Valence and Arousal: A Comparison of Two Sets of Emotional Facial Expressions

DIRK ADOLPH
University of Würzburg, University of Düsseldorf

GEORG W. ALPERS
University of Würzburg, University of Eichstätt, University of Mannheim

Facial expressions are often used in emotion research. Although they may differ in several relevant features such as the intensity of the facial expressions, the picture sets have not been compared systematically. Because the intensity of expressions is thought to determine the level of emotional arousal induced by the stimulus, the first aim of this study was to test whether 2 frequently used sets of emotional facial expressions induce different levels of perceived arousal. Furthermore, we tested whether the sex of the actor modulates arousal ratings. Participants viewed facial expressions from the NimStim set (more intense expressions) and the Karolinska Directed Emotional Faces set (less intense expressions). Female expressions from the Karolinska Directed Emotional Faces but male expressions from the NimStim set were rated as more emotionally arousing. We conclude that less intense female expressions but more intense male expressions may be more potent in inducing emotional responses. This study may encourage researchers to further compare the properties of picture sets.

The basic emotions happiness, surprise, anger, fear, disgust, and sadness are reflected in universal facial expressions (Ekman & Friesen, 1971). These expressions can be decoded largely independent of cultural background, and they induce distinct emotional states in the viewer (Ekman & Davidson, 1994; Izard, 1992). Therefore, in emotion research, visual stimuli of emotional facial expressions are often preferred over more complex pictures such as those from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2005) because they are more physically uniform across emotion categories (Alpers & Gerdes, 2007). Emotional facial expressions are often presented in research on emotion recognition (for an overview, see Adolphs, 2002). They are also used often to induce specific emotions in research on facial mimicry and emotion contagion (i.e., the tendency to experience and express another person's emotion) (Hess & Blairy, 2001; Lundqvist & Dimberg, 1995; Schneider, Gur, Gur, & Muenz, 1994). Moreover, emotional facial expressions are used to elicit emotional responses, such as valence- and arousal-dependent approach and avoidance behavior with the affect-modulated startle reflex (Balaban, 1995; Hess, Sabourin, & Kleck, 2007; Springer, Rosas, McGetrick, & Bowers, 2007), emotion-specific facial action (for a review on facial electromyography, see Dimberg & Öhman, 1996), or emotion-related changes in autonomic activity (e.g.,

American Journal of Psychology
Summer 2010, Vol. 123, No. 2 pp. 209–219 •

heart rate changes and electrodermal activity; Johnsen, Thayer, & Hugdahl, 1995; Vrana & Gross, 2004). Finally, emotional facial expressions are often used in studies that investigate neuronal activity in emotionally relevant brain areas (Fitzgerald, Angstadt, Jelsone, Nathan, & Phan, 2006).

However, several sets of emotional facial expressions are currently available, including the Pictures of Facial Affect (Ekman & Friesen, 1976), the Japanese and Caucasian Facial Expressions of Emotion (JACFEE; Biehl et al., 1997), the Karolinska Directed Emotional Faces (KDEF; Lundqvist, Flykt, & Öhman, 1997), and the NimStim set (Tottenham et al., 2009), but these picture sets are anything but uniform. Some are black and white (Pictures of Facial Affect); others are color photographs (KDEF, NimStim, JACFEE). For some, the emotional expression is posed by professional or trained, well-instructed actors (e.g., NimStim, KDEF); others use untrained actors and did not give any explicit instruction (Mazurski & Bond, 1993). Some show Caucasian actors (KDEF, Pictures of Facial Affect); others use actors of different ethnicity (JACFEE, NimStim). In some, pictures appear outdated (Pictures of Facial Affect); in others styles are more up to date (NimStim, KDEF). Moreover, some sets show more intense emotional expressions (e.g., NimStim, Pictures of Facial Affect); others show less expressively posed expressions (e.g., KDEF).

Nevertheless, with respect to the aforementioned studies, these different sets aim to induce strong emotional responses in the viewer. To our knowledge, no study has directly examined whether different picture sets induce comparable emotional responses. Emotional reactions are sensitive to numerous characteristics of the stimulus. For example, different picture contents induce different levels of emotional valence and arousal. Emotional valence and arousal modulate the characteristics and the intensity of emotional responses, also along discrete emotion categories (Nyklicek, Thayer, & Van Doornen, 1997; Schlosberg, 1952; Thayer & Miller, 1988; Watson & Tellegen, 1985). This has been shown for musical (North & Hargreaves, 1995), scenery (e.g., from the International Affective Picture System set; Bradley, Codispoti, Cuthbert, & Lang, 2001; Cuthbert, Bradley, & Lang, 1996), olfactory (Ehrlichman, Brown Kuhl, Zhu, & Warrenburg, 1997), and facial stimuli (Russell & Bullock, 1985). For emotional facial expressions, modulatory influences of arousal have been shown even within a single picture set. Johnsen et al. (1995) demonstrated that arousal ratings of particular pictures of the Pictures of Facial Affect correlated positively with the skin conductance response that was evoked by the picture. Moreover, it has been proposed that the neural circuits involved in emotional processing, such as the amygdala, may respond more to stimuli that directly elicit strong emotions (Nakamura et al., 1999) and are probably more sensitive to the intensity than to the valence of the presented stimulus (Anderson et al., 2003).

In the same vein, for emotional scenes, it has been demonstrated that more extreme picture content (i.e., more intense stimuli) induces more extreme levels of subjective arousal and physiological responses (Cuthbert et al., 1996). For facial expressions, using positron emission tomography, Morris et al. (1996) showed an increasing response of emotionally relevant brain structures to faces with increasing emotional intensity. Similarly, decoding accuracy of facially expressed emotions decreases when participants are exposed to naturally occurring, more subtle expressions, in contrast to more extremely posed expressions (Hess, Blairy, & Kleck, 1997; Motley & Camden, 1988).

Thus, it seems logical to conclude that facial expressions that are apparently more intense should induce more intense emotional states and also lead to superior decoding accuracy. However, numerous variables can influence the processing of emotional facial expressions. For example, it has been argued that male expressions are perceived as more dominant, whereas female expressions are perceived as more submissive (Hess, Adams, & Kleck, 2005), and negative facial expressions of male actors and positive expressions of female actors are usually rated as more intense (Hess et al., 1997). Furthermore, some authors reported that ratings of perceived intensity in response to several emotional expressions might be influenced by an interaction of the sex of the actor and the sex of perceiver (Hess et al., 1997). This suggests a more complex relationship between self-reported intensity in response to an emotional facial expression and the expression intensity of the perceived face. Therefore, it seems unlikely that more intense expressions in general induce more intense emotions, although they may lead to greater decoding accuracy.

EXPERIMENT

The present study tested whether more intense expressions induce a more intense emotional experience and whether the sex of the actor influences the experience of the emotions. Moreover, we examined the influence of these variables on the accuracy with which the emotions can be decoded. Because few studies directly compare different sets of facial expressions, we presented pictures from two sets of emotional facial expressions that apparently differ in the intensity of posed emotional expressions, the Karolinska Directed Emotional Faces and the NimStim set, instead of manipulating the intensities of one single set of facial expressions. The apparent differences in the expressions' intensity were confirmed in a pilot study.

Aside from the differences in intensity, both sets are up to date and include the six basic emotions identified by Ekman and Friesen (1971) as well as neutral facial expressions. Both sets are frequently used, and they are similar on several relevant physical characteristics: The pictures are professional color photographs, well controlled for background and luminance, and portray trained actors. The most apparent difference is the intensity of emotional facial expressions. Whereas the actors in the KDEF set show more subtle, real-life expressions, those in the NimStim set show more intense, exaggerated expressions.

In line with adaptation-level theory (see Helson, 1964), it has been shown that the order in which facial stimuli from opposing valence categories are presented (e.g., first negative and then positive) can influence self-report ratings of emotion (Thayer, 1980a, 1980b). In brief, adaptation-level theory states that prior experience of a stimulus creates a reference point against which later stimuli are judged, so there is a strong chance that order effects will occur. In our study, presenting a set of more (or less) intense (and therefore more or less emotionally evocative) stimuli first probably influenced the ratings on a subsequently presented set of stimuli. Therefore, we counterbalanced presentation order of picture sets (i.e., t1 KDEF, t2 NimStim; t1 NimStim, t2 KDEF) and included presentation order as a between-participant factor in our analyses. This design allowed us to directly compare the effects of within-participant and between-participant experimental designs on the intensity of emotional experiences. Because we wanted to rule out the possibility of further order effects due to the mixed presentation of female and male facial expressions, we included sex of the actor as a between-participant factor. This further enabled us to assess the effects of the sex of the actor separately.

Our main aim was to assess the relationship between the intensity of emotional facial expressions and the perceived level of emotional arousal they induce. By operationalizing more and less intense emotional expressions with pictures from two recent picture sets, we were able to compare the overall potency of these picture sets in inducing strong emotional responses. We included neutral faces; pictures from two negative emotion categories, depicting threatening and fear-relevant stimuli (i.e., fearful and angry facial expression); and happy and surprised expressions.

METHOD

Participants

A total of 121 psychology students (69 female) received course credit for their participation (age range 19–40 years, *M* = 25.0, *SD* = 5.0). Participants were within the normal range for trait anxiety, *M* = 38.9, *SD* = 8.5 on the State–Trait Anxiety Inventory (STAI; Laux, Schaffner, Glanzmann, & Spielberger, 1981). All participants had normal or corrected-to-normal vision and affirmed understanding of task instructions. Informed consent was obtained.

Stimulus Material and Apparatus

One hundred ten slides of 22 (11 male and 11 female) actors of European descent, each with an angry, fearful, neutral, happy, or surprised facial expression, were randomly selected from the NimStim[1] and the KDEF (slides from the NimStim set depicted open mouths). The color slides were adjusted for size (resulting in a visual angle of 13° × 10°) and were comparable in luminance. Participants viewed stimuli on a 15-in. thin film transistor display. Picture delivery was controlled with the Presentation software (Neurobehavioral Systems Inc., Albany, New York).

Pilot Study

To quantify the differences in the intensity with which the emotions are expressed in the two picture sets, we first compared intensity ratings to a subset of KDEF and NimStim faces in a pilot study. Eleven (6 female) participants (mean age 28.4 years, *SD* = 5.4), who did not participate in the main experiment, were tested individually. Twenty-four slides, including three male and female actors from the NimStim and KDEF set displaying the emotions anger, fear, happiness, and surprise, were chosen randomly from the 110 slides used in our main study. Neutral expressions were excluded from the pilot study because they do not express an emotion. The faces were presented in random order on a 15-in. thin film transistor display for 2.5 s each, immediately followed by the rating scale. Picture presentation was controlled with Presentation software. The participants were asked to indicate for each of the 24 emotional expressions the intensity with which each face expressed the displayed emotion (range: 1 = *not at all,* 9 = *extremely intensely*). Mean intensity ratings were calculated for each set, and the data were then entered into a *t* test. Cohen's effect size *d* was calculated. Results indicate that expressions from the NimStim set are indeed perceived as more intense than those from the KDEF set, $t(10) = 6.34$, $p < .001$, $d = 5.23$ (Figure 1).

Procedure

For the main study, written informed consent was obtained. Participants then viewed facial stimuli in two blocks of 55 slides each. The first half of the participants ($N = 61$, 33 women) viewed slides of female actors, and the other half ($N = 60$, 36 women) viewed slides of male actors. The presentation order was counterbalanced across participants (i.e., t1 KDEF, t2 NimStim; t1 NimStim, t2 KDEF). Within each block the sequence of facial expressions was randomized. Each picture was presented for 3 s and was immediately followed by three rating scales. First, for each picture participants were asked to rate their emotional reactions; that is, how they felt during exposure to the expression (scale 1–9: 1 = *very unpleasant,* 5 = *neutral,* 9 = *very pleasant*); second, they rated how emotionally aroused they were (scale 1–9: 1 = *not at all aroused,* 9 = *very aroused*). Afterwards, they performed a seven-alternative forced-choice categorization task on the presented expression. They were asked to indicate the particular emotion displayed by that face in choosing one out of seven specific labels (i.e., *anger, fear, disgust, sadness, surprise, happiness,* and *neutral*).

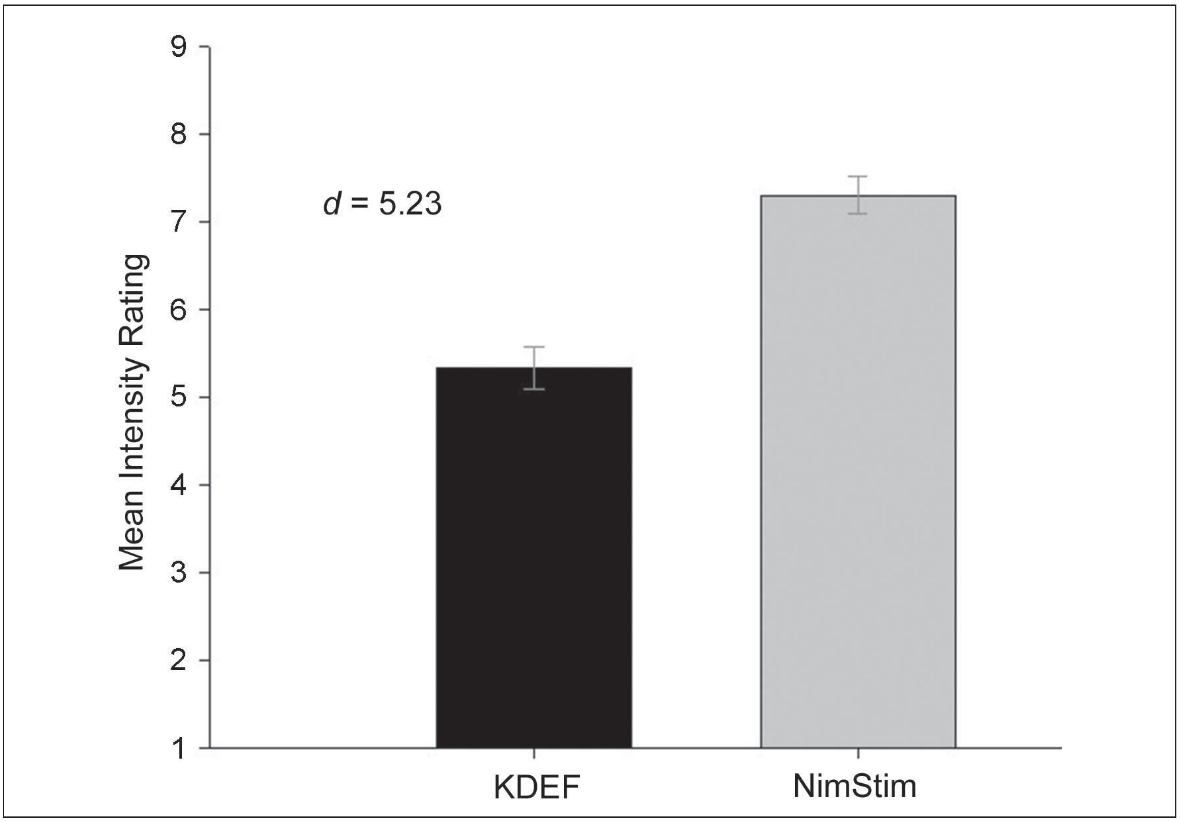

FIGURE 1. Mean intensity ratings (<+/–>*SEM*) for NimStim and KDEF faces from the pilot study (*N* = 11)

At the end of the session participants completed the trait form of the STAI.

Data Reduction and Analysis

To examine differences in decoding accuracy and differences in mean valence and arousal ratings between the two sets, separate ANOVAs were run for each dependent variable. Our main aim was to compare the two picture sets according to the level of emotional arousal they induce. Thus, the within-participant factor stimulus set (NimStim, KDEF) and the between-participant factors sex of the actor (male, female) and presentation order (t1 KDEF, t2 NimStim; t1 NimStim, t2 KDEF) were entered into the ANOVA. In addition, to account for possible effects of the five emotion categories, we also included a within-participant factor emotion (angry, fearful, neutral, surprised, happy). For all effects, Cohen's effect size f was calculated (Cohen, 1988). Huynh–Feldt corrections of degrees of freedom were applied, and corrected p values are reported. An alpha level of 5% was used for all statistical tests. Cohen's effect size d was calculated for follow-up t tests on overall differences between the sets.

Because viewing pictures from one set influences the emotional experience of the other set, only data from t1 would give valid information about the emotional properties of the two sets (because in that case participants have not seen the other set before). Therefore, follow-up analyses on effects involving the factor presentation order were performed for all dependent variables, including data from t1 and t2 together, and in a separate analysis with data from t1 only.

RESULTS

Decoding Accuracy

OVERALL EFFECTS

Mean decoding accuracy for both sets ranged between 69% and 78% (Table 1) and was therefore well above chance (i.e., 20%). Overall (data from t1 and t2), the ANOVA indicates that faces from the NimStim set were decoded more accurately than those from the KDEF set, $F(1, 116) = 20.72$, $p < .001$, $f = 0.42$ (main effect for stimulus set). This is consistent for faces posed by female, $t(58) = 3.83$, $p < .001$, $d = 0.48$, and male actors, $t(60) = 2.29$, $p = .027$, $d = 0.22$.

There was a significant interaction for set by presentation order, $F(1, 116) = 5.01$, $p = .027$, $f = 0.21$, and a significant interaction for stimulus set by sex of the actor, $F(1, 116) = 4.19$, $p = .043$, $f = 0.19$. Follow-up analyses of these interactions indicate that at t1 female faces from the NimStim set were identified more accurately than those from the KDEF set, $t(57) = 2.06$, $p = .044$, $d = 0.53$, and there were no differences for male actors, all ps > .10 (see Table 1).

DIFFERENCES BETWEEN THE FIVE EMOTIONS

There was a significant main effect for emotion, $F(4, 464) = 165.10$, $p < .001$, $f = 1.19$ (i.e., happy > angry > neutral = surprise > fearful, all ps < .001; comparison of neutral with surprise, $p > .200$). There were also significant interactions for stimulus set by emotion, $F(4, 464) = 17.42$, $p < .001$, $f = 0.39$, and stimulus set by emotion by sex of the actor, $F(4, 464) = 5.07$, $p = .001$, $f = 0.21$. Finally, there was a trend for an in-

TABLE 1. Decoding accuracy, valence, and arousal ratings as a function of stimulus set, sex of the actor, and presentation order

		Female actors						Male actors					
		t1 and t2		t1 only		t2 only		t1 and t2		t1 only		t2 only	
		M	*SD*	*M*	*SD*	*M*	*SD*	M	*SD*	*M*	*SD*	*M*	*SD*
Decoding accuracy (%)	KDEF	70.42	14.99	69.03	16.16	71.76	13.90	74.34	11.15	74.31	8.64	74.36	13.42
	NimStim	77.20	13.17	75.82	7.89	78.62	17.05	76.96	12.71	75.09	14.99	78.77	9.95
Valence	KDEF	4.65	0.48	4.49	0.57	4.80	0.17	4.65	0.40	4.66	0.46	4.64	0.35
	NimStim	4.49	0.45	4.66	0.46	4.32	0.39	4.58	0.47	4.59	0.40	4.57	0.54
Arousal	KDEF	3.70	1.50	4.17	1.40	3.24	1.48	3.63	1.53	3.40	1.38	3.86	1.67
	NimStim	4.00	1.49	3.51	1.36	4.49	1.48	3.88	1.61	4.10	1.62	3.66	1.59

Note. *Valence and arousal ratings range from 1 to 9. KDEF = Karolinska Directed Emotional Faces.*

teraction for stimulus set by emotion by presentation order, $F(4, 464) = 2.42$, $p = .057$, $f = 0.14$.

Follow-up t tests for these interactions indicate that (including data from t1 and t2) for female expressions, decoding accuracy was higher in the NimStim set for angry, $t(58) = 7.83$, $p = .001$, fearful, $t(58) = 3.64$, $p = .001$, and, as a trend, for happy expressions, $t(58) = 1.85$, $p = .070$, than in the KDEF set. There were no significant differences for surprised and neutral expressions between the stimulus sets, $p > .10$.

For male expressions from the NimStim set decoding accuracy was higher for fearful, $t(58) = 5.10$, $p < .001$, and happy, $t(58) = 1.87$, $p = .066$, slides but lower for surprised slides, $t(58) = 3.39$, $p = .001$, compared with male expressions from the KDEF set.

Finally, decoding accuracy was higher for NimStim expressions when regarding data from t1 only: Angry, $t(57) = 3.46$, $p = .001$, and fearful, $t(57) = 1.90$, $p = .063$, female expressions, and fearful male expressions, $t(59) = 2.62$, $p = .011$, from the NimStim set were decoded more accurately than corresponding expressions from the KDEF set.

Valence

OVERALL EFFECTS

Table 1 shows mean valence ratings for male and female actors of both sets. Overall, faces from the KDEF set were rated as more positively than those from the NimStim set, $F(1, 117) = 9.41$, $p = .003$, $f = 0.28$ (main effect for stimulus set). There was a main effect for presentation order, $F(1, 117) = 5.16$, $p = .025$, $f = 0.21$, and a significant interaction for presentation order by sex of the actor, $F(1, 117) = 5.30$, $p = .023$, $f = 0.21$.

Follow-up analyses show that regarding data from t1 and t2 together, female but not male ($p > .10$) faces from the KDEF set were rated as more positively than those from the NimStim set, $t(59) = 2.68$, $p = .009$, $d = 0.34$. However, regarding data from t1 only, there were no differences between the sets, all ps $> .10$.

DIFFERENCES BETWEEN THE FIVE EMOTIONS

Regarding data from t1 and t2, valence ratings also differed for the five emotion categories, $F(4, 468) = 318.60$, $p < .001$, $f = 1.65$ (main effect for emotion, i.e., angry < fearful < neutral < surprised < happy, all ps < 0.001). There was also an interaction for stimulus set by emotion, $F(4, 468) = 6.29$, $p < .001$, $f = 0.23$. A significant interaction including the factors stimulus set, emotion, and presentation order was evident, $F(4, 468) = 5.35$, $p = .001$, $f = 0.21$, as was a trend toward a significant interaction including stimulus set, emotion, and sex of the actors, $F(4, 468) = 2.21$, $p = .076$, $f = 0.14$. These indicate that the perceived valence of the five emotion categories differed as a function of sex of the actor, stimulus set, and presentation order. Follow-up analyses of these interactions show that overall (i.e., including data from t1 and t2) female actors with angry, $t(59) = 3.76$, $p < .001$, and surprised, $t(59) = 3.26$, $p = .002$, facial expressions were perceived as more positive in the KDEF set than in the NimStim set. For male actors (t1 and t2), angry, $t(60) = 2.52$, $p = .014$, and fearful, $t(69) = 2.55$, $p = .013$, expressions from the NimStim set were perceived as more negative than those from the KDEF set. However, in t1 fearful female faces from the KDEF set were rated as more negative than fearful female faces from the NimStim set, $t(58) = 2.80$, $p = .007$. There were no other differences in t1 between the sets according to the five emotion categories.

Arousal

OVERALL EFFECTS

Arousal ratings ranged between 3.4 and 4.8 (on a 9-point scale) for both stimulus sets, indicating that faces from both sets were perceived as mildly arousing (see Table 1). Overall (including data from t1 and t2), faces from the KDEF set were rated as less arousing than those from the NimStim set, $F(1, 117) = 21.43$, $p < .001$, $f = 0.43$ (main effect for stimulus set). However, subjective ratings of arousal for the two sets differed as a function of presentation order and sex of the actor, $F(1, 117) = 6.93$, $p = .010$, $f = 0.24$.

Regarding data from t1 and t2 together, male, $t(60) = 3.41$, $p = .001$, $d = 0.16$, and female, $t(59) = 3.24$, $p = .002$, $d = 0.20$, faces from the NimStim set were rated as more arousing than those from the KDEF set. However, regarding data from t1 only, female faces from the KDEF set, $t(58) = 1.83$, $p = .073$, $d = 0.48$, but male faces from the NimStim set were rated as more arousing, $t(59) = 1.83$, $p = .073$, $d = 0.47$ (see Table 1).

DIFFERENCES BETWEEN THE FIVE EMOTIONS

Arousal ratings differed for the five emotion categories, $F(4, 468) = 102.00$, $p < .001$, $f = 0.93$ (main effect for emotion). Follow-up t tests revealed that (including data from t1 and t2) angry, fearful, and happy expressions were rated equally arousing, all *p*s > .10. However, surprised expressions were rated as less arousing than angry, fearful, and happy expressions, all *p*s < .001, and neutral expressions were rated as less arousing than all emotional expressions, all *p*s < .001. There was also a significant interaction for stimulus set by emotion, $F(4, 468) = 9.55$, $p < .001$, $f = 0.28$, and a significant interaction including the factors stimulus set, emotion, and presentation order, $F(4, 468) = 9.28$, $p < .001$, $f = 0.28$, indicating that the perceived arousal of the emotional expressions differed as a function of sex of the actor, stimulus set, and presentation order.

Follow-up analyses of these interactions indicate that overall (including data from t1 and t2) female actors from the NimStim set showing angry, $t(59) = 5.70$, $p < .001$, fearful, $t(59) = 2.49$, $p = .016$, and, as a trend, happy, $t(59) = 1.77$, $p = .082$, facial expressions were rated as more arousing than those from the KDEF set. There were no differences between the KDEF and NimStim set in arousal ratings for neutral and surprised facial expressions, all *p*s >.10. Angry, $t(60) = 3.91$, $p < .001$, fearful, $t(60) = 2.73$, $p = .008$, surprised, $t(60) = 2.28$, $p = .026$, and neutral, $t(60) = 1.96$, $p = .054$, male expressions from the NimStim set were rated as more arousing than those from the KDEF. However, mean arousal ratings for happy male expressions did not differ between the two sets, $t(60) = 1.40$, $p = .168$.

For data from t1 only, fearful, $t(58) = 2.04$, $p = .046$, surprised, $t(58) = 2.30$, $p = .025$, and, as a trend, happy, $t(58) = 1.79$, $p = .079$, female expressions but not angry, $t(58) = .87$, $p = .390$, and neutral expressions, $t(58) = 1.47$, $p = .148$, from the KDEF set were rated as more arousing than those from the NimStim set. For male expressions follow-up *t* tests of this interaction indicate that when they were seen at t1, surprised, $t(59) = 2.03$, $p = .047$, neutral, $t(59) = 2.28$, $p = .026$, and fearful, $t(59) = 1.83$, $p = .071$, expressions from the NimStim set were rated as more arousing than those from the KDEF set. There were no differences between the stimulus sets in arousal ratings for angry and happy expressions at t1, all *p*s > .10.

DISCUSSION

In psychological research, emotional facial expressions are often used as probes for emotion recognition and to elicit emotions in the viewer, although available picture sets have not been directly compared according to their emotion-eliciting properties. We assessed valence and arousal ratings and decoding accuracy for two widely used sets of photographs, NimStim and KDEF, which differ in the intensities of the expressions. We also included presentation order of the two picture sets as a factor in our analyses.

Overall, more intensely posed facial expressions from the NimStim set elicited higher emotional arousal, were rated differently in terms emotional valence, and were more accurately identified. However, presentation order strongly influenced subjective ratings. Consequently, we also analyzed data from t1 only (i.e., participants had not seen faces from the other set before). These analyses showed that more intensely posed male faces (NimStim) and less intensely posed female faces (KDEF) elicited higher levels of subjective arousal.

Decoding Accuracy

In general, facial expressions were identified accurately, that is above chance, for both male and female actors from both picture sets. Decoding accuracy was higher for the more intense faces (NimStim; regarding data from t1 and t2 together), replicating previous findings (Hess et al., 1997; Motley & Camden, 1988). Given that knowing what other people feel is an important element of everyday social interaction (Hess et al., 1997), this result further supports the finding that people are better at reading discrete emotions from more intense expressions.

However, our data indicate that decoding accuracy may not only depend on expression intensity but may vary with expression intensity and sex of the actor. When we excluded possible contrast effects by examining data from t1 only, decoding accuracy did not differ between the more intense (NimStim) and the less intense (KDEF) male faces, but it was significantly lower for the less intense female faces (KDEF) than for the more intense faces (NimStim). Hess et al. (1997) pointed to a systematic decoding bias for low-intensity female emotional expressions as compared to low-intensity male expressions. Therefore, the current findings on decoding accuracy have important implications for future studies: Those using less intense emotional expressions, such as those from the KDEF set, may be more prone to decoding biases than studies using more intensely posed emotional expressions, such as those from the NimStim set.

Decoding accuracy was highest for happy expressions, replicating previous research (e.g., Ekman & Friesen, 1971; Hess et al., 1997). Fearful and surprised expressions had low levels of correct classifications (including both sexes and both sets). Almost 30% of those expressions were erroneously classified as showing disgust or surprise. This is plausible because these expressions share some relevant features (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003), and poor classification results of fearful facial expressions have been reported previously (e.g., Eisenbarth, Alpers, Segre, Calogero, & Angrilli, 2008).

Self-Reported Valence

Overall, concerning self-reported valence, male faces from the KDEF and the NimStim set were rated similarly, and female expressions from the KDEF set were rated slightly more positive than those from the NimStim set (data from t1 and t2). But when data from t1 were analyzed separately, there were no differences between the sets.

Analysis of the effects of the five emotions showed that overall (t1 and t2) valence ratings differed between the sets for male actors (i.e., fearful and angry expressions from the NimStim set were rated as more negative than those from the KDEF) and for female actors: Female angry and surprised pictures of the NimStim set were rated as more negative than those from the KDEF set, explaining the overall differences in valence ratings. When we eliminated the influence of presentation order, ratings of self-report valence differed for fearful female expressions only, that is, fearful faces from the KDEF set were rated as more negative than fearful female faces from the NimStim set.

However, together with the high levels of decoding accuracy, our results indicate that expressions of both sets induce qualitatively comparable emotional experiences. Despite the differences in female fearful expressions, valence ratings show high degrees of similarity between the two sets.

Self-Reported Arousal

Emotional expressions were rated as more arousing than neutral expressions, for female and male actors, showing that the participants were more emotionally engaged when viewing emotional facial expressions than when viewing neutral ones. These results have been reported previously for emotional facial expressions (e.g., Johnsen et al., 1995).

However, there were also differences between the sets. Overall (t1 and t2), facial expressions of the NimStim set were clearly rated as more arousing than those from the KDEF set. At first glance, this may suggest a linear relationship between expression intensity and self-report arousal. But, considering subjective ratings at t1 alone (as in usual experiments with only one block of picture presentation), the less intense female expressions from the KDEF set were rated as more arousing, and the more intense male expressions from the NimStim set were rated as more arousing. This shows that the sex of the actor influences the perceived intensity of experienced emotions. (Effect sizes for this effect are $d = 0.48$ for female faces and $d = 0.47$ for male faces, suggesting a robust effect.)

For female actors, the less intense expressions caused more intense emotional arousal. In everyday life, facial expressions are generally more subtle (Elfenbein, Marsh, & Ambady, 2002), and people are experts in interpreting naturally occurring facial expressions (Noller, 1985). Because the NimStim expressions show more intense emotions and are more strongly posed than those that are frequently encountered in everyday life, the participants may have felt less emotionally involved. This effect may have emerged for women because women are described to be generally more facially expressive of most emotions than men are (Brody & Hall, 2000), and these stereotypic expectations in the receiver may have led to more emotional contagion even to less intense expressions of emotion. This effect was strongest for less intense fearful, surprised, and happy female expressions from the KDEF set, which are rated as more arousing than similar female expressions from the NimStim set. This is plausible, because women are thought to smile more than men (Briton & Hall, 1995), and fear is reported even more often by women than by men (Brody, 1999). The more intense male expressions from the NimStim set were rated as more arousing then the less intense expressions from the KDEF set (data from t1 only). This may have occurred because men are more likely to show more intense (negative) emotions than women (see Hess, Adams, & Kleck, 2004; Hess et al., 2005). This effect was most pronounced in fearful, surprised, and neutral male expressions. Interestingly, a recent study suggests that a neutral facial expression is perceived as a sign of dominance in men (Hareli, Shomrat, & Hess, 2009). In the present study the more intense neutral expressions may have been perceived as even more dominant and therefore led to more emotional engagement. This is also supported by the ratings of emotional valence, showing that neutral male expressions from the NimStim set were rated as more negative than neutral male expressions from the KDEF set.

Taken together, our results suggest that more intense expressions may be more accepted in men, and extremely posed expressions may be perceived as even more unnatural and therefore less arousing in women.

Implications

Our results show that emotional reactions to facial expressions, although depicting identical emotional content, are not invariant. Ratings of two professionally posed and photographed sets of emotional expressions differed with respect to decoding accuracy, self-reported emotion arousal, and emotional valence. Furthermore, the results varied with experimental approach (data from t1 only = one block of picture presentation in a between-participant design; data from t1 and t2 = counterbalanced blocks of picture presentation in a within-participant design) and the sex of the actor.

The findings on our counterbalanced block presentation are generally consistent with adaptation-level theory (Helson, 1964), which suggests that a person's current emotional state provides a context for judgments of the emotional states induced by subsequent stimuli (Manstead, Wagner, & MacDonald, 1983). Our data indicate that emotional responding to emotional facial expressions may be similarly affected by such contrast effects.

This is especially important because some studies used different sets of emotional facial picture material in one experimental design (e.g., Hess et al., 2005; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004), whereas others used a single set. Moreover, physiological reactions can be influenced by contrast effects, such as those observed in the present study (Alpers & Adolph, 2006; Spangler, Emlinger, Meinhardt, & Hamm, 2001). Finally, even blood flow within emotional brain circuitry seems to be sensitive to contrast effects (Somerville, Kim, Johnstone, Alexander, & Whalen, 2004).

The finding that not all sets of emotional facial expressions are equal has implications for future research. Selection of the stimulus material regarding sex of the actor seems to be very important for researchers using emotional facial expressions. If only male faces are used, more intense expressions seem more appropriate. When using female expressions, more subtle, less intense emotions may be more appropriate. Researchers should be also concerned about the methods and experimental design in which emotional expressions are presented: Using within- or between-participant designs may lead to different results concerning decoding accuracy, emotional valence, and arousal.

NOTES

We are grateful for the support of our work by Paul Pauli. We would like to thank Sarah Heston for proofreading the manuscript.

Address correspondence about this article to Georg W. Alpers, Department of Psychology, University of Würzburg, Marcusstrasse 9-11, 97070 Würzburg, Germany (e-mail: alpers@psychscience.org).

1. Numbers of the actors from the NimStim set were 01F_O, 02F_O, 03F_O, 05F_O, 06F_O, 07F_O, 08F_O, 09F_O, 10F_O, 17F_O, 18F_O, 20M_O, 21M_O, 22M_O, 23M_O, 24M_O, 25M_O, 29M_O, 31M_O, 35M_O, 36M_O, 37M_O. Numbers of the actors from the KDEF set were AF01, AF05, AF07, AF08, AF11, AF13, AF14, AF17, AF19, AF20, AF22, AM01, AM02, AM03, AM05, AM06, AM07, AM10, AM13, AM21, AM30, AM34. Tables with mean valence and arousal ratings of the rated expressions and the five emotion categories can be obtained from the authors.

REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12,* 169–177.

Alpers, G. W., & Adolph, D. (2006). Startle and automatic nervous system modulation while viewing emotional scenes or emotional facial expressions. *Psychophysiology, 43,* S7.

Alpers, G. W., & Gerdes, A. B. (2007). Here is looking at you: Emotional faces predominate in binocular rivalry. *Emotion, 7,* 495–506.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience, 6,* 196–202.

Balaban, M. T. (1995). Affective influences on startle in five-

month-old infants: Reactions to facial expressions of emotion. *Child Development, 66,* 28–36.
Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., et al. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21,* 3–21.
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.
Briton, N., & Hall, J. (1995). Beliefs about female and male nonverbal communication. *Sex Roles, 32,* 79–90.
Brody, L. R. (1999). *Gender, emotion, and the family*. Cambridge, MA: Harvard University Press.
Brody, L. R., & Hall, J. (2000). Gender, emotion and expression. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 338–349). New York: Guilford.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology, 33,* 103–111.
Dimberg, U., & Öhman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion, 20,* 149–182.
Ehrlichman, H., Brown Kuhl, S., Zhu, J., & Warrenburg, S. (1997). Startle reflex modulation by pleasant and unpleasant odors in a between-subjects design. *Psychophysiology, 34,* 726–729.
Eisenbarth, H., Alpers, G. W., Segre, D., Calogero, A., & Angrilli, A. (2008). Categorization and evaluation of emotional faces in psychopathic women. *Psychiatry Research, 159,* 189–195.
Ekman, P., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions.* New York: Oxford University Press.
Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–129.
Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.
Elfenbein, H. A., Marsh, A. A., & Ambady, N. (2002). Emotional intelligence and the recognition of emotion from facial expressions. In L. F. Barrett & P. Salovey (Eds.), *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 37–59). New York: Guilford.
Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *NeuroImage, 30,* 1441–1448.
Hareli, S., Shomrat, N., & Hess, U. (2009). Emotional versus neutral expression and perceptions of social dominance and submissiveness. *Emotion, 9,* 378–384.
Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper & Row.
Hess, U., Adams, R. B., Jr., & Kleck, R. E. (2004). Facial appearance, gender, and emotion expression. *Emotion, 4,* 378–388.
Hess, U., Adams, R. B., Jr., & Kleck, R. E. (2005). Who may frown and who should smile? Dominance, affiliation, and the display of happiness and anger. *Cognition & Emotion, 19,* 515–536.
Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology, 40,* 129–141.
Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior, 21,* 241–257.
Hess, U., Sabourin, G., & Kleck, R. E. (2007). Postauricular and eyeblink startle responses to facial expressions. *Psychophysiology, 44,* 431–435.
Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion–cognition relations. *Psychological Review, 99,* 561–565.
Johnsen, B. H., Thayer, J. F., & Hugdahl, K. (1995). Affective judgment of the Ekman faces: A dimensional approach. *Journal of Psychophysiology, 9,* 193–202.
Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport, 14,* 2317–2322.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Digitized photographs, instruction manual and affective ratings.* Technical report A-6. Gainesville: University of Florida.
Laux, L., Schaffner, P., Glanzmann, P., & Spielberger, C. D. (1981). *Das State–Trait Angstinventar (STAI).* Weinheim, Germany: Beltz Testgesellschaft.
Lundqvist, L. O., & Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology, 9,* 203–211.
Lundqvist, L. O., Flykt, A., & Öhman, A. (1997). *The Karolinska Directed Emotional Faces: KDEF*. Stockholm: Karolinska Institutet.
Manstead, A. S. R., Wagner, H. L., & MacDonald, C. J. (1983). A contrast effect in judgements of own emotional state. *Motivation and Emotion, 7,* 279–290.
Mazurski, E. J., & Bond, N. W. (1993). A new series of slides depicting facial expressions of affect: A comparison with the pictures of facial affect series. *Australian Journal of Psychology, 45,* 41–47.
Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383,* 812–815.
Motley, M. T., & Camden, C. T. (1988). Facial expression

of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication, 52,* 1–22.

Nakamura, K., Kawashima, R., Ito, K., Sugiura, M., Kato, T., Nakamura, A., et al. (1999). Activation of the right inferior frontal cortex during assessment of facial emotion. *Journal of Neurophysiology, 82,* 1610–1614.

Noller, P. (1985). Video primacy: A further look. *Journal of Nonverbal Behavior, 9,* 28–47.

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology, 14,* 77–93.

Nyklicek, I., Thayer, J. F., & Van Doornen, L. J. P. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology, 11,* 304–321.

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology, 48,* 1290–1298.

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology, 44,* 229–237.

Schneider, F., Gur, R. C., Gur, R. E., & Muenz, L. R. (1994). Standardized mood induction with happy and sad facial expressions. *Psychiatry Research, 51,* 19–31.

Somerville, L. H., Kim, H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biological Psychiatry, 55,* 897–903.

Spangler, G., Emlinger, S., Meinhardt, J., & Hamm, A. (2001). The specificity of infant emotional expression for emotion perception. *International Journal of Psychophysiology, 41,* 155–168.

Springer, U. S., Rosas, A., McGetrick, J., & Bowers, D. (2007). Differences in startle reactivity during the perception of angry and fearful faces. *Emotion, 7,* 516–525.

Thayer, J. F., & Miller, M. L. (1988). Further evidence for the independence of hedonic level and emotional intensity. *Personality and Individual Differences, 9,* 425–426.

Thayer, S. (1980a). The effect of expression sequence and expressor identity on judgments of the intensity of facial expressions. *Journal of Nonverbal Behavior, 5,* 71–79.

Thayer, S. (1980b). The effect of facial expressions sequence upon judgements of emotion. *Journal of Social Psychology, 111,* 305–306.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168,* 242–249.

Vrana, S. R., & Gross, D. (2004). Reactions to facial expressions: Effects of social context and speech anxiety on responses to neutral, anger, and joy expressions. *Biological Psychology, 66,* 63–78.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience, 7,* 1271–1278.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.