed to Assaf Kron, Department of Psychology, The Institute of Information Processing and Decision Making, University of Haifa, Mount Carmel, Haifa 31905, Israel. E-mail: akron@psy.haifa.ac.il

and feelings-naïve self-reports, in which participants were asked to report their feelings, but were not explicitly presented with the distinction between feelings and knowledge. We compared the ability of the three types of self-report data to predict facial electromyography, heart rate, and electrodermal changes in response to affective stimuli. This experimental design serves two research questions: First, as will be explained in detail below, our distinction between affective and semantic valence predicts that feelings-focused instructions will outperform knowledge-focused instructions when modeling facial and autonomic affective reactions. Second, it examines the degree to which explicit distinction between feelings and knowledge improves the modeling of facial and autonomic response in comparison to naïve instructions.

## Working Definitions

### Affective Response

The affective reaction can be thought of as activation of multicomponent response channels that indexes the occurrence of an event as pleasant and/or unpleasant (e.g., Bradley, Codispoti, Cuthbert, & Lang, 2001; Dolan, 2002), with a degree of arousal (e.g., Russell, 1980, 2003; Barrett & Russell, 1999, but see also Kron, Goldstein, Lee, Gardhouse, & Anderson, 2013). The affective response includes various components such as action tendencies, autonomic pattern, attribution, appraisal and conscious feelings. What we term here *affective response* is compatible with previous definitions of emotion (e.g., Lang, Greenwald, Bradley, & Hamm, 1993) that emphasize the response being brief and object-related (see Beedie, Terry, & Lane, 2005 for review). We use the term *affect* instead of *emotion* to emphasize that we are working in the theoretical context of a dimensional model with bipolar valence and arousal dimensions (Russell, 2003).

### Feelings

Feelings are one specific component of the affective response (see Barrett, Mesquita, Ochsner, & Gross, 2007; Frijda, 2005; Lambie & Marcel, 2002, for reviews) defined here as the conscious experience of X, where X in our case is affect. As was mentioned above we are working in the context of the bipolar valence arousal model. The bipolar valence arousal model is a well-known dimensional model that defines the affective feelings space with two dimensions of bipolar valence (ranged from pleasant to unpleasant) and arousal (ranged from low to high activation; Russell, 2003). The relation between the arousal and bipolar-valence dimensions is a matter of debate with three main alternatives: first, bipolar valence and arousal are separate dimensions, each with its own intensity (Barrett & Russell, 1999); second, arousal is the intensity of bipolar valence (Bradley et al., 2001); third, arousal is not a separate dimension from bipolar valence nor its intensity; but rather the sum of intensities of separate unipolar dimensions of pleasant (PL) and unpleasant (UN; Kron et al., 2013; Kron, Pilkiw, Banaei, Goldstein, & Anderson, 2015). In this study we assume the third model; specifically, that when valence is measured in two unipolar scales, one for pleasant feelings (PL) and another for unpleasant feelings (UN), arousal is the sum of these feelings (arousal = PL + UN) and bipolar valence is the difference between them (bipolar valence = PL − UN). Supporting this view, we found that a separate arousal scale had no advantage over PL + UN in predicting skin conductance response and bipolar valence showed no advantage over PL − UN in predicting EMG corrugator activation (Kron et al., 2013, 2015).

### Affective Valence

Affective valence, the property of being positive and/or negative, is used to describe the affective response channels. These include autonomic (e.g., heart rate and affective modulation of startle response—Bradley et al., 2001; Bradley, Lang, & Cuthbert, 1993), facial (e.g., corrugator and zygomaticus muscles; Bradley et al., 2001), behavioral (e.g., approach avoidance instigation; Kron et al., 2014), and experiential (e.g., the valence self-report scale; Bradley & Lang, 1994) changes that constitute the affective response to stimuli.

### Semantic Valence

The term *semantic valence* does not refer to the affective response but to the representation of general stored knowledge about the valence of objects and events (Osgood, 1952). This framework is compatible with the taxonomy of episodic and semantic memory (e.g., Schacter, Wagner, & Buckner, 2000; Tulving, 1984, 1993; Wheeler, Stuss, & Tulving, 1997): knowledge about valence of events can be episodic, that is, related to a specific episode at a particular time and place ("We had a car accident 2 years ago, it was horrible."), or alternatively, semantic, that is, representing general conceptual impersonal knowledge about an event ("car accidents are horrible"). In the current study, the term semantic valence refers to semantic rather than episodic knowledge about an event.

### Feelings-Focused and Knowledge-Focused Self-Reports

The distinction between affective and semantic valence is important for precise measurement and interpretation of self-reported data, as self-reports can be a proxy for experiential feelings in some cases, and reflect semantic knowledge in others. It is widely assumed that valence is a fundamental part of the structure of the human meaning (semantic) system (Osgood, 1952). Semantic representation of any given valence value might be composed of several facets, including the participant's beliefs about his or her own feelings (Robinson & Clore, 2002a, 2002b); the belief about the potential of the stimuli to change core affect (Russell, 2003); beliefs related to gender stereotypes (Shields, 1987; Widiger & Settle, 1987; Williams & Bennett, 1975), and cultural stereotypes (e.g., Diener, Diener, & Diener, 1995); as well as beliefs about feelings in specific situations (Kahneman, 1999). These facets can converge into a decision about one valence value. This valence value is a semantic evaluation of objects as being negative or positive in a way that reflects the "cognitive structure of affect" (Russell, 1980).

Self-report routines are inconsistent in regard to the distinction between affective and semantic aspects of valence. Some procedures specifically instruct participants to report their feelings, such as the self-assessment manikin (Bradley & Lang, 1994), the semantic differential-emotion instructions (Mehrabian & Russell, 1974); The Affect Grid (Russell, Weiss, & Mendelsohn, 1989) and the Positive and Negative Affect Schedule (Watson & Clark, 1999). Other self-report instructions are ambiguous regarding this issue. For example, participants in Ito, Cacioppo, and Lang (1998) were told that the experiment assesses their opinion and reaction to the stimuli, which may imply a partial semantic component. Finally, some self-report instructions ask to evaluate the stimuli and not the feelings (e.g., Gerber et al., 2008; Tottenham et al., 2009). Asking participants to evaluate if stimuli are positive or negative can also imply semantic and not affective evaluations (Russell, 1980).

Merely specifying throughout the instructions that reports should refer to feelings or objects is not always enough to guarantee an optimal distinction between reporting on affective and semantic information. Robinson and Clore (2002b) suggested an accessibility model that predicts when participants tend to report about affective feelings or semantic evaluations. The accessibility model suggests that with instructions to report about one's own feelings, the more feelings ("experiential knowledge" in Robinson's and Clore's terms) are accessible, the more the report will be based on one's own affective experience. However, when information about feelings is not accessible, it is replaced with nonexperiential knowledge. For example, according to Robinson and Clore, when there is temporal proximity between the affect-inducing event and self-reports (i.e., self-reports about current feelings), the report is more likely to include information about feelings. In contrast, when participants are asked to provide retrospective self-reports, or given considerable time after the affective response occurred (e.g., how did you feel 5 weeks ago?), information about the actual experience is no longer available and participants are more likely to rely on semantic knowledge.

Limited access to experiential information is not restricted to cases of retrospective reports; as it can occur even for a real-time rating. For example, when the intensity of the affective response is substantially low, it is harder to detect feelings (Karmon-Presser, Sheppes, & Meiran, 2018). Consequently, access to experiential information is limited and, according to the accessibility model, self-reports are likely to include stronger semantic components (see also Levenson, 2003). The potential leakage of semantic information into self-reports of feelings with low intensity is relevant to self-reports in a laboratory setting where, for ethical reasons, affect intensity is limited.

We developed two self-report instruction procedures, feelings-focused and knowledge-focused, in order to address these concerns and reduce the interference between semantic and experiential information in self-reported feelings (Itkes et al., 2017; Kron et al., 2015). The full instructions appear in the Method section and SOM 1 in online supplemental material. The main difference between these instructions and traditional routines is that they directly communicate the distinction between the affective- and semantic-content of the reports to the participant. After demarcating feelings and knowledge, the feelings-focused instructions encourage participants to report about their actual feelings rather than semantic knowledge about the content of the stimulus. In contrast to the feelings-focused instructions, the knowledge-focused instructions encourage participants to report about the content of the stimuli and not their feelings.

## The Current Study

The aim of the current study is to examine two research questions that stem from the distinction between affective versus semantic valence. First, will feelings-focused reports show stronger associations with facial and autonomic affective responses than knowledge-focused reports? Second, do feelings-focused reports have an advantage over feelings-naïve instructions in predicting facial and autonomic response?

There are at least two critical differences that are categorically distinct between affective response and semantic knowledge. First, the affective response includes (but is not limited to) a change in feelings, autonomic response and facial activity, but not necessarily a change in semantic knowledge. Furthermore, the components of the affective response, consisting of feelings as well as other changes related to affective valence, are all time sensitive—they increase with stimulus onset and attenuate after stimulus offset (e.g., Bradley et al., 2001; Lang et al., 1993). However, semantic knowledge that represents meaning and ideas is not expected to be time sensitive or dramatically change with exposure to stimuli. We have previously shown that repeated exposure to stimuli does not change semantic knowledge about valence, but does change the measures that are related to the affective response, supporting the disentanglement of knowledge from affect (Itkes et al., 2017).

In this study, we focus on four response channels: facial expressions, which will be collected by surface electromyography from the areas above the corrugator supercilii and zygomaticus major muscles; heart rate as measured by cardiac deceleration in response to stimulus intake; electrodermal change; and self-reports about feelings. Next we describe each response channel in more detail.

Facial expressions are assumed to be a response channel both in the discrete approach (discriminating between different emotional episodes; Ekman, 1993 but see also Aviezer, Trope, & Todorov, 2012) and the dimensional approach (varying with valence dimension; Lang et al., 1993; Larsen, Norris, & Cacioppo, 2003). The function of facial expressions in affective responses is a matter of controversy with evidence for a role in both social nonverbal communication (Frith, 2009) and modulation of sensory acquisition (Susskind et al., 2008). However, not all facial expressions in response to affect-inducing events are a part of the affective response. Facial expressions can be classified to those that originate from the primary motor cortex, which are more voluntarily controlled, and are assumed to reflect strategic reactions or display rules (Cole, 1986; Müri, 2016), versus expressions that originate in the extrapyramidal motor system, to which there is no voluntary control and thus reflect the affective response (Larsen et al., 2003; Rinn, 1984). In this study we use facial electromyographic (EMG) measures from the area above the zygomaticus major and corrugator supercilii muscles. The zygomaticus major pulls the corners of the mouth back and activates a smile; accordingly, it was found to be more activated in pictures that elicit positive feelings than in pictures that elicit negative feelings (e.g., Kron et al., 2013; Larsen et al., 2003). The corrugator supercilii draws the brows medially into a frown and is activated more in response to pictures that elicit negative feelings than to pictures that elicit positive feelings (e.g., Kron et al., 2013; Larsen et al., 2003). We predict stronger associations between corrugator and zygomaticus EMG activations with feelings-focused instructions than both knowledge-focused and feelings-naïve instructions.

The second response channel that will be examined is cardiac changes, specifically cardiac deceleration. Cardiac changes are monitored by both sympathetic and parasympathetic innervations and are the most reported autonomic measure in emotion research (Kreibig, 2010). The exact pattern of cardiac changes is dependent on the type of affective stimuli. Specifically, cardiac deceleration is observed for 2–3 seconds after the onset of visual stimuli (e.g., Bradley et al., 2001) and is usually interpreted as part of the orienting response (Palomba, Angrilli, & Mini, 1997) and linked to the allocation of attention during "stimulus intake" (Lacey & Lacey, 1978). Cardiac acceleration is thought to be related to stronger affective responses and is observed in imagination and social tasks (e.g., Gollnisch & Averill, 1993).

Accordingly, a pattern of cardiac deceleration rather than acceleration is expected with the type of stimulus presented in the current study (e.g., Bradley et al., 2001). The affective modulation of cardiac deceleration is usually associated with the valence dimension along with the more consistent finding that cardiac deceleration is more pronounced for negative than positive stimuli (Palomba et al., 1997). The deceleration effect in response to neutral versus positive stimuli is not consistent. Some studies showed a linear relationship, whereas response to neutral stimuli falls in between those of positive and negative stimuli (Lang et al., 1993; Palomba et al., 1997). Interestingly, other studies that used the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997) show stronger cardiac deceleration in responses for both positive and negative stimuli in comparison to negative stimuli (Bradley et al., 2001). Finally, other studies show no difference between cardiac deceleration in response to positive and neutral stimuli (Itkes et al., 2017). Intersecting the previous results of cardiac deceleration in response to affective visual stimuli suggests only one consistent finding: when classifying the valence axis to the categorical structure of negative, neutral, and positive values, pictures with a negative valence value result in stronger cardiac deceleration than positive pictures. In conclusion, cardiac deceleration is part of the early stages of the affective response and is most likely related to the allocation of attention to visual stimuli with some consistent mediating effect of stimulus valence.

The third response channel that will be examined is electrodermal activation. Electrodermal activation is modulated by sympathetic activation, is assumed to be part of the orienting response (Maltzman & Boyd, 1984), and normally occurs due to a change in the environment and in response to novel stimuli (e.g., Furedy, 1968; Williams, 1963). Affective states increase skin conductance magnitude in response to novel, high-arousal positive and negative stimuli (e.g., Kron et al., 2013, 2015; Lang et al., 1993). This pattern is frequently interpreted as affective modulation of the orienting response and part of the affective response (Bradley et al., 2001).

The fourth response channel is feelings. Although feelings are not easy to define (see Barrett et al., 2007; Frijda, 2005 for reviews) they constitute the response channel that, for laypeople, is more associated with an affect. For most people, feelings are the main and sometimes only interface with their affective response. As was mentioned above in the working definition section, we refer to feelings as the conscious experience of X, where X in our case is affect. Since some definitions of feelings also include unconscious aspects (Damasio, 1999), it is important to emphasize that here we restrict the use of the term feelings to conscious experiences of affect. Given the dimensional model we are working with, we here focus on pleasant and unpleasant feelings. Specifically, feelings will be operationalized using a two-dimensional bivariate valence model with one continuous axis for positive valence and a second potentially (but not necessarily) independent axis for negative valence (see Kron et al., 2013, 2015).

We expect a degree of association between feelings-focused reports, zygomaticus and corrugator activity, and electrodermal changes to the extent that they are all part of the same affective response. However, unlike feelings, we assume that semantic knowledge about the valence of stimuli is not part of the affective response. We consequently predict that to the extent that knowledge-focused self-reports indeed reflect more of semantic knowledge, it will show a weaker association with autonomic and facial reactions than feelings-focused reports.

The experiment was designed to compare the predictive value of the three different sets of self-report instructions for facial electromyography, heart rate response, and galvanic skin response. Three sets of self-report instructions were compared: feelings-focused, knowledge-focused, and feelings-naïve. In the feelings-focused instructions condition, participants went through a procedure that instructed them to report their subjective feelings (and not semantic knowledge about the stimuli). In the knowledge-focused condition, participants underwent a procedure that instructed them to report the content of a picture and not their actual feelings. In the third condition, feelings-naïve, participants were instructed to report their feelings without implementing any distinction between affective and semantic valence and any special procedure, hence the name *feelings-naïve*. Participants in the study looked at still pictures and were asked to report the valence (both positive and negative) according to the instruction conditions. Heart rate, galvanic skin response, and facial EMG were collected.

# Method

## Experimental Practice

Sample size was a priori determined to 35 participants per condition relying on previous studies which used this specific design (Kron et al., 2013, 2015). Termination of data collection was a priori determined to be in $n = 35$ per condition. Analyses were planned a priori and were similar (except minor adaptation to the current research question to Kron et al., 2013, 2015). All data that was analyzed is reported.

## Participants

One hundred five undergraduate students (35 participants in each instruction condition) from the University of Haifa participated in this study, in return for either course credit or monetary compensation. All participants were native speakers of Hebrew. Male and female participants were independently randomly assigned separately to the three conditions to ensure a similar ratio of males to females in each condition. Table 1 describes the number of participants that were not analyzed in each response channel and the reason for data exclusion.

Table 1
*Number of Participants Excluded From Analysis and Reason for Each Response Channel*

| Response channel | Number of excluded participants | Reason |
|---|---|---|
| Self-reports | 0 | N/A |
| EDA | 0 | N/A |
| fEMG corrugator | 1 | Extensive unrelated facial movements |
| fEMG zygomaticus | 1 | Extensive unrelated facial movements |
| HR | 5 | malfunction of recording |

*Note*. EDA = Electro Dermal Activation; FEMG = Facial Electro-MyoGraphy; HR = Heart Rate.

## Stimuli

Seventy two images were selected from the International Affective Picture System (IAPS; Lang et al., 1997). Selection was done using an in-house algorithm that randomly selects a sample of 72 images such that the resulting two-dimensional (valence and arousal) shape of the selected sample is the same as the original shape of the IAPS set. All images were distributed across this shape in a uniform manner and possible combinations of valence and arousal were represented (see Figure 1).

## Self-Report Instructions

**Shared components.** As was mentioned in the Working Definitions section, the current study is done in the context of a dimensional model. Based on previous studies (e.g., Kron et al., 2013, 2015), we assume that when valence is measured in two unipolar scales, one for pleasant feelings (PL) and another for unpleasant feelings (UN), arousal is the sum of these feelings (arousal = PL + UN) and bipolar valence is the difference between them (bipolar valence = PL − UN). In the three instructions, we used separate scales for pleasant and unpleasant feelings and report results in the main text in terms of PL + UN (arousal) and PL − UN (bipolar valence). Results in terms of unipolar pleasure and displeasure scales are reported in SOM 2 and SOM 3 in online supplemental material.

**Feelings-focused.** Full instructions protocol can be found in SOM 1 in online supplemental material. Based on Kron et al. (2015), self-report instructions were developed to reflect the participant's subjective feelings, as opposed to evaluations based on semantic knowledge, expectations, or beliefs. To that end, we used three rating scales ranging from 0 (*none*) to 8 (*high*): general feelings scale (rating the most intensely experienced value of any type of affective feelings, such as arousal, pleasure, displeasure, or any other feeling); pleasure scale (rating feelings of pleasure, happiness, and/or any other pleasant feelings); and displeasure scale (rating feelings of displeasure, sadness). In the beginning of the instruction procedure, participants were informed about the distinction between *feeling* and *knowing*. In particular, we emphasized two cases: (1) confusing the evaluation of feelings with the evaluation of picture content (e.g., you feel an unpleasant/negative feeling vs. the content of the picture is unpleasant/negative), and (2) confusing feelings with beliefs or expectations about what one "should feel" while looking at the picture. Next, participants were familiarized with the three scales. We used the first scale to frame the task as an "affect detection task" and asked participants to report any kind of affective feelings, if they detected any. Participants were told to think of this scale as a volume knob that indicates the intensity of their affect and that their response should be based on whether they detected any feelings (e.g., pleasant, unpleasant, arousing, etc.). If no feelings were detected, they were asked to press [0]. If the participants did detect feeling, they were asked to rate the intensity of that feeling. The purpose of this scale was to reduce the accessibility bias—reporting about semantic knowledge in the absence of strong feelings— by legitimizing cases in which no feelings were experienced. When a participant detected feelings (i.e., rated their affective feelings as nonzero), they used the next two scales to rate how positive/negative those feelings were.

**Feelings-naïve.** Participants were instructed to rate their feelings using only the pleasant and unpleasant scales.

**Knowledge-focused.** These instructions were developed to ascertain the participants' report of the semantic evaluation of the content of the event. To achieve that, we used the same pleasure and displeasure scales and instructions as in the self-report about feelings, but this time participants were instructed to rate how positive/negative the content of the picture was and not their subjective feelings. See SOM 1 in online supplemental material for translated instructions and the rating scales.

## Reliability

In accordance with previous measurements of reliability for dimensional scales (e.g., bipolar and arousal scales), split-half reliabilities were calculated (Lang et al., 1997; Moors et al., 2013). Each of the three instruction groups was randomly split into halves. The correlation between the mean self-reports for each picture of the two halves was computed and presented in Table 2. Mean self-reports for each picture are available in SOM 4 in online supplemental material.

## Physiological Data Acquisition

Physiological data was recorded and amplified with a multichannel BioNex 8-slot chassis (MindWare Technologies, Grah-

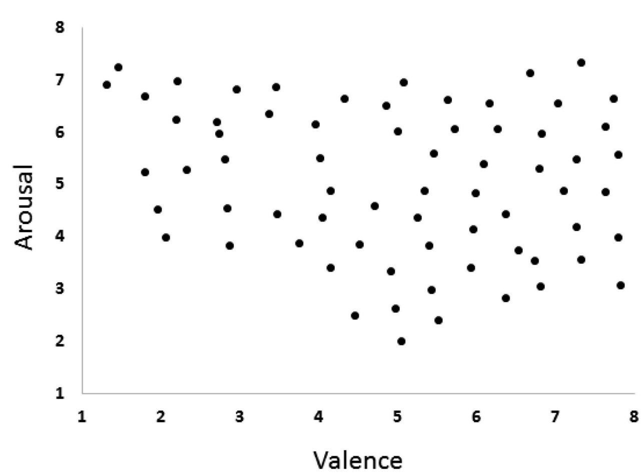


*Figure 1.* Distribution of the selected pictures along the International Affective Picture System (IAPS) bipolar valence arousal space.

Table 2
*Reliability Index (Pearson's Correlation Between Two Halves) for Each of the Four Scales in Each of the Instruction Types*

| Instructions/Scale | Valence | Arousal | Pleasure | Displeasure |
|---|---|---|---|---|
| Feelings-focused | .95 | .86 | .89 | .96 |
| Knowledge-focused | .97 | .79 | .97 | .96 |
| Feelings-naïve | .98 | .80 | .95 | .97 |

anna, OH) equipped with a two BioNex 4-channel bio potential amplifier (Model 50−371102–00). All data was sampled at 1,000 Hz and transmitted to a computer for viewing and storage using MindWare acquisition software BioLab 2.4. The experiment was designed using E-Prime 2 professional software (Schneider, Eschman, & Zuccolotto, 2002), run on an HP PC and a 23" color monitor.

**Facial electromyography (EMG).** Surface EMG was recorded from the areas above the zygomaticus major and corrugator supercilii muscles on the left side of the face (Fridlund, & Cacioppo, 1986) with 4 mm miniature Beckman Ag/AgCl electrode pairs (1 cm between electrodes), filled with the designated gel. Before electrode application, Nuprep (Weaver and Company, Aurora, CO) was applied to the designated skin sites to lower interelectrodes impedance to 10kW. Note that although we aimed for impedance under 10kW, for some participants this value cannot be reached without damaging the skin. In such cases, we also included the cases where impedance was under 30kW.

**Heart Rate (HR).** HR was extracted from an electrocardiogram (ECG) signal that was recorded using two electrodes placed on the right collar bone and the 10th left rib.

**Electrodermal activation (EDA).** EDA was recorded using two 1/1.5 foam disposable electrodes placed on the palm (thenar and hypothenar eminence) of the left hand.

## Design and Procedure

The research received approval from a research ethics committee. Participants were tested individually in a quiet room. Upon arrival, they were asked to sign a consent form; then they were connected to the facial EMG electrocardiogram (ECG) and EDA electrodes and were randomly assigned to one of the three instruction groups (feelings-focused [F-F], knowledge-focused [K-F], feelings-naïve [F-N]). Participants were seated approximately 60 cm from the computer monitor and were asked to sit without making extensive movements or touching their face. Participants were instructed according to the self-report instruction condition (F-F, K-F, or F-N) they were allocated to and completed a short (three trials) practice run. In the main experiment, 72 pictures were presented in random order. Each picture was presented for 6 seconds. After each picture, participants were asked to provide a self-report about the picture. After the rating scales, a black screen was presented for an average of 8.5 seconds. A hidden video camera recorded each participant's face during the EMG recording to remove movement artifacts. The participants were informed about the recording at the end of the experiment and were asked to provide their consent for using it. Upon refusal, the video recording was deleted.

## Preprocessing and Data Reduction

**Standardization.** Rescaling into a mean of zero and a standard deviation of 1 was performed on all measures (self-reports, facial EMG, HR, and EDA scores). Transformation was performed on the final statistics, that is, after the change score was computed (for physiological measures) or after self-report scales were combined—and done separately for each participant.

**Self-report.** We calculated a self-report bipolar valence score: the two unipolar self-report scores of pleasure and displeasure were converted into a single bipolar valence score (positive minus negative) for the purpose of fluency of reading the data analysis (Kron et al., 2015; Larsen et al., 2003). Results in terms of unipolar pleasure and displeasure scales are reported in SOM 2 and SOM 3 in online supplemental material.

**Facial EMG.** Preprocessing and quantification for change scores of the area above Zygomaticus Major (aaZM) and the area above Corrugator Supercilii (aaCS) are identical and are described here together.

***Preprocessing.*** Prior to preprocessing, artifact removal was done by inspecting the video recording. The following artifacts were removed (with experimenter blind to experimental conditions): yawning, lip licking and biting, scratching, and similar unrelated movements. Preprocessing was done with MATLAB R2014a (MathWorks Inc.). EMG signals were rectified by absolute value and fed into a 20–450 Hz Butterworth band-pass filtered (Butter, filtfilt, MATLAB).

***Signal quantification.*** EMG change score was computed as the mean activation (of zygomaticus or corrugator) during 6 seconds of picture presentation divided by the mean activation in the 2 seconds (baseline) prior to picture presentation. To ensure results are not specific to division by the baseline, results with measures that use subtraction (instead of division) from the baseline are presented in SOM 5 in online supplemental material.

**Heart rate.**

***Preprocessing.*** The heart rate was extracted from the ECG signal and processed offline using MindWare Technology's HRV 3.0.25 software. HR was filtered using a high-pass filter of 5 Hz. Artifacts were manually removed.

***Signal quantification.*** Heart rate score was computed as the mean heart rate following the first 4 seconds after stimulus onset subtracted from the mean heart rate in the 2 seconds prior to stimulus presentation.

Compatible with previous literature (Palomba et al., 1997), unlike the area above Corrugator Supercilli (aaCS) and the area above Zygomaticus Major (aaZM) that were modeled with a continuous valence scale, heart rate was compared between two binary conditions that quantify only positive and negative pictures. We ignored neutral stimuli in this analysis since cardiac deceleration in response to neutral stimuli is inconsistent, showing less deactivation than positive stimuli at times (Bradley et al., 2001), sometimes falling between positive and negative stimuli (Lang et al., 1993; Palomba et al., 1997), and in other cases showing no difference from positive stimuli (Itkes et al., 2017). In conclusion, the only consistent finding is that with the classification of valence axis to a categorical structure of positive, neutral, and negative values, pictures with negative valence value result in stronger cardiac deceleration than positive pictures. Classifying the continuous valence scale into a negative versus positive binary measure requires two cut-off points: a threshold under which a score is classified as negative, and a threshold above which a score is classified as positive. To make sure the results are not threshold-specific, heart rate analysis was performed for 4 thresholds: 0—scores below zero are negative and scores above zero are positive; −1, 1—scores below −1 are negative and scores above 1 are positive; −2, 2—scores below −2 are negative and scores above 2 are positive; and −3, 3—scores below −3 are negative and scores above 3 are positive.

**Electrodermal activation.**

***Preprocessing.*** EDA signal was processed with EDA analysis software (MindWare V3.01). Raw EDA was filtered (High pass: 1 Hz).

***Signal quantification.*** Skin conductance response was computed using two time windows; the first time window was used to define event-related responses and was set to signal change that occurs within three seconds starting one second after picture onset. The threshold of "signal change" was set at 0.01 mv. The second time window was used to quantify the Skin Conductance Response (SCR). SCR was quantified by subtracting SC minimum level from SC maximum level during an 8-second window starting from picture offset. To make sure results are not specific to the 8-second time window, additional analysis with a 6-second time window is now presented in SOM 5 in online supplemental material.

## Analytical Strategy and Statistical Analysis

The main purpose of this study was to compare the association between self-report evaluations and physiological response using the three instruction sets as different conditions. Since the context of this experiment is the IAPS pictures space that is mapped on the dimension of bipolar valence and arousal, the physiological measures that were chosen for the comparison between the instructions are correlated with either bipolar valence or arousal. Specifically, initial attenuation of HR and facial corrugator activation were previously found to be linearly related to bipolar valence scores (e.g., Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Greenwald, Cook, & Lang, 1989; Kron et al., 2013; Lang et al., 1993; Larsen et al., 2003), facial zygomaticus activation shows consistent quadratic relation to bipolar valence scores (e.g., Greenwald et al., 1989; Kron et al., 2013; Lang et al., 1993; Larsen et al., 2003) and SCR has a relation to arousal scores (Bradley et al., 2001; Greenwald et al., 1989; Kron et al., 2013; Lang et al., 1993).

To the aim of comparing the association between the three self-report evaluations and physiological response, we performed two types of analyses. The first analysis estimated the degree to which self-report data predicted the intensity of each of the signals (facial EMG, EDA, and HR). This analysis was done using linear mixed modeling (proc mixed, SAS) and had a similar model structure for all analyses: dependent variable—change score of physiological signal (corrugator EMG/zygomaticus EMG/heart rate/electrodermal activity); fixed effect—self-report scale and condition of instruction; random factors—participants and interaction of participants with rating scale; classification factor—participants and instruction conditions.

The second analysis examined the degree to which each self-report scale discriminated between binary measures of activation versus no activation. Discriminability was estimated by computing the area under the curve (AUC) of the receiver operating characteristic (ROC), an index taking into account sensitivity as function of false positive rates (Metz, 1978). To this aim, the continuous physiological signal was converted into binary measures reflecting activation versus no activation. To make sure that results are not dependent on a specific threshold used to convert the signal, four thresholds were computed (1, 0.75, 0.5, and 0.25 standard deviations above the mean activation). For example, in the case of 0.75 cut-off, intensity that was below 0.75 standard deviation above the mean was coded as "off" (no activation), while intensity above 0.75 standard deviation was coded as "on" (activation). Comparing areas under the curves of the different instruction conditions was estimated using chi square statistic (Gönen, 2007).

# Results

## General Reactivity

We first compared the average activation of each physiological channel among the three types of instructions. This initial analysis estimates potential effect of type of instruction on the physiological reactivity. Table 3 summarizes the mean and standard deviation of each physiological reactivity for each instruction condition. No differences between the three groups were found in Electrodermal Activity, $F(2, 7182) = 0.11$, *ns*; area above the corrugator, $F(2, 7139) = 1.57$, *ns*; area above the zygomaticus, $F(2, 6929) = 1.51$, *ns*; nor in heart rate, $F(2, 3656) = 0.18$, *ns.*

Table 3
*Means and Standard Deviations (in Parentheses) of Physiological Activity for Each Instruction Condition*

| Response channel | Feelings-Focused | Knowledge-Focused | Feelings-Naïve |
|---|---|---|---|
| Electrodermal activation | .032 (.125) | .033 (.134) | .039 (.172) |
| Area above corrugator | 1.089 (.819) | 1.083 (.473) | 1.155 (.708) |
| Area above zygomaticus | 1.313 (1.521) | 1.171 (.908) | 1.389 (2.306) |
| Heart rate | −1.34 (12.099) | −.66 (14.305) | −1.93 (14.210) |

## Area Above Zygomaticus Major (aaZM)—EMG Intensity Prediction

**Quadratic relationship.**[1] First, we compared the quadratic relationship between bipolar valence (Pos–Neg) from the three types of instructions with aaZM EMG (remember that aaZM is the mean activation of zygomaticus during 6 seconds of picture presentation divided by mean activation in the 2 seconds before picture presentation; confidence intervals of the following contrasts are depicted in Figure 2.1). Feelings-focused bipolar valence outperformed knowledge-focused bipolar valence in quadratic prediction of aaZM EMG intensity $F(1, 4488) = 26.48$, $p < .0001$. The difference between feelings-naïve bipolar valence and knowledge-focused bipolar valence was not significant $F(1, 4293) = 2.16$, *ns*. Feelings-focused bipolar valence outperformed feelings-naïve bipolar valence in predicting aaZM EMG intensity $F(1, 4547) = 9.80$, $p < .0018$.

Next, we estimated the simple effects of the quadratic relationship between each of the three bipolar valence (Pos–Neg) scales and zygomaticus activation (confidence intervals for the following contrasts are depicted in Figure 2.2). Feelings-focused bipolar valence was positively quadratically associated with zygomaticus intensity, $F(1, 34) = 67.75$, $p < .0001$; knowledge-focused bipolar valence was positively quadratically associated with zygomaticus intensity, $F(1, 30) = 6.48$, $p < .0163$; feelings-naïve bipolar valence was positively quadratically associated with zygomaticus intensity, $F(1, 31) = 13.07$, $p < .0011$.

**Linear relationship.** Here we compared the linear relationship between bipolar valence (Pos–Neg) of the three types of instructions with aaZM EMG intensity (confidence intervals for the following contrasts are depicted in Figure 2.3). Feelings-focused bipolar valence outperformed knowledge-focused bipolar valence in linear prediction of aaZM EMG intensity $F(1, 4554) = 5.16$, $p < .02$. The difference between feelings-naïve bipolar valence and knowledge-focused bipolar valence was not significant $F(1, 4356) = 3.11$, *ns.* Feelings-focused bipolar valence was not significantly different from feelings-naïve bipolar valence in predicting aaZM EMG intensity, $F(1, 4614) = 2.10$, *ns.*

Next, we estimated the simple effects of the linear relationships between each of the bipolar valence (Pos–Neg) types and aaZM (confidence intervals for the following contrasts are depicted in Figure 2.4). Feelings-focused bipolar valence was positively linearly associated with zygomaticus intensity, $F(1, 34) = 26.10$, $p < .0001$; knowledge-focused bipolar valence was positively linearly associated with zygomaticus intensity, $F(1, 30) = 10.59$, $p < .0028$; feelings-naïve bipolar valence was positively linearly associated with zygomaticus intensity, $F(1, 31) = 9.43$, $p < .0.0044$.

## Area Above Zygomaticus Major (aaZM)—EMG Discriminability Analysis

As mentioned above, a discriminability analysis involves determining the threshold on which zygomaticus' continuous EMG activation is converted into a binary measure of activation versus no activation (see Method, Analytical Strategy and Statistical Analysis sections). To make sure that the results are not dependent on a specific threshold that was used to convert zygomaticus activation into a binary on–off measure, four thresholds were computed (1, 0.75, 05, and 0.25 standard deviations above the mean activation). Each of the dark gray/light gray/black dots in Figure 3.1 is an area under the curve (AUC) for a specific threshold.

AUC of all instruction sets in all cut-offs showed discriminability higher than chance (see confidence intervals in Figure 3.1). Feelings-focused bipolar valence showed significantly higher discriminability between zygomaticus activation versus no-activation than feelings-knowledge bipolar valence in all cut-offs: 0.25 *SD*—$x^2(1) = 12.9$, $p < 0.0003$; 0.5 *SD*—$x^2(1) = 11.5$, $p < 0.0006$; 0.75 *SD*—$x^2(1) = 6.9$, $p < 0.008$; 1 *SD*—$x^2(1) = 7.3$, $p < 0.006$. Feelings-focused bipolar valence showed higher discriminability than feelings-naïve bipolar valence—the difference was significant in two cut-offs: 0.25 *SD*—$x^2(1) = 6.5$, $p < 0.01$; 0.5 *SD*—$x^2(1) = 4.1$, $p <$ *SD*—$x^2(1) = 1.1$ *ns*; 1 *SD*—$x^2(1) = 1.3$ *ns*; feelings-naïve bipolar valence showed higher discriminability than knowledge-focused bipolar valence but the differences were not significant. .25 *SD*—$x^2(1) = 1.7$ *ns*; 0.5 *SD*—$x^2(1) = 2$ *ns*; 0.75 *SD*—$x^2(1) = 2.4$ *ns*; 1 *SD*—$x^2(1) = 2.2$ *ns*.

## Area Above Corrugator Supercilii (aaCS)—EMG Intensity Prediction

We first compared the linear relationship between bipolar valence (Pos–Neg) of the three types of instructions with aaCS EMG intensity (remember that aaCS mean activation of corrugator during 6 seconds of picture presentation divided by mean activation in the 2 seconds before picture presentation; confidence intervals of the following contrasts are depicted in Figure 4.1). Feelings-focused bipolar valence outperformed knowledge-focused bipolar valence in predicting aaCS intensity, $F(1, 4761) = 14.88$, $p < .0001$**;** feelings-naïve bipolar valence outperformed knowledge-focused bipolar valence in predicting aaCS intensity, $F(1, 4563) = 4.19$, $p < .04$. We found no significant difference between feelings-focused bipolar valence and feelings-naïve bipolar valence $F(1, 4614) = 2.87$, $p < .09$.

Next, we estimated the simple effects of the linear relationship between bipolar valence (Pos–Neg) scale of each of the self-report type and aaCS (confidence intervals of the simple effects are depicted in Figure 4.2). Feelings-focused bipolar valence was positively linearly associated with aaCS intensity, $F(1, 34) = 71.28$, $p < .0001$; knowledge-focused bipolar valence was positively linearly associated with aaCS intensity, $F(1, 33) = 29.33$, $p < .0001$; feelings-naïve bipolar valence was positively linearly associated with aaCS intensity, $F(1, 31) = 41.73$, $p < .0.0001$.

## Discriminability of Area Above Corrugator Supercilii (aaCS) Activation

Estimating discriminability of aaCS activation was examined separately for the attenuation of activation in aaCS (usually characterizing a reaction to positive stimuli) and for the increase in aaCS activation (characterizing a reaction to negative stimuli; Kron et al., 2013; Lang et al., 1993; Larsen et al., 2003). In aaCS

[1] Note that estimation of quadratic relationship is done only for zygomaticus activity. In this type of design (i.e., when stimuli are selected to cover the entire IAPS space), zygomaticus activity (but not corrugator) has a strong quadratic component. (e.g., Kron, Pilkiw, Banaei, Goldstein, & Anderson, 2015; Larsen, Norris, & Cacioppo, 2003).

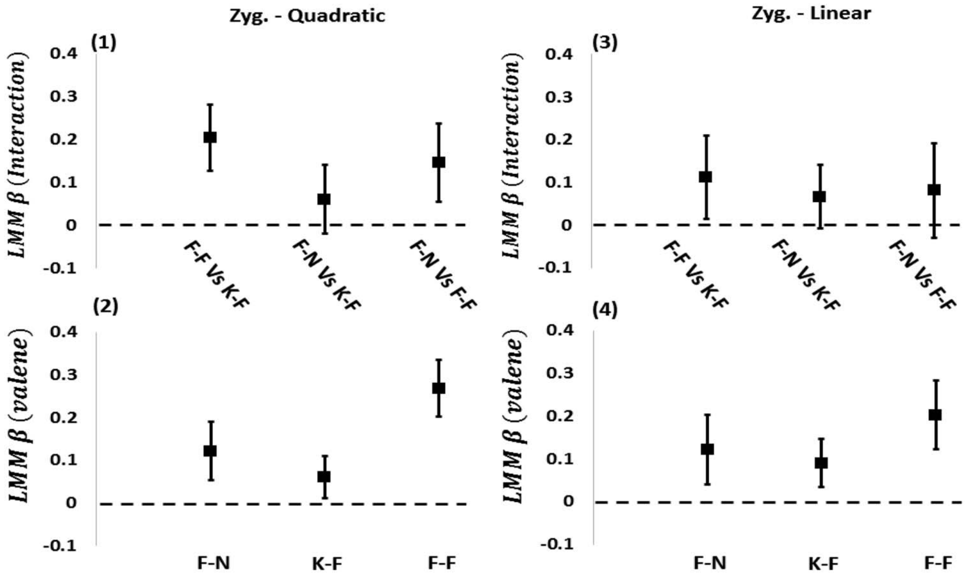


*Figure 2.* F-F–Feelings-Focused, K-F–Knowledge-Focused, F-N–Feelings-Naïve, LMM–Linear Mixed Model, aaZM-area above Zygomaticus Major. (2.1) Confidence intervals (á = 0.05) for linear mixed model's coefficients of interaction between two quadratic components. (2.2) Confidence intervals (á = 0.05) for linear mixed model coefficients of quadratic association between valence and aaZM intensity. (2.3) Confidence intervals (á = 0.05) for linear mixed model interaction coefficients between two linear components. (2.4) Confidence intervals (á = 0.05) for linear association coefficients between valence and aaZM intensity.

attenuation, the continuous scores were converted into binary scores of aaCS attenuation versus both activation and nonreaction. In aaCS increase, the continuous scores were converted into binary scores of aaCS increase versus both decrease and no-difference.

**Area above corrugator supercilii (aaCS)—Attenuation.** AUC of all bipolar valence (Pos—Neg) sets in all cut-offs showed discriminability higher than chance (see confidence intervals in Figure 5.1). Feelings-focused bipolar valence showed significantly higher discriminability between aaCS activation versus no-activation than Knowledge-focused bipolar valence for all thresholds: 0.25 *SD*—$x^2(1) = 27.2$, $p < 0.00001$; 0.5 *SD*—$x^2(1) = 29.4$, $p < 0.00001$; 0.75 *SD*—$x^2(1) = 36.5$, $p < 0.00001$; 1 *SD*—$x^2(1) = 30.9$, $p < 0.00001$. Feelings-focused bipolar valence showed significantly higher discriminability than feelings-naïve bipolar valence across all thresholds except 0.25 SD: 0.25 *SD*—$x^2(1) = 3.6$, $p < 0.057$; 0.5 *SD*—$x^2(1) = 7.7$, $p < 0.005$; 0.75 *SD*—$x^2(1) = 14$, $p < 0.0001$**;** 1 *SD*—$x^2(1) = 15.6$, $p < 0.0001$; feelings-naïve bipolar valence showed higher discriminability than

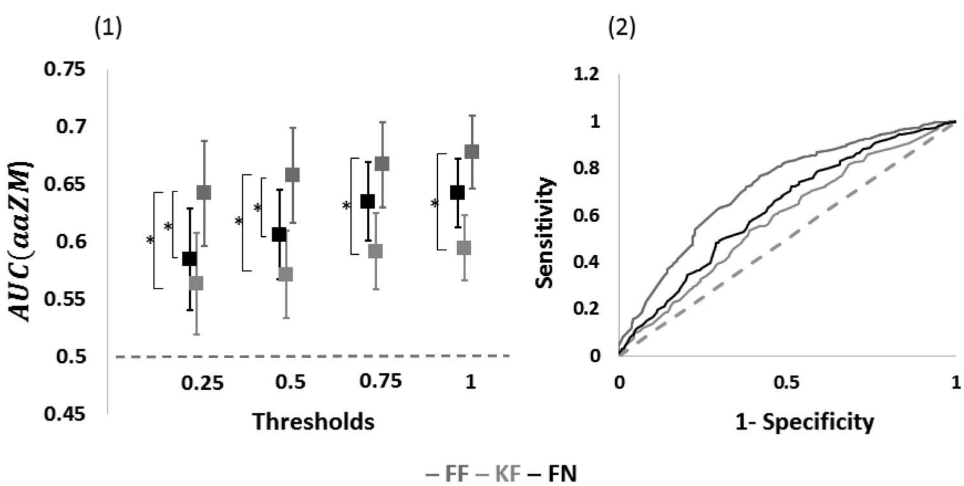


*Figure 3.* F-F–Feelings-Focused, K-F–Knowledge-Focused, F-N–Feelings-Naïve, AUC–Area Under the Curve. (3.1) Squares are areas under the curve with confidence intervals (á = 0.05). The horizontal axis is thresholds used to divide zygomaticus activation into active versus nonactive. (3.2) A graphic description of three Receiver Operating Characteristic (ROC) curves for cut-off of 0.5 standard deviations.

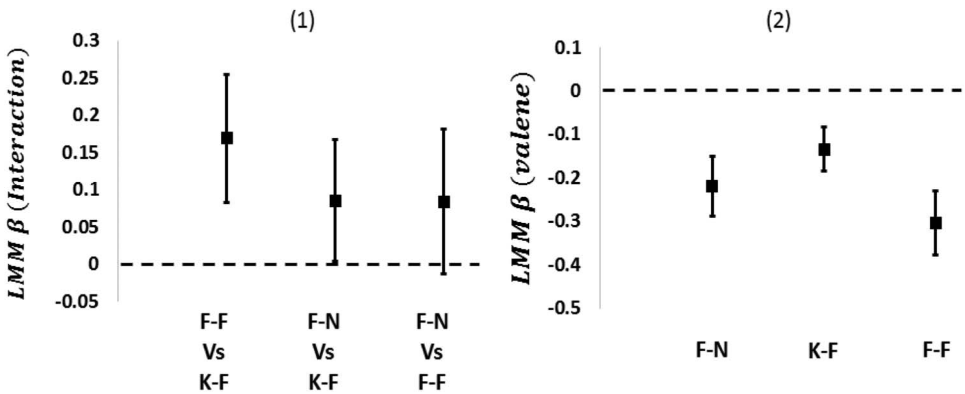


*Figure 4.* F-F–Feelings-Focused, K-F–Knowledge-Focused, F-N–Feelings-Naïve, LMM–Linear Mixed Model, aaCS-area above the Corrugator Supercilii. (4.1) Confidence intervals (á = 0.05) for linear mixed model coefficients of interaction between two bipolar valence scales predicting activation in aaCS. (4.2) Confidence intervals (á = 0.05) for linear mixed model coefficients of linear association between valence and aaCS intensity.

knowledge-focused bipolar valence in all cut-offs except −1 standard deviations. 0.25 *SD*—$x^2(1) = 10.9$, $p < 0.001$; 0.5 *SD*—$x^2(1) = 6.1$, $p < 0.01$; 0.75 *SD*—$x^2(1) = 5$, $p < 0.02$; 1 *SD*—$x^2(1) = 2.5$, *ns*.

**Area above corrugator supercilii (aaCS)—Increase.** AUC of all bipolar valence sets in all cut-offs showed discriminability higher than chance (see confidence intervals in Figure 5.3). Feelings-focused bipolar valence showed significantly higher discriminability between aaCS activation versus no-activation than knowledge-focused bipolar valence for all thresholds: 0.25 *SD*—$x^2(1) = 14.6$, $p < 0.0001$; 0.5 *SD*—$x^2(1) = 11.8$, $p < 0.0005$; 0.75 *SD*—$x^2(1) = 9.5$, $p < 0.002$**;** 1 *SD*—$x^2(1) = 6.5$, $p < 0.01$; feelings-naïve bipolar valence showed higher discriminability than knowledge-focused bipolar valence to all thresholds: 0.25 *SD*—$x^2(1) = 7.4$, $p < 0.006$; 0.5 *SD*—$x^2(1) = 4.8$, $p < 0.02$; 0.75 *SD*—$x^2(1) = 7.3$, $p < 0.006$**;** 1 *SD*—$x^2(1) = 4.1$, $p < 0.04$. No

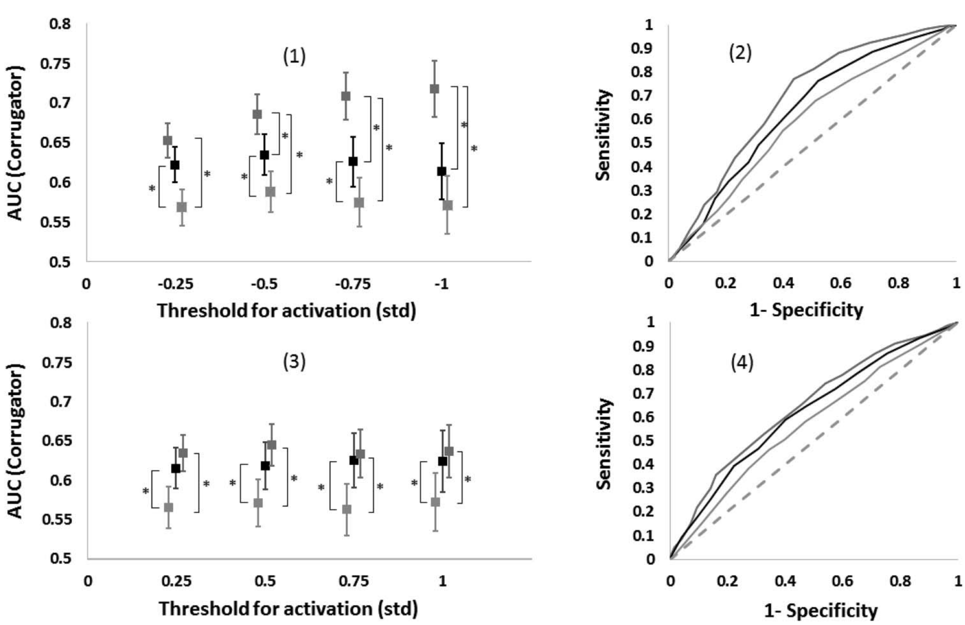


*Figure 5.* F-F–Feelings-Focused, K-F–Knowledge-Focused, F-N–Feelings-Naïve, AUC–Area under the Receiver Operating Characteristic (ROC) curve. (5.1) Area above Corrugator Supercilii (aaCS)–attenuation. Squares represent areas under the curves and are surrounded by confidence intervals (á = 0.05). The horizontal axis shows the thresholds used to divide corrugator activation into attenuation versus no attenuation. (5.2) A graphic description of three ROC curves for cut-off of −0.5 standard deviations. (5.3) Area above Corrugator Supercilii (aaCS)–increase. Squares represent areas under the curve and are surrounded with confidence intervals (á = 0.05). The horizontal axis shows the thresholds used to divide corrugator activation into increase versus no increase. (5.4) A graphic description of three ROC curves for cut-off of 0.5 standard deviations.

difference was found between feelings-focused bipolar valence and feelings-naïve bipolar valence.

## Heart Rate

Heart rate was compared between two binary conditions that quantify only positive and negative valence and ignore neutral (see Preprocessing and Data Reduction section). As can be clearly seen in Figure 6, the differences between positive and negative valence were significant for feelings-focused reports across all four cut-offs: for cut-off of 0 (i.e., negative defined as bipolar valence scores lower than 0 and positive for scores higher than 0), $t(1511) = 3$, $p < .002$; for cut-offs of $-1$,1 (i.e., negative defined as bipolar valence scores lower than $-1$ and positive for scores higher than 1), $t(1307) = 3.35$, $p < .0008$; for cut-offs of $-2$,2 (i.e., negative defined as bipolar valence scores lower than $-2$ and positive for scores higher than 2), $t(1075) = 2.9$, $p < .004$; for cut-offs of $-3$,3 (i.e., negative defined as bipolar valence scores lower than $-3$ and positive for scores higher than 3), $t(846) = 2.8$, $p < .005$. The difference between positive and negative valence was significant for feelings-naïve for three of the four cut-offs: for cut-offs of 0, $t(1872) = 1.61$, *ns*; for cut-offs of $-1$,1, $t(1686) = 2$, $p < .04$; for cut-offs of $-2$,2, $t(1465) = 2.82$, $p < .004$; for cut-offs of $-3$,3, $t(1236) = 2.43$, $p < .01$. No significant difference between positive and negative valence was found for knowledge-focused instructions. No significant differences were found between feelings-focused, knowledge-focused, and feelings-naïve instructions for any threshold.

## Electrodermal (EDA) Intensity Prediction

We first compared the linear relationship between arousal (Pos + Neg) of the three types of instructions with EDA intensity. Feelings-focused arousal outperformed knowledge-focused arousal in predicting EDA intensity, $F(1, 4755) = 5.34$, $p < .02$, See Figure 7.1 for confidence intervals. The difference between feelings-naïve arousal and knowledge-focused arousal in predicting EDA intensity was significant, $F(1, 4612) = 4.4$, $p < .03$. We found no significant difference between feelings-focused arousal and feelings-naïve arousal $F(1, 4683) = 0.11$.

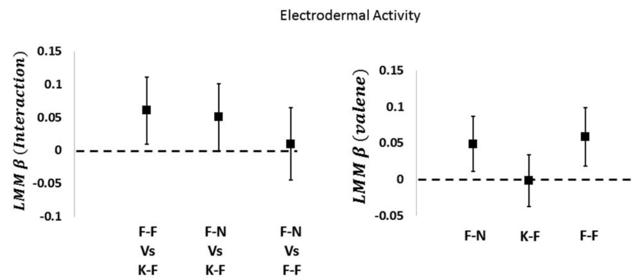


*Figure 7.* F-F–Feelings-Focused, K-F–Knowledge-Focused, F-N–Feelings-Naïve, LMM–Linear Mixed Model. (7.1) Confidence intervals (á = 0.05) for linear mixed models interaction coefficients. (7.2) Confidence intervals (á = 0.05) for linear association coefficients between valence and EDA intensity.

Next, we estimated the simple effects of linear relationships between each of the arousal (Pos + Neg) of self-report instruction sets and EDA, see Figure 7.2 for confidence intervals. The linear association between arousal and EDA activation was significant for feelings-focused reports, $F(1, 2413) = 9.66$, $p < .002$, and for feelings-naïve reports, $F(1, 2270) = 8.77$, $p < .0.003$, see Figure 7.2 for confidence interval. Knowledge-focused arousal reports were not significantly associated with EDA $F(1, 2342) = 0.01$, *ns*.

## Discriminability of Electrodermal Activation

Figure 8 presents confidence intervals for the three arousal (Pos + Neg) instructions sets in all four binary EDA activation cut-offs (activation vs. nonactivation). Feelings-focused arousal showed significant above-chance discriminability in all cut-offs;

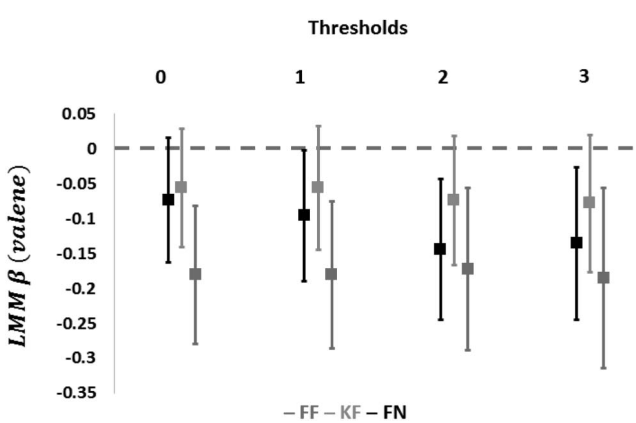


*Figure 6.* F-F–Feelings-Focused, K-F–Knowledge-Focused, F-N–Feelings-Naïve, *LMM* â (*valence*)—linear mixed model coefficient for binary valence index of HR. Squares represent LMM coefficients for the binary valence index and are surrounded by confidence intervals (á = 0.05). The horizontal axis is the thresholds used to divide the continuous valence scale into binary measures.

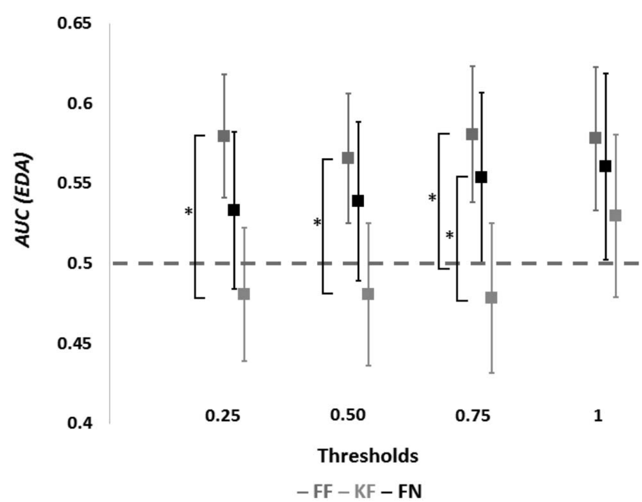


*Figure 8.* F-F–Feelings-Focused, K-F–Knowledge-Focused, F-N–Feelings-Naïve, AUC Area under the Receiver Operating Characteristic (ROC) curve. Discriminability of electrodermal activation (EDA). Squares represent areas under the curve and are surrounded by confidence intervals (á = 0.05). The horizontal axis contains the thresholds used to divide EDA into activation versus no activation.

feelings-naïve arousal showed significant above-chance discriminability in two cut-offs (1 *SD*, 0.75 *SD*); knowledge-focused arousal showed no significant above-chance discriminability in all cut-offs. Significant differences were found between ROC curves of feelings-focused and knowledge-focused in three cut-offs (0.25 *SD*, 0.5 *SD*, and 0.75 *SD*): 0.25 *SD*—$x^2(1) = 11.7$, $p < 0.0006$; 0.5 *SD*—$x^2(1) = 9.2$, $p < 0.002$; 0.75 *SD*—$x^2(1) = 11$, $p < 0.0009$; 1 *SD*—$x^2(1) = 1.9$, *ns*. Feelings-naïve arousal outperformed knowledge-focused in only one cut-off (0.75 *SD*), $x^2(1) = 4.1$, $p < 0.04$. No significant difference was found between feelings-focused and feelings-naïve arousal scales.

## Discussion

The aim of this study was to examine two research questions that stem from the distinction between affective and semantic representations of valence. First, we examined whether feelings-focused reports are more in tune with facial and autonomic affective reactions than knowledge-focused reports. Second, we examined whether feelings-focused reporting has an advantage over feelings-naïve reporting in modeling facial and autonomic responses. To this end, three types of instructions were examined: feelings-focused instructions, centered around reporting affective feelings; knowledge-focused instructions, centered around reporting semantic knowledge; and feelings-naïve instructions, in which participants received no special instructions other than rating their feelings. The relationship between self-reports and both physiological signal intensity and signal discriminability were examined.

In terms of signal intensity, feelings-focused instructions had a clear advantage over knowledge-focused instructions. Specifically, feelings-focused instructions showed significantly better performance in predicting facial muscle activation in areas above the zygomaticus and corrugator as well as electrodermal activation. However, the difference between feelings-focused and knowledge-focused instructions was not significant in the case of heart rate. Yet, feelings-focused instructions showed a significant relationship with heart rate deceleration while knowledge-focused instructions did not.

For the discrimination analysis, the area under the ROC curve provides an index of how informative the self-report score is in discriminating between "activation" and "no activation" of the physiological signal. The area under the curve takes both sensitivity and false positive rates under different decision rules into account. Overall, the accuracy of self-reports in discriminating physiological signal activation was low and ranged from *poor* to *fair* for all sets of instructions. This finding is not surprising given the low reliability of physiological signals and, consequently, the low coherence usually found across different emotional response systems, in particular physiological and experiential measures (Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). Importantly, our discriminability analysis suggests a significant advantage for feelings-focused instructions over knowledge-focused instructions in predicting facial EMG response and electrodermal activation.

The advantage of feelings-focused instructions over knowledge-focused instructions in predicting both signal intensity and activation status supports the theoretical distinction between semantic and affective representations of pleasure and displeasure. This distinction suggests that "being positive/negative" can be features of the affective response—"affective valence," or knowledge about the stimulus—"semantic valence." The working assumption of this study was that affective response operates on multiple response channels that work together. These include autonomic responses (heart rate deceleration or electrodermal change), facial expressions, and feelings; all traditionally assumed to be response channels of the affective response. Semantic knowledge, on the other hand, is assumed to be a different representation type that is not part of the affective response. The fact that feelings-focused reports better predicted facial and autonomic measures than knowledge-focused reports supports this assumption and the theoretical distinction between semantic and affective valence. These results are consistent with previous work in which we showed that repeated exposure to a stimulus results in the attenuation of measures that reflect an affective response (feelings-focused instructions, autonomic measures, and facial expressions), but does not affect measures that relate to semantic valence such as knowledge-focused reports (Itkes et al., 2017).

This study serves a larger-scale scientific endeavor which aims to characterize the differences between affective and semantic representations of affect and emotion. These include the theoretical distinction between "affective quality" and "core affect" (Russell, 2003), experiential and nonexperiential information of self-reports (Robinson & Clore, 2002a), cold versus hot emotional processes (Schaefer et al., 2003), and "affective gut reaction" to "propositional belief" (Gawronski & Bodenhausen, 2011). This endeavor is important, since without an explicit distinction between affective and semantic representations, experimental results might be difficult to interpret. Today, many tasks that are interpreted as affective actually involve semantic valence (Itkes et al., 2017). The results of the current study not only provide support for the theoretical distinction between affective and semantic valence, but also for the validity of the feelings-focused and knowledge-focused instructions. The advantage of feelings-focused over knowledge-focused in predicting physiological signals suggests that, with the right instructions, participants can meaningfully distinguish and consistently report about affective and semantic aspects of valence.

The feelings-naïve condition, in which participants are merely instructed to rate their feelings without the support of a procedure that distinguishes between feelings and semantic knowledge, enabled us to evaluate the relative contribution of feelings-focused instructions beyond the more "traditional" instruction routine. Given the consistent results across EDA, HR, and facial EMG, it is safe to conclude that the performance of feelings-naïve instructions falls in between those of feelings-focused and knowledge-focused. The difference between feelings-focused and feelings-naïve is not as strong as the difference between feelings-focused and knowledge-focused; and in the same vein, the difference between feelings-naïve and knowledge-focused is not as clear as the difference between feelings-focused and knowledge-focused instructions.

One limitation of the current study is the sample size that consisted of 105 participants; 35 in each instruction condition. In terms of comparing association between self-reports and physiological activity between the three conditions, such a sample size might result in an underpowered design. In this regard, we suggested taking a more cautious approach and interpreted the results with a bird's-eye view: that is, to give more weight to the general trend that emerges beyond the different physiological measures (better prediction ability to feelings-focused reports than knowledge focused reports with feelings-naïve falling in between) and less weight to the difference between the different physiological measures (differences between zygomaticus EMG signals and HR). A second limitation of this study is that comparison between instructions was done in a very specific context—passive viewing of still pictures and modeling the response on dimensional space (i.e., bivariate pleasure and displeasure). It will be important to examine the generalization of these results in other contexts, such as self-relevant stimuli, and/or stimuli that elicit discrete emotions (e.g., sadness, disgust).

Self-reports of valence are critical in emotion research. They serve as a measure of affective feelings (Barrett et al., 2007), as means to allocate stimuli to experimental conditions by standardized norms (Lang et al., 1997), and as means to model physiological (e.g., Lang et al., 1993) and neural (e.g., Phan et al., 2003) signals. Feelings-focused instructions emphasize two aspects that may improve the quality of reports about actual feelings rather than semantic knowledge. First, they educate participants regarding the distinction between reporting about affective feelings and reporting about semantic knowledge. Second, feelings-focused instructions legitimize providing low ratings for stimuli with high semantic valence value and, in so doing, potentially reducing the effect of social desirability biases (Crowne & Marlowe, 1960). It will be beneficial to the scientific community to examine if the advantage of the feelings-focused instruction set is more relevant to stimuli with relatively low affective intensity but high semantic value, such as words or pictorial stimuli. With such stimuli, the experience of feelings is vague and sometimes even entirely absent and consequently, without the adequate instructions, participants may report semantic knowledge rather than actual feelings (Levenson, 2003; Robinson & Clore, 2002a).

Finally, in this work we assume that, in the knowledge-focused instructions, participants answer self-reports by providing semantic information. That is, that when providing knowledge-focused value on the pleasant and unpleasant scales, they rely on general, impersonal, stored knowledge about the valence of objects. However this design cannot rule out that, although participants evaluate impersonal events (i.e., the IAPS picture), they still use episodic information—for example, relying on similar events that happened to them at a specific time and place. It will be fruitful to disentangle these two types of knowledge in future research and to examine if episodic and semantic components behave in the same way.

## References


Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338,* 1225–1229. http://dx.doi.org/10.1126/science.1224313

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality, 40,* 35–55. http://dx.doi.org/10.1016/j.jrp.2005.08.006

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology, 58,* 373–403. http://dx.doi.org/10.1146/annurev.psych.58.110405.085709

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science, 8,* 10–14. http://dx.doi.org/10.1111/1467-8721.00003

Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. *Cognition and Emotion, 19,* 847–878. http://dx.doi.org/10.1080/02699930541000057

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1,* 276–298. http://dx.doi.org/10.1037/1528-3542.1.3.276

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25,* 49–59. http://dx.doi.org/10.1016/0005-7916(94)90063-9

Bradley, M. M., Lang, P. J., & Cuthbert, B. N. (1993). Emotion, novelty, and the startle reflex: Habituation in humans. *Behavioral Neuroscience, 107,* 970–980. http://dx.doi.org/10.1037/0735-7044.107.6.970

Cole, P. M. (1986). Children's spontaneous control of facial expression. *Child Development, 57,* 1309–1321. http://dx.doi.org/10.2307/1130411

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349–354. http://dx.doi.org/10.1037/h0047358

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology, 52,* 95–111. http://dx.doi.org/10.1016/S0301-0511(99)00044-7

Damasio, A. R. (1999). The feeling of what happens: Body and emotion in the making of consciousness. *New York Times Book Review, 104,* 8–8.

Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology, 69,* 851–864. http://dx.doi.org/10.1037/0022-3514.69.5.851

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science, 298,* 1191–1194. http://dx.doi.org/10.1126/science.1076358

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384–392. http://dx.doi.org/10.1037/0003-066X.48.4.384

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23,* 567–589. http://dx.doi.org/10.1111/j.1469-8986.1986.tb00676.x

Frijda, N. (2005). Emotion experience. *Cognition and Emotion, 19,* 473–497. http://dx.doi.org/10.1080/02699930441000346

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364,* 3453–3458. http://dx.doi.org/10.1098/rstb.2009.0142

Furedy, J. J. (1968). Novelty and the measurement of the GSR. *Journal of Experimental Psychology, 76,* 501–503.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59–127. http://dx.doi.org/10.1016/B978-0-12-385522-0.00002-0

Gerber, A. J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., . . . Peterson, B. S. (2008). An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces. *Neuropsychologia, 46,* 2129–2139. http://dx.doi.org/10.1016/j.neuropsychologia.2008.02.032

Gollnisch, G., & Averill, J. R. (1993). Emotional imagery: Strategies and correlates. *Cognition and Emotion, 7,* 407–429. http://dx.doi.org/10.1080/02699939308409196

Gönen, M. (2007). *Analyzing receiver operating characteristic curves with SAS.* Cary, NC: SAS Institute.

Greenwald, M. K., Cook, E. W., & Lang, P. J. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology, 3,* 51–64.

Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A., & Kron, A. (2017). Dissociating affective and semantic valence. *Journal of Experimental Psychology: General, 146,* 924–942. http://dx.doi.org/10.1037/xge0000291

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality and Social Psychology Bulletin, 24,* 855–879. http://dx.doi.org/10.1177/0146167298248006

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 85–105). New York, NY: Russell Sage Foundation.

Karmon-Presser, A., Sheppes, G., & Meiran, N. (2018). How does it "feel"? A signal detection approach to feeling generation. *Emotion, 18,* 94–115. http://dx.doi.org/10.1037/emo0000298

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84,* 394–421. http://dx.doi.org/10.1016/j.biopsycho.2010.03.010

Kron, A., Goldstein, A., Lee, D. H. J., Gardhouse, K., & Anderson, A. K. (2013). How are you feeling? Revisiting the quantification of emotional qualia. *Psychological Science, 24,* 1503–1511. http://dx.doi.org/10.1177/0956797613475456

Kron, A., Pilkiw, M., Banaei, J., Goldstein, A., & Anderson, A. K. (2015). Are valence and arousal separable in emotional experience? *Emotion, 15,* 35–44. http://dx.doi.org/10.1037/a0038474

Kron, A., Pilkiw, M., Goldstein, A., Lee, D. H., Gardhouse, K., & Anderson, A. K. (2014). Spending one's time: The hedonic principle in ad libitum viewing of pictures. *Emotion, 14,* 1087–1101. http://dx.doi.org/10.1037/a0037696

Lacey, B. C., & Lacey, J. I. (1978). Two-way communication between the heart and the brain: Significance of time within the cardiac cycle. *American Psychologist, 33,* 99–113. http://dx.doi.org/10.1037/0003-066X.33.2.99

Lambie, J. A., & Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review, 109,* 219–259. http://dx.doi.org/10.1037/0033-295X.109.2.219

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical manual and affective ratings* (Report No. 8-A). Gainesville, FL: Center for the Study of Emotion and Attention.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30,* 261–273. http://dx.doi.org/10.1111/j.1469-8986.1993.tb03352.x

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology, 40,* 776–785. http://dx.doi.org/10.1111/1469-8986.00078

Levenson, R. W. (2003). Autonomic specificity and emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in affective science: Handbook of affective sciences* (pp. 212–224). New York, NY: Oxford University Press.

Maltzman, I., & Boyd, G. (1984). Stimulus significance and bilateral SCRs to potentially phobic pictures. *Journal of Abnormal Psychology, 93,* 41–46. http://dx.doi.org/10.1037/0021-843X.93.1.41

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5,* 175–190. http://dx.doi.org/10.1037/1528-3542.5.2.175

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge, MA: MIT Press.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8,* 283–298.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A. L., . . . Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods, 45,* 169–177. http://dx.doi.org/10.3758/s13428-012-0243-8

Müri, R. M. (2016). Cortical control of facial expression. *The Journal of Comparative Neurology, 524,* 1578–1585. http://dx.doi.org/10.1002/cne.23908

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin, 49,* 197–237. http://dx.doi.org/10.1037/h0055737

Palomba, D., Angrilli, A., & Mini, A. (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *International Journal of Psychophysiology, 27,* 55–67. http://dx.doi.org/10.1016/S0167-8760(97)00751-4

Phan, K. L., Taylor, S. F., Welsh, R. C., Decker, L. R., Noll, D. C., Nichols, T. E., . . . Liberzon, I. (2003). Activation of the medial prefrontal cortex and extended amygdala by individual ratings of emotional arousal: A fMRI study. *Biological Psychiatry, 53,* 211–215. http://dx.doi.org/10.1016/S0006-3223(02)01485-3

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95,* 52–77. http://dx.doi.org/10.1037/0033-2909.95.1.52

Robinson, M. D., & Clore, G. L. (2002a). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128,* 934–960. http://dx.doi.org/10.1037/0033-2909.128.6.934

Robinson, M. D., & Clore, G. L. (2002b). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology, 83,* 198–215. http://dx.doi.org/10.1037/0022-3514.83.1.198

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. http://dx.doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172. http://dx.doi.org/10.1037/0033-295X.110.1.145

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57,* 493–502. http://dx.doi.org/10.1037/0022-3514.57.3.493

Schacter, D. L., Wagner, A. D., & Buckner, R. L. (2000). *Memory systems of 1999*. New York, NY: Oxford University Press.

Schaefer, A., Collette, F., Philippot, P., van der Linden, M., Laureys, S., Delfiore, G., . . . Salmon, E. (2003). Neural correlates of "hot" and "cold" emotional processing: A multilevel approach to the functional anatomy of emotion. *NeuroImage, 18,* 938–949. http://dx.doi.org/10.1016/S1053-8119(03)00009-0

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh, PA: Psychology Software.

Shields, S. A. (1987). Women, men, and the dilemma of emotions. In P. Shaver & C. Hendrick (Eds.), *Review of personality and social psychology: Vol. 7. Sex and gender* (pp. 229–250). Newbury Park, CA: Sage.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience, 11,* 843–850. http://dx.doi.org/10.1038/nn.2138

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168,* 242–249. http://dx.doi.org/10.1016/j.psychres.2008.05.006

Tulving, E. (1984). Précis of elements of episodic memory. *Behavioral and Brain Sciences, 7,* 223–238. http://dx.doi.org/10.1017/S0140525X0004440X

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science, 2,* 67–70. http://dx.doi.org/10.1111/1467-8721.ep10770899

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. Retrieved from https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1011&context=psychology_pubs

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin, 121,* 331–354. http://dx.doi.org/10.1037/0033-2909.121.3.331

Widiger, T. A., & Settle, S. A. (1987). Broverman et al. revisited: An artifactual sex bias. *Journal of Personality and Social Psychology, 53,* 463–469. http://dx.doi.org/10.1037/0022-3514.53.3.463

Williams, J. A. (1963). Novelty, GSR, and stimulus generalization. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 17,* 52–61. http://dx.doi.org/10.1037/h0083270

Williams, J. E., & Bennett, S. M. (1975). The definition of sex stereotypes via the adjective check list. *Sex Roles, 1,* 327–337. http://dx.doi.org/10.1007/BF00287224

Received May 17, 2018
Revision received November 29, 2018
Accepted November 30, 2018 ■