Journal of Experimental Social Psychology 77 (2018) 143–154

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Implicit evaluations of faces depend on emotional expression and group membership☆

Andrea Paulus[a,*,1], Dirk Wentura[b,**,1]

[a] Technical University of Munich, Germany
[b] Department of Psychology, Saarland University, Germany

Keywords:
Implicit evaluation
Face perception
Emotional expression
Group membership
Evaluative priming

ARTICLE INFO

Handling editor: Ursula Hess

ABSTRACT

Faces carry a lot of information influencing evaluative reactions, such as emotional expression, age, or group membership. Even though, typically, many of these aspects will be present in a face concurrently, only few studies have examined automatic evaluative reactions to faces that vary on more than one dimension. As an exception, two recent priming studies examined the concurrent influence of group membership and emotional expression. Quite astoundingly, they leave the reader with two divergent outcomes: while Weisbuch and Ambady (2008) observed an interactive influence of emotional expression and group membership on evaluative reactions, Craig et al. (2014) found that group membership did not contribute to the implicit evaluation of positive and negative emotional expressions. In order to shed light on this matter, we conducted three high-powered experiments using prime images of highly relevant in-group and out-group members expressing happiness and fear. We furthermore varied the social context of the priming task in order to give the "interaction hypothesis" a chance. However, we found no evidence for the interaction reported by Weisbuch and Ambady. In contrast to Craig et al., we found that both emotional expression and group membership independently contributed to implicit evaluations. Differences are discussed in terms of relevance of the employed groups, test power, and the time-scale of underlying processes.

Faces are important carriers of information: They inform interaction partners about many aspects that are (ostensibly) relevant for social interaction, such as emotional state, age, sex, and ethnicity. Accordingly, numerous studies have demonstrated that those features influence the immediate evaluation of a face (e.g., Bar-Anan & Nosek, 2014; Carroll & Young, 2005; Degner & Wentura, 2011; Fazio, Jackson, Dunton, & Williams, 1995; Nosek, Banaji, & Greenwald, 2002; Rohr, Degner, & Wentura, 2012). However, faces are often characterized by the presence of several evaluation-relevant features at once (e.g., emotional expression *and* ethnicity). Therefore, the question arises: when encountering another person, which of the many potentially relevant facial features are *automatically* assessed and evaluated? Does one evaluative aspect, if present, dominate the other ones? Do different aspects independently contribute to an evaluation? Or does the overall evaluation result from the interaction of several aspects?

Given that in real life the presence of more than one evaluative feature in a face is probably the norm rather than the exception, it is surprising that these questions have rarely been addressed in experimental research. Most studies examining the influence of facial features on automatic evaluative responses with so-called implicit measures have only varied one evaluative aspect while holding others constant (see, e.g., Carroll & Young, 2005; Degner & Wentura, 2011). The few studies varying more than one evaluative aspect often examined the influence of attention and/or task instruction on evaluative responses (Craig, Lipp, & Mallan, 2014; Gawronski, Cunningham, LeBel, & Deutsch, 2010; Imhoff, Schmidt, Bernhardt, Dierksmeier, &

☆ This research was supported by a grant from the German Research Foundation (WE 2284/14-1) to Dirk Wentura and Andrea Paulus.
* Correspondence to: A. Paulus, Technical University of Munich, Chair of Psychology, Arcisstraße 21, 80333 München, Germany.
** Correspondence to: D. Wentura, Saarland University, Department of Psychology, Campus A2.4, D-66123 Saarbruecken, Germany.
E-mail addresses: andrea.paulus@wi.tum.de (A. Paulus), wentura@mx.uni-saarland.de (D. Wentura).
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.jesp.2018.04.004
Received 27 November 2017; Received in revised form 3 April 2018; Accepted 8 April 2018

Banse, 2011). Gawronski and colleagues, for example, studied automatic evaluative responses to faces varying in race (i.e., black/white) and age (i.e., young/old), employing an affective misattribution procedure (AMP) as well as an evaluative priming paradigm. They showed that—depending on task—one or both aspects of the face influenced affective responses. However, in these studies participants' attention was always directed to one of the two aspects (i.e., race or age) by the instruction to keep track of the number of faces in a certain feature category.

However, there also exist a couple of studies that have examined automatic evaluative responses to faces varying on more than one evaluative aspect without directing attention to one of the two aspects (Craig et al., 2014; Weisbuch & Ambady, 2008). These studies varied both emotional expression and group membership of the faces to examine their (potentially) interactive influence on automatic evaluative responses. However, they yielded conflicting results: while Weisbuch and Ambady (2008) observed an interactive influence of group membership and emotional expression on evaluative responses, Craig et al. (2014) only found an influence of emotional expression. Group membership did not influence reactions in any way. This divergence is quite astounding, especially given the similarities between the two studies: Both studies employed happy and fearful expressions of white and black individuals as stimuli,[2] and both measured evaluative responses with an evaluative priming paradigm (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; for a review, see Klauer & Musch, 2003). In the evaluative priming paradigm, participants categorize positive and negative targets as fast as possible based on valence. Targets are preceded by briefly presented primes. Typically, if valence of prime and target match, responses are faster and/or more accurate than in the case of a non-match (for a meta-analysis, see Herring et al., 2013). The paradigm can thus be used to assess the automatic evaluation of attitude objects (Fazio et al., 1995; for a meta-analysis, see Cameron, Brown-Iannuzzi, & Payne, 2012).

Given the divergence of results, it is not surprising that different conclusions have been drawn regarding the processes underlying automatic evaluative responses to emotional in-group and out-group faces. An implication of Weisbuch and Ambady (2008) is that group membership and emotional expression are not only both processed automatically, but that the two dimensions predict evaluative responses in an interactive way. The basic assumption is that group membership influences the social meaning of the expression: Happy in-group faces should be interpreted as benevolent (i.e., signaling a safe social environment and/or a desire to affiliate), whereas happy out-group faces should be interpreted as potentially malevolent (i.e., signaling rival superiority). By contrast, fearful in-group faces should be interpreted as signaling an unsafe social environment, whereas fearful out-group faces should be interpreted as signaling inferiority. The results support the authors' assumptions. In the priming task, in which participants' task was to evaluate positive and negative target images, Weisbuch and Ambady found an evaluative priming effect for white faces (in white participants; i.e., for in-group faces), corresponding to what would be expected prima facie: happy primes led to an automatic positive evaluation whereas fearful expressions led to an automatic negative evaluation. For black faces (i.e., out-group faces), however, the pattern reversed, in accordance with a priori expectations. In this case, happy expressions led to a *negative* evaluation whereas fearful expressions led to a *positive* evaluation. Interestingly, similar results were also obtained with different dependent variables, such as tone of voice (Weisbuch & Ambady, 2008) and approach and avoidance reactions (Paulus & Wentura, 2014).

Craig et al. (2014) also employed the evaluative priming paradigm using white and black faces with happy and fearful expressions as prime stimuli. However, in stark contrast to Weisbuch and Ambady (2008), they predominantly found emotion-based priming effects (i.e., happy primes facilitated responses to positive targets and fearful prime faces facilitated responses to negative targets, irrespective of group membership). Group-based priming effects were only observed in a modified task, where race was made salient by instruction (i.e., after each target response, participants had to name the race of the individual displayed in the prime image).[3] In this task, there was an effect of in-group-favoritism/out-group-derogation (in addition to the emotion priming effect), that is, in-group faces facilitated positive target responses, whereas out-group faces facilitated negative responses, irrespective of emotional expression. The effect of emotional expression (as described above) was also observed. Interestingly, however, there was not even a slight indication of an interaction between emotional expression and group membership (as Weisbuch & Ambady, 2008, had found).

Taken together, it is currently not clear at all how faces displaying more than one evaluative aspect influence automatic evaluative responses. Therefore, the research questions of (1) whether emotional expression and social group membership are both automatically processed dimensions, and, importantly, (2) whether the two aspects interactively influence automatic target responses in an evaluative priming paradigm, are still open questions demanding further investigation. Apart from the clarification of the above-mentioned empirical discrepancies, there are additional reasons why we consider this a worthwhile endeavor.

First, apart from the priming studies presented above, there exists manifold evidence for an interactive effect of emotional expression and group membership from studies using measures other than automatic evaluative responses. To illustrate, an interaction between these two factors has been found with approach and avoidance reactions (Paulus & Wentura, 2014), facial muscle activity (Bourgeois & Hess, 2008; van der Schalk et al., 2011; but see Sachisthal, Sauter, & Fischer, 2016), the startle reflex (Paulus, Musial, & Renn, 2014; Paulus, Renn, & Wentura, 2018), mood contagion (Epstude & Mussweiler, 2009), and emotion or group recognition (Hugenberg, 2005; Hugenberg & Bodenhausen, 2003). Many of the results observed in these studies suggest that group membership influences the social meaning of emotional expression. Paulus and Wentura, for example, found that emotions (fear and happiness) expressed by in-group members elicited concordant behavior (i.e., approach for happy faces, avoidance for fearful faces), whereas emotions expressed by out-group members activated the reverse pattern. Thus, Weisbuch and Ambady's (2008) evaluative priming results are corroborated by other findings in the literature.

Second, apart from the incompatible results, low experimental power and other methodological details make it difficult to draw strong conclusions from Weisbuch and Ambady (2008) and Craig et al. (2014). The key 2 (group: in-group vs. out-group) × 2 (emotion: happy vs. fear) interaction found by Weisbuch and Ambady (Experiment 1) was associated with an effect size of $d_Z = 0.40$ (see below for details). Thus, the power to replicate this effect was only $1 - \beta = 0.61$ for Experiment 2 of Weisbuch and Ambady (who found an effect) and $1 - \beta = 0.55$ and 0.59 for Experiments 1 and 2, respectively, of Craig et al. (who did not find the effect). Given these power values, one successful and two unsuccessful replication attempts is not an unlikely outcome. Furthermore, Weisbuch and Ambady employed only 48 priming trials. More

[2] In the present manuscript we focus solely on those studies which are directly comparable (i.e., those which employed black and white individuals showing happy and fearful expressions as stimuli), that is, Experiment 2 of Craig et al. (2014) and Experiments 1 and 2 of Weisbuch and Ambady (2008). To be precise, however, both groups report additional studies in their respective papers. The article by Craig et al. (2014) encompassed two further experiments with the pattern of results being almost the same to the one of Experiment 2: In Experiment 1, they used emotional Black and White faces, but happy and angry emotional expressions; in Experiment 3, they used happy and fearful emotional expressions, but young and old faces. Weisbuch and Ambady (2008) had three further experiments to test their theory. However, in these studies the experimental paradigms were different from evaluative priming.

[3] The study also included a version in which the emotional expression was made salient. In this version, only a main effect for emotional expression occurred.

typical trial numbers range from 100 to 300 in the evaluative priming paradigm (see Wentura & Degner, 2010), so this detail is also noteworthy. Taken together, we argue that a replication of Weisbuch and Ambady's as well as Craig et al.'s priming studies with sufficient power and a larger number of trials was warranted.

Furthermore, Craig et al. (2014) themselves noted that the out-group employed in their experiments might not have been an ideal choice in order to demonstrate an influence of group membership on emotional expression: Just like Weisbuch and Ambady (2008), Craig et al. used images of African-American and White-Caucasian individuals as primes. However, in contrast to Weisbuch and Ambady, Craig et al.'s participants were White-Caucasian Australians. Prejudice against African-American individuals might not be as wide-spread in this sample as it is in a White-Caucasian American sample. The authors themselves stated that "…it may be the case that attitudes relating to race are not chronically activated or contextually relevant to the same extent as they are in an American sample." (p. 875). This observation might also be able to explain the divergence between the findings of Craig et al. and other authors, who showed an influence of group membership on evaluative responses (e.g., Degner & Wentura, 2011). Therefore, the replication should use an out-group that is unequivocally relevant in the culture the data is collected in.

A last argument for another examination of the interplay between emotional expression and group membership is that the hypothesis put forward by Weisbuch and Ambady (2008; and other researchers employing other paradigms) presume a social interaction context (albeit implicitly). The authors argue from the perspective of an observer who evaluates individuals with regard to their relational qualities: "A smiling out-group member is negative *for me* because it demonstrates superiority *in relation to me*; a fearful out-group member is positive *for me* because it demonstrates weakness *in relation to me*." Therefore, one might argue that some kind of interactive context is necessary for the interplay of emotion and group to emerge. Even though the experiments by Weisbuch and Ambady and Craig et al. (2014) do not seem to differ in regard to the social context of their experiments, we believe a fair test of the "interaction hypothesis" should put the task into an interactive context.

Taken together, based on these assumptions we argue that it was warranted to re-examine the interaction between group membership and emotional expression in an experiment with sufficient power, priming-specific parameters, a relevant out-group, and an interactive context. To this end, we conducted three experiments.

1. Overview

In three experiments, we used facial images of White-Caucasian and Middle-Eastern young men, showing a happy or fearful emotional expression. Group choice was based on previous findings that in Germany, men of Middle-Eastern appearance are evaluated more negatively than White-Caucasian men (e.g., Degner & Wentura, 2011; Degner, Wentura, Gniewosz, & Noack, 2007; Neumann & Seibt, 2001; Wagner, van Dick, Pettigrew, & Christ, 2003) and are associated with low warmth (e.g., Asbrock, 2010). Most importantly, Paulus and Wentura (2014) found an interaction between group membership and emotional expression using the exact same materials, albeit in an approach/avoidance task.

Experiment 1 tested which of the effects reported by Craig et al. (2014) and Weisbuch and Ambady (2008) would replicate in a standard evaluative priming task with a relevant out-group. However, in order to boost the (potential) influence of group membership, and in accordance with Craig et al., we made group membership more salient by administering a group membership categorization task that preceded the evaluative priming task. This allowed us to enhance race salience while keeping the priming task in its usual format.

To foreshadow, the results were clear-cut: both emotional expression and group membership produced priming effects (in the expected direction); there was no hint of an interaction. Thus, results were comparable to those obtained by Craig et al. (2014) with their race-focused task. In Experiment 2, we therefore discarded the group membership task preceding the evaluative priming task to see whether we can obtain a group effect without any emphasis on group membership. Moreover, we attempted to give the task a social dimension by using human trait adjectives as targets. We found essentially the same results as in Experiment 1, that is, both emotion as well as group membership produced priming effects (in the expected direction). Thus, this experiment showed an automatic prejudice effect without a task emphasis on group membership (i.e., race). Again, the interaction failed to be significant. In Experiment 3, we replicated Experiment 2 but created an even stronger social-evaluative context, in order to put the interaction hypothesis to a final test.

In all three experiments, we used the response-window version of the evaluative priming paradigm (Greenwald, Draine, & Abrams, 1996; Draine & Greenwald, 1998). In this variant, participants are instructed to respond during a rather short time window. Importantly, this implements a response deadline that is shorter than typical response times, thus producing quite large error rates. The error rates then constitute the main dependent variable. The rationale is that the response window homogenizes participants' choice of speed-accuracy criterion and thereby avoids a dispersion of priming effects across response-time and error measures. Although the technique was initially introduced for the exploration of masked priming effects (i.e., primes of which participants are unware), it can be successfully used with visible primes as well. The technique has been successfully employed in basic studies on evaluative priming (de Paula Couto & Wentura, 2017; Klauer, Mierke, & Musch, 2003; Klauer, Musch, & Eder, 2005) as well as in studies on implicit attitudes (Cunningham, Preacher, & Banaji, 2001; Degner & Wentura, 2009; Degner & Wentura, 2010, 2011; Degner et al., 2007; Frings & Wentura, 2003; Otten & Wentura, 1999; Wentura, Kulfanek, & Greve, 2005). We decided to use a stimulus onset asynchrony (SOA) of prime and target of 200 ms (see Wentura & Degner, 2010), since this SOA typically leads to robust priming effects.

All the data as well as the material is openly accessible at https://osf.io/9p6t5. We report all measures, manipulations and exclusions for our studies.

2. Experiment 1

In Experiment 1, we made group membership more salient compared to a standard evaluative priming task. Unlike Craig et al. (2014), who used a race-focused task, we enhanced salience by means of a group categorization task that preceded the evaluative priming task, in order to keep the priming task in its usual format.

2.1. Method

2.1.1. Participants

Participants were 83 non-psychology undergraduate students (41 females, 42 males; age $Md = 22$ years, range: 18–31) from Saarland University, recruited on campus and via an electronic sign-up system. Participants received four euros for participation. All participants were White-Caucasian and native speakers of German.[4] Experimenters were explicitly instructed to recruit native speakers and White-Caucasian participants. However, they did not directly address these restrictions during the recruitment process in order to not direct attention to the intergroup context of our experiments. Therefore, accidently, eight non-native speakers and three non-Caucasian participants were also allowed to participate, but their data were discarded.

For power calculations, we used two anchor values: First, we

[4] Being native speaker of German was an a priori criterion to secure full understanding of instructions and target words in Experiment 2 and 3. To keep inclusion criteria constant across experiments, we also applied this criterion to Experiment 1.

calculated a reasonable estimate[5] for the 2 (group membership) × 2 (emotional expression) interaction effect found in Weisbuch and Ambady's (2008) Experiment 1 (i.e., the experiment using supraliminally presented primes) as $d_Z = 0.40$. Second, the emotion effect on priming scores in Experiment 2 of Craig et al. (2014; i.e., the experiment using fear and joy as primes) was estimated as $d_Z = 0.58$. With $N = 83$, an effect of $d_Z = 0.40$ can be detected with $1 - \beta = 0.95$ ($\alpha = 0.05$). Power calculations were done with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) and the sample size was determined before any data analysis.

2.1.2. Design

We employed a 2 (prime group membership: Middle-Eastern vs. White-Caucasian) × 2 (prime emotional expression: happy vs. fearful) × 2 (target valence: positive vs. negative) design, with all factors varied within participants.

2.1.3. Materials

Prime stimuli were taken from Paulus and Wentura (2014). For the evaluative priming task, we selected fear and happiness expressions from 10 White-Caucasian men and 10 Middle-Eastern men. For the initial group categorization task, we selected neutral expressions of the same individuals. Pictures were taken from the Radboud Faces Database (Langner et al., 2010), the Amsterdam Dynamic Facial Expression Set (van der Schalk et al., 2011), and our own collection (Paulus, Rohr, Neuschwander, Seewald, & Wentura, 2012; for details of the selection procedure see Paulus & Wentura, 2014). All images were headshots with a straight head orientation and gaze directed at the viewer. Images were edited in a way that the face and top of the neck were shown on a white background. They measured 400 × 530 pixels and were presented on a CRT display set to a resolution of 1024 × 786 pixels.

Target stimuli were taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008). We selected 10 positive and 10 negative images according to norm values of valence and arousal.[6] We only selected pictures with a non-social content. The size of the presented images was 400 × 300 pixels.

2.1.4. Procedure

Participants were tested in groups of two to six; they were seated at individual computers, separated by partition walls. The experiment was implemented in E-Prime (Version 2.0) and presented on standard PCs with 17″ CRT monitors with a refresh rate of 100 Hz. All instructions were given on the computer screen.

In the initial group categorization task, participants were informed that faces of young men would be presented on the screen and that their task was to decide as quickly as possible whether the depicted person was of Turkish/Arabic or German/Western-European descent. The exact procedure was as follows: first, a white circle was presented on a black background at the bottom of the screen. Participants were instructed to click on the circle with the mouse. As soon as the circle was clicked on, a face image was presented in the middle of the screen and the labels “German/Western-European” and “Turkish/Arabic” appeared above and below the circle. Assignment of labels to positions was counterbalanced. Participants' task was to select the correct label as quickly as possible by clicking on it. As soon as the label was clicked on, the face image disappeared. The next trial started with presentation of a white circle again. The categorization task comprised 40 trials (i.e., each face was presented twice).

In the priming task, participants were told that they had to classify briefly presented pictures based on valence. The responses “negative” and “positive” were assigned to the A-key and to the L-key of a standard German QWERTZ keyboard. Response assignment was counterbalanced. The priming task used the response-window technique developed by Draine and Greenwald (1998). The beginning of a trial was marked by a centrally displayed fixation cross that remained on the screen for 1000 ms. It was replaced by the prime stimulus, which was presented for 100 ms, followed by a 100 ms blank screen and the target picture (i.e., the SOA was 200 ms). The prime stimulus was therefore clearly visible (i.e., presented supraliminal). The target remained on screen for 300 ms and was replaced by three blue exclamation marks that signaled the beginning of the response window. The response interval was 150 ms; participants were instructed to press the correct key within the response window (i.e., 300–450 ms post target onset). If the response fell in the response window, the exclamation marks changed color (from blue to yellow), indicating a valid response. The yellow exclamation marks were shown for 250 ms. The change of color occurred irrespective if the answer was correct or not. If the response was too slow, the exclamation marks disappeared (without changing color). If a response was given too quickly, the target disappeared immediately and a blank screen was shown for 1300 ms. A new trial started after an inter-trial-interval of 1000 ms (irrespective of response). The response window was automatically adjusted to individual performance after each block of trials. That is, if the error rate in one block was above 45% and the RT median was 100 ms or more above the midpoint of the response window (i.e., initially, > 475 ms), the response window was delayed by 33 ms. If, however, the error rate was below 20% and the RT median was below the response window midpoint plus 100 ms (i.e., initially, < 475 ms), the response window was brought forward by 33 ms.

Participants first completed two blocks of 20 practice trials, with each target picture presented once. Prime images were the White-Caucasian and Middle-Eastern neutral-expression faces from the group categorization task. Practice blocks served not only to familiarize participants with the task, but also to train their ability to hit the response window and to allow for response-window adjustment. During the practice blocks, participants received direct error feedback after each trial.

The main part of the priming task consisted of 320 trials, separated into eight blocks. Across all blocks, each prime stimulus appeared four times with a negative and a positive target, respectively. No trial-by-trial feedback was provided during experimental blocks, but summarized feedback was given at the end of each block, detailing the rate of correct responses, rate of accurately timed responses, and median RT for the just-completed block. After the priming task, demographics were obtained via questionnaire. Finally, participants were fully debriefed and thanked for their participation. Debriefing included a funneled debriefing in which we asked participants about their perceptions of the procedure as well as their hypotheses regarding the experiment. These data were not analyzed.

2.2. Results

To correct for anticipatory responses and lapses of attention, trials with response latencies below 100 ms or above 1000 ms were excluded from analyses (8.7% of all trials). The main analysis focused on the error rates. Reaction times were only analyzed for exploratory reasons. Mean error rate was 18.0% ($SD = 9.9$). Mean error rates and response times across all conditions are provided in Table A.1 (Appendix).

Priming scores (Craig et al., 2014; Weisbuch & Ambady, 2008) were calculated by subtracting average error rates of positive targets from average error rates of negative targets for each prime category (i.e., there were priming scores for happy and fearful Caucasian and Middle-

[5] Unfortunately, Weisbuch and Ambady (2008) only reported the overall 3 (emotion: happy vs. neutral vs. fear) × 2 (race) interaction, followed by tests of simple race effects for the different emotions, but not the crucial 2 (emotion: happy vs. fear) × 2 (race) interaction. However, making the plausible assumption that the difference score variable RT(Happy/White) − RT(Happy/Black) is uncorrelated with the difference score variable RT(Fearful/White) − RT(Fearful/Black), one can infer $d_Z = 0.40$ on the basis of the two reported *t*-values.

[6] The final selection included IAPS pictures # 1440, 1460, 1604, 1710, 1750, 5200, 5825, 5829, 5831, and 5833 for positive images and # 1111, 1202, 1275, 1525, 9000, 9295, 9300, 9570, 9830, and 9901 for negative images.

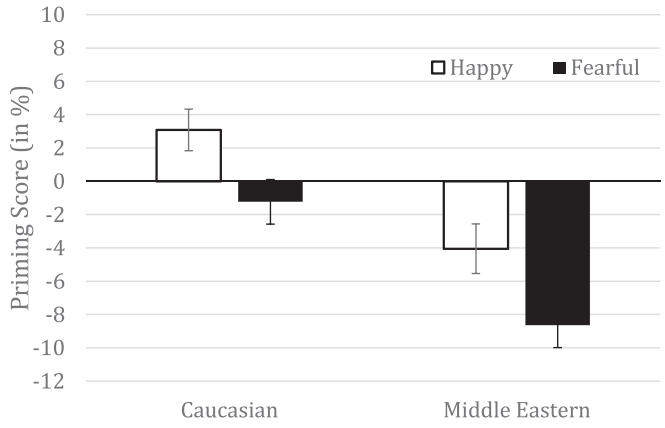

Fig. 1. Priming scores (i.e., error rates for negative targets minus error rates for positive targets) across emotional expression and group membership conditions (Experiment 1; error bars depict one standard error above/below the mean).

Eastern faces). Positive scores indicate relatively positive evaluations and negative scores indicate relatively negative evaluations.[7]

Fig. 1 shows the priming scores for the 2 × 2 prime conditions. A 2 (group: Middle-Eastern vs. White Caucasian) × 2 (emotion: happy vs. fearful) repeated measures ANOVA with priming scores as the dependent variable yielded two significant main effects, $F(1,82) = 26.03$, $p < .001$, $\eta_p^2 = .241$ ($d_Z = 0.56$: 95% CI [0.36–0.73])[8] for group, and $F(1,82) = 21.68$, $p < .001$, $\eta_p^2 = .209$ ($d_Z = 0.51$; 95% CI [0.31–0.72]) for emotion, but no indication of an interaction, $F < 1$. That is, we found the expected evaluative priming effect for emotions: happy faces (relatively) facilitated responses to positive target pictures; fearful faces facilitated responses to negative target pictures (in relative terms). Second, collapsed across emotions, the mean priming score for White-Caucasian faces exceeded the effect for Middle-Eastern faces. That is, we found the expected, arguably prejudice-related, priming effect for groups: relatively speaking, White Caucasian faces facilitated responses to positive pictures; Middle-Eastern faces facilitate responses to negative target pictures.

A corresponding analysis on mean RT priming scores (only correct trials with the same outlier correction as the error rates; see Table A.1) yielded two significant main effects as well, $F(1,82) = 8.12$, $p = .006$, $\eta_p^2 = .090$ ($d_Z = 0.31$; 95% CI [0.08–0.55]) for group, and $F(1,82) = 23.46$, $p < .001$, $\eta_p^2 = .222$ ($d_Z = 0.53$; 95% CI [0.34–0.72]) for emotion; these main effects corresponded in direction to the ones found for errors. There was no indication of an interaction, $F < 1$.

2.3. Discussion

The results of Experiment 1 were clear-cut, and suggest that in a standard evaluative priming task, both the emotional expression and the group membership of facial expressions are automatically extracted, influencing target responses. Both effects were in the expected direction: Results indicated that happy faces as well as White-Caucasian faces are automatically evaluated as more positive compared to fearful and Middle-Eastern faces, respectively. However, there was not the slightest hint of an interactive influence of the two factors along the lines of Weisbuch and Ambady (2008) findings.

The results were more in line with the pattern found by Craig et al. (2014) with their race-focus task. However, we found the group membership effect even in a standard evaluative priming task, that is, in a task that does not direct attention to group membership (i.e., race). Nevertheless, it might be that we induced a comparable attentional focus on race by means of the group categorization task that preceded the priming task. In order to examine if a group-membership priming effect would occur without attention being directed to this factor, we repeated the experiment without the group categorization task.

Furthermore, the targets in Experiment 2 were adjectives describing human traits; this was done to add a subtle social dimension to the priming task. As we argued before, the interaction hypothesis proposed by Weisbuch and Ambady (2008) puts the evaluation of the faces in a social context: The authors argued that participants evaluated the faces in terms of their social significance. In other words, this assumes that for the interaction between group membership and emotional expression to arise, participants have to ask themselves (at least implicitly) what implication a face might carry for them. We assumed that the evaluation of typical human traits as good or bad would induce a social-interactive perspective.[9]

Additionally, we introduced type of adjective as a further factor into the design. Peeters (1983; Peeters & Czapinski, 1990) highlighted the fact that the valence of an attribute depends on perspective. For example, an attribute like *intelligent* is unconditionally positive from the perspective of the *possessor* (i.e., the intelligent person), but intelligence can be misused and is thus not unequivocally positive for the intelligent person's social environment. Likewise, an attribute like *tolerant* is unconditionally positive for *others* interacting with the tolerant person, but not necessarily for the tolerant person themselves, as their tolerance may be taken advantage of. The same holds for negative attributes. For example, an attribute like *lonely* is unconditionally negative from the perspective of the lonely person, whereas an attribute like *mean* is unconditionally negative from the perspective of others interacting with the mean person. Attributes such as *intelligent* and *lonely* can thus be termed possessor-relevant, whereas attributes such as *tolerant* and *mean* can be termed other-relevant (Wentura, Rothermund, & Bak, 2000). We were interested if type of adjective would further moderate the (expected) interaction between group membership and emotional expression.

3. Experiment 2

3.1. Method

3.1.1. Participants

Participants were 83 non-psychology undergraduate students (40 females, 43 males; age *Md* = 22.5 years, range: 18–32) from Saarland University, recruited on campus and via an electronic sign-up system. Participants received four euros for participation. Following the procedure outlined in Experiment 1, the data of seven participants were excluded because they were not German native speakers ($n = 1$) or White-Caucasian ($n = 5$), or because the rate of trials with response times in the range of 100 to 1000 ms was below 60% ($n = 1$).

With regard to test power, power was (again) oriented on $d_Z = 0.40$ (i.e., the interaction effect found by Weisbuch & Ambady, 2008). The effect sizes of both effects found in our Experiment 1 (i.e., the main effects of emotion and ethnicity) exceeded this value. Thus, there was no need to increase sample size with regard to a replication of these effects. Power calculations were done with G*Power (Faul et al., 2007) and the sample size was determined before any data analysis.

3.1.2. Design, materials, and procedure

Design, materials, and procedure were identical to Experiment 1 with the following two exceptions: (1) The group categorization task

[7] Emphasis is on "relatively" because due to a potential target main effect – for example, a bias towards more errors for positive targets, irrespective of prime type – these priming scores cannot be interpreted in absolute terms.

[8] We used bootstrapping (using the R package BootES, Kirby & Gerlanc, 2013) with 5000 resamples to estimate the confidence intervals.

[9] It should be noted that Craig et al. (2014) also used adjectives as targets (drawn from Fazio et al.'s (1995) evaluative priming study). However, those were not predominantly human trait adjectives.

Table 1
Priming scores (i.e., error rates for negative targets minus error rates for positive targets in %; standard errors in parentheses) as a function of prime group, prime emotion, and target relevance (Experiment 2).

	Prime			
	Caucasian		Middle-Eastern	
	Happy	Fearful	Happy	Fearful
Target				
Possessor-relevant	17.9 (2.0)	6.3 (2.1)	11.7 (2.0)	4.1 (1.8)
Other-relevant	9.3 (1.8)	−8.7 (2.0)	3.0 (2.1)	−11.6 (2.2)

preceding the priming task was discarded. (2) Positive and negative adjectives were used as targets. Target words were either possessor-relevant or other-relevant (Peeters, 1983; see also Degner & Wentura, 2011; Wentura et al., 2000). However, participants were again instructed to categorize the valence of the targets. This extension resulted in a 2 (prime group membership: Middle-Eastern vs. White-Caucasian) × 2 (prime emotional expression: happy vs. fearful) × 2 (target valence: positive vs. negative) × 2 (target relevance: possessor vs. other) design, with all factors varied within participants. There were five target words for each combination of valence and relevance. Target words were taken from Degner and Wentura (2011); two words were replaced due to their semantic neighborhood to emotion terms (see Table A.2 in the Appendix section).

3.2. Results

Data trimming followed the procedure of Experiment 1 (6.0% of all trials were excluded because response latencies were below 100 ms or above 1000 ms). Mean error rate was 30.5% (SD = 9.0). Mean error rates and response times across all conditions are provided in Table A.1 (Appendix).

Priming scores (i.e., error rates for negative targets minus error rates for positive targets) are given in Table 1. A 2 (group: Middle-Eastern vs. White-Caucasian) × 2 (emotion: happy vs. fearful) × 2 (target relevance: possessor vs. other) repeated measures ANOVA with priming scores as the dependent variable again yielded significant main effects of group and emotion, $F(1,82) = 14.87$, $p < .001$, $\eta_p^2 = .153$ ($d_Z = 0.42$; 95% CI [0.20–0.64]) for group, and $F(1,82) = 73.23$, $p < .001$, $\eta_p^2 = .472$ ($d_Z = 0.94$; 95% CI [0.73–1.15]) for emotion. The two-way interaction between group membership and emotional expression missed the criterion of significance, $F(1,82) = 3.00$, $p = .087$, $\eta_p^2 = .035$ ($d_Z = 0.19$; 95% CI [−0.03–0.40]). This pattern was not moderated by a three-way interaction, $F < 1$. Fig. 2 shows the mean priming scores for this (non-significant) interaction collapsed over target relevance.

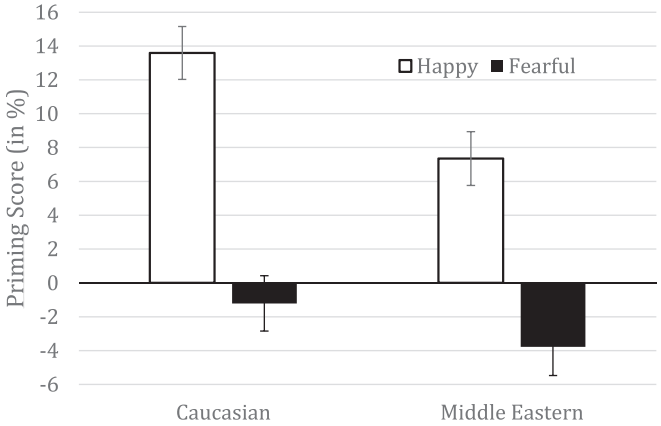

Fig. 2. Priming scores (i.e., error rates for negative targets minus error rates for positive targets) across emotional expression and group membership conditions (Experiment 2; error bars depict one standard error above/below the mean).

The pattern of results is essentially the same as in Experiment 1: First, collapsed over groups, the mean priming score for happy faces exceeded the score for fearful faces. That is, we found the expected evaluative priming effect for emotion: happy faces (relatively) facilitated responses to positive target words; fearful faces (relatively) facilitated responses to negative target words (in relative terms). Second, collapsed over emotions, the mean priming score for White-Caucasian faces exceeded the one for Middle-Eastern faces. That is, we found the expected prejudice-related priming effect for group: White-Caucasian faces facilitated responses to positive target words; Middle-Eastern faces facilitated responses to negative target words.

With some caution, one might interpret the interaction between group membership and emotional expression. There are two reasons for this: (1) The interaction test in a within-participants design with a numerator df of 1 is formally equivalent to a one-sample t-test (with t = squareroot(F) and $p_F = p_{t,\ two\text{-}tailed}$). In order to transform the interaction result of the F-test in a t-test we have to calculate the critical difference variable (i.e., priming scores difference happy-fearful for Whites minus priming scores difference happy-fearful for Middle-Eastern), and test it against zero. This results in $t(82) = 1.73$, $p = .044$ (one-tailed). (2) The direction of this difference variable corresponds to the effect found by Weisbuch and Ambady (2008). Thus, it could be considered significant in a one-tailed t-test. However, in contrast to Weisbuch and Ambady's finding, the interaction is only of the ordinal type: the emotion priming effect is smaller for Middle-Eastern faces compared to Caucasian faces, but it is significant for both groups, $t(82) = 8.44$, $p < .001$, $d_Z = 0.93$ (95% CI [0.72–1.13]) for Caucasian faces, and $t(82) = 5.73$, $p < .001$, $d_Z = 0.63$ (95% CI [0.42–0.83]) for Middle-Eastern faces.

As mentioned earlier, target relevance did not qualify the interaction between group and emotion. However, the emotion priming effect was significantly moderated by target relevance, $F(1,82) = 12.64$, $p < .001$, $\eta_p^2 = .134$ ($d_Z = 0.39$; 95% CI [0.15–0.61]); for the sake of completeness: $F < 1$ for the interaction of group and target relevance). The emotion priming effect was larger for other-relevant targets compared to possessor-relevant targets; however, it was significant for both types, $t(82) = 8.60$, $p < .001$, $d_Z = 0.94$ (95% CI [0.71–1.19]) for other-relevant targets, and $t(82) = 5.78$, $p < .001$, $d_Z = 0.63$ (95% CI [0.42–0.85]) for possessor-relevant targets.

A corresponding analysis on mean RT priming scores (see Table A.1) yielded a non-significant main effect of group, $F < 1$, and a significant main effect of emotion, $F(1,82) = 20.50$, $p < .001$, $\eta_p^2 = .200$ ($d_Z = 0.50$; 95% CI [0.29–0.70]). This main effect corresponded in direction to the one found for errors. The main effects of group and emotion were not qualified by relevance, both Fs < 1. The two-way interaction of group and emotion was non-significant, $F(1,82) = 2.05$, $p = .156$, $\eta_p^2 = .024$ ($d_Z = 0.16$; 95% CI [−0.07–0.36]); it was not further qualified by relevance, $F < 1.15$.

3.3. Discussion

Experiment 2 yielded essentially the same results as Experiment 1. Both the variation in group membership as well as the variation in

emotional expression produced priming effects in the expected direction. Numerically, the group priming effect was a bit smaller in Experiment 2 compared to Experiment 1, a result which might be attributed to the discarding of the group categorization phase in Experiment 2. We will return to this issue in the General discussion section.

A slight, but potentially meaningful difference between the two sets of results concerns the interaction between group membership and emotional expression. In Experiment 2, the emotion-related priming effect (i.e., the priming effect showing an automatic positive evaluation of happy expressions relative to fearful expressions) tended to be smaller for out-group faces than in-group faces; this pattern fits with the results of Weisbuch and Ambady (2008). However, the interaction is far from the disordinal form that made Weisbuch and Ambady's results so striking.

One potential reason for an interaction pattern to emerge might lie in the social dimension of Experiment 2's priming task. In contrast to Experiment 1, the target stimuli in Experiment 2 were adjectives describing human traits, and participants evaluated these adjectives as positive or negative. It is reasonable to assume that this task activated a social-interactive mindset in participants, in which evaluation of interaction partners might play a fundamental role. However, as this interaction was not significant in Experiment 2, we decided to examine in a further experiment whether the interaction would be strengthened if the social character of the evaluation was more strongly emphasized in the task.

4. Experiment 3

Experiment 3 was a replication of Experiment 2 with two changes: First, we restricted the target word set to other-relevant adjectives. Second, and more important, we instructed participants to categorize targets from an other-relevant perspective.

4.1. Method

4.1.1. Participants

Participants were 88 non-psychology undergraduate students (46 females, 42 males; age *Md* = 22 years, range: 18–33) from Saarland University, recruited on campus and via an electronic sign-up system. Participants received four euros for participation. The data of five participants was excluded because German was not their mother tongue.

With regard to test power, power was (again) oriented on $d_Z = 0.40$ (i.e., the interaction effect found by Weisbuch and Ambady, 2008). The effect sizes of both effects found in our Experiment 1 and Experiment 2 (i.e., the main effects of emotion and ethnicity) exceeded this value. Thus, there was no need to increase sample size with regard to a replication of these effects. The slight increase from $N = 83$ [Exp. 1, 2] to $N = 88$ was only due to pragmatic reasons of recruitment: Since, typically, a few recruited participants do not show up for an experimental session, experimenters recruited a surplus of participants. Power calculations were done with G*Power (Faul et al., 2007) and the sample size was determined before any data analysis.

4.1.2. Design, materials, and procedure

Design, materials, and procedure were identical to Experiment 2 with the following two exceptions: First, we only used words of the other-relevant type, adding five positive and five negative adjectives to the target set of Experiment 2. Second, the evaluative priming task focused more explicitly on the evaluation of a social context. Thus, we informed participants that the "…adjectives name person attributes. Your task is to decide whether it would be good or bad for you if your interaction partner had this attribute. Thus, imagine that the adjectives refer to attributes of persons with whom you interact. The adjectives might name good (e.g., "peaceful") or bad (e.g., "aggressive") attributes of your interaction partner." This instruction explicitly directed participants' attention to the evaluation of a social interaction. Participants were instructed to press a "good" or "bad" key accordingly.

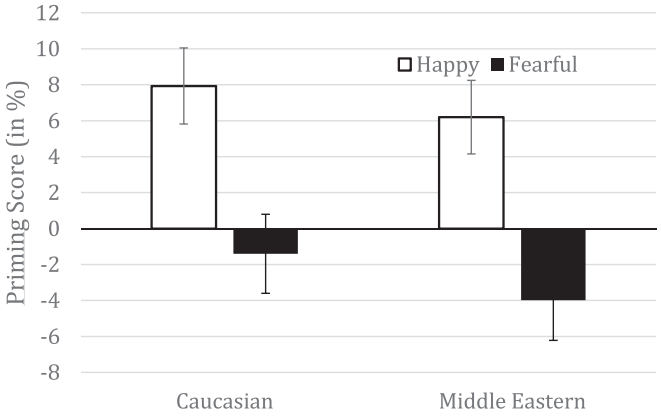

Fig. 3. Priming scores (i.e., error rates for negative targets minus error rates for positive targets) across emotional expression and group membership conditions (Experiment 3; error bars depict one standard error above/below the mean).

4.2. Results

Data trimming followed the procedure of Experiments 1 and 2 (4.6% of all trials were excluded because response latencies were below 100 ms or above 1000 ms). Mean error rate was 31.1% (*SD* = 8.7). Mean error rates and response times across all conditions are provided in Table A.1 (Appendix).

Priming scores (i.e., error rates for negative targets minus error rates for positive targets) are depicted in Fig. 3. A 2 (group: Middle-Eastern vs. White Caucasian) × 2 (emotion: happy vs. fearful) repeated measures ANOVA with priming scores as the dependent variable yielded a main effects of group which was associated with $F(1,87) = 3.20$, $p = .078$, $\eta_p^2 = .035$ ($d_Z = 0.19$; 95% CI [−0.03–0.39]) and a significant main effect for emotion, $F(1,87) = 38.27$, $p < .001$, $\eta_p^2 = .305$ ($d_Z = 0.66$; 95% CI [0.47–0.85]) for emotion. The two-way interaction was non-significant, $F < 1$. We consider the main effect of group to be significant since the main effect test in a *F*-test in a within-participants design with numerator *df* of 1 directly corresponds to a *t*-test (with t = squareroot(F) and $p_F = p_{t,\ two\text{-}tailed}$) comparing the mean priming score for White primes with the mean priming score for Middle Eastern primes (collapsed across emotions). Thus, since this difference has a positive mean (which is the expected sign for the group priming effect), it is significant in a one-tailed *t*-test, $t(87) = 1.79$, $p = .039$ (one-tailed).

That is, first, we found the expected evaluative priming effect for emotion: happy faces facilitated responses to positive target words; fearful faces facilitated responses to negative target words. Second, we found the expected prejudice-related priming effect for group: White-

Caucasian faces (relatively) facilitated responses to positive target words; Middle-Eastern faces (relatively) facilitated responses to negative target words.

A corresponding analysis on mean RT priming scores (see Table A.1) yielded a non-significant main effect of group, $F(1,87) = 2.19$, $p = .071$ (one-tailed), $\eta_p^2 = .025$ ($d_Z = 0.16$; 95% CI [−0.05–0.37]), and a significant main effect of emotion, $F(1,87) = 16.07$, $p < .001$, $\eta_p^2 = .156$ ($d_Z = 0.43$; 95% CI [0.21–0.63]). Both main effects corresponded in direction to the effects for errors. The two-way interaction was non-significant, $F < 1$.

4.3. Discussion

The results were again essentially the same as in Experiments 1 and 2: the factors group membership and emotional expression produced independent priming effects, with one—emotion—reflecting the relative positivity of happy facial expressions compared to fearful expressions and the other—group—reflecting the relative positivity of in-group members compared to out-group members. Admittedly, the group effect was somewhat smaller in Experiment 3 compared to the preceding experiments. We will report across-experiments analyses in the next section to clarify whether we should assume meaningful differences between experiments.

Notably, the two-way interaction between group membership and emotional expression was again clearly non-significant. Thus, the altered instruction for the evaluative decision task that emphasized the social character of the evaluation did not strengthen the marginal interaction pattern found in Experiment 2, and thus we did not find the disordinal interaction reported by Weisbuch and Ambady (2008).

5. Across-experiments analyses

To get a better indication of which between-experiments differences deserve further discussion, we conducted a 2 (group: Middle-Eastern vs. White-Caucasian) × 2 (emotion: happy vs. fearful) × 3 (experiment) mixed ANOVA, with the first two factors varied within participants and experiment as a between-participants factor, and priming scores as the dependent variable. The analysis yielded significant main effects of group and emotion, $F(1,251) = 39.93$, $p < .001$, $\eta_p^2 = .137$ ($d_Z = 0.39$; 95% CI [0.27–0.50]) for group, and $F(1,251) = 128.16$, $p < .001$, $\eta_p^2 = .338$ ($d_Z = 0.69$; 95% CI [0.58–0.80]) for emotion. The group × emotion interaction was non-significant, $F < 1$. Both main effects were significantly moderated by experiment, $F(2,251) = 4.32$, $p = .014$, $\eta_p^2 = .033$ for the group × experiment interaction, and $F(2,251) = 9.39$, $p < .001$, $\eta_p^2 = .070$ for the emotion × experiment interaction. The three-way interaction was non-significant, $F(2,251) = 1.57$, $p = .210$, $\eta_p^2 = .012$.

To further explore these interactions, we conducted two planned orthogonal contrasts (Helmert) for both the group × experiment and the emotion × experiment effects. Experiment 1 differed from Experiments 2 and 3 in target modality and the presence of a group categorization task; furthermore, Experiments 2 and 3 both featured a (slight) social task dimension. Thus, Helmert interaction contrast 1 compared Experiment 1 with Experiments 2 and 3 collapsed. This contrast was significant for the group × experiment interaction, $t(252) = 2.69$, $p = .008$, $d = 0.36$ (95% CI [0.10–0.63]), as well as the emotion × experiment interaction, $t(238.423) = 4.650$, $p < .001$, $d = 0.53$ (95% CI [0.32–0.74]), with the emotion effect being smaller in Experiment 1 compared to Experiments 2/3 and the group effect being larger in Experiment 1 compared to Experiments 2/3. Helmert contrast 2 compared Experiment 2 with Experiment 3, that is, the two word-target experiments without a group categorization phase. This contrast was non-significant for both the group × experiment interaction, $t(169) = 1.28$, $p = .202$, $d = 0.20$ (95% CI [−0.11–0.50]), and the emotion × experiment interaction, $t(169) = 1.53$, $p = .129$, $d = 0.23$ (95% CI [−0.08–0.54]). For Experiments 2 and 3 collapsed, the emotion priming effect was $d_Z = 0.79$ (95% CI [0.65–0.93]) and the group-based effect was $d_Z = 0.30$ (95% CI [0.15–0.44]).

In light of these findings, how should we evaluate the empirical evidence with regard to our central research questions? They were: (1) Is there evidence for an interactive influence of group membership and emotional expression in the evaluative priming paradigm, along the lines of Weisbuch and Ambady (2008) finding? (2) Is there evidence for a group membership priming effect in a standard form of the evaluative priming task (i.e., a version without a group focus task as used by Craig et al., 2014)?

To answer these questions, we calculated Bayes factors on the basis of all three experiments (and additionally, for the second research question, just based on Experiments 2 and 3, as these had no group categorization phase). With regard to our first research question, the Bayes factor in favor of the null hypothesis (i.e., “there is no interaction between group membership and emotional expression”) was $BF_{01} = 10.88$. This can be considered “strong evidence” according to Jeffreys (1961, p. 432; see also Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). With regard to our second research question, the Bayes factor in favor of the presence of a group membership priming effect was $BF_{10} = 3.0 \times 10^6$ based on all three experiments, and $BF_{10} = 98.31$ based on only Experiments 2 and 3. The former can be considered “extreme evidence,” the latter “very strong evidence” (Jeffreys, 1961; Wagenmakers et al., 2011).[10]

6. General discussion

We started with the observation that only few studies have examined how faces varying on more than one evaluative dimension influence automatic evaluative responses. Two recent articles that examined this issue (without directing participant attention to one of the two dimensions) left readers with conflicting conclusions. While Weisbuch and Ambady (2008) found evidence that group membership and emotional expression are instantaneously amalgamated in the evaluation process, the results of Craig et al. (2014) suggested that group membership only enters the evaluation process if explicit membership categorization is added to the evaluative priming task.

We reinvestigated the issue in three well-powered experiments. The experiments were alike in several aspects: Methodologically, we always employed a response-window version of an evaluative priming task with an SOA of 200 ms. Based on the literature, these choices are adequate in order to obtain evaluative priming effects in general, and involuntary, non-strategic evaluation of primes in particular (Degner, 2009; Wentura & Degner, 2010, but see Teige-Mocigemba & Klauer, 2013). Across all experiments, we used the evaluative priming paradigm in its standard form (i.e., without a secondary task directing attention to one of the two evaluative dimensions, as in Craig et al., 2014), to test if group membership effects would emerge in this task. Finally, in all three experiments we used prime faces from an in-group (White-Caucasian) and an out-group (Middle-Eastern) that were relevant to our (German) sample. Prejudice-related evaluative priming effects with these groups had been found before, using neutral-looking faces (Degner & Wentura, 2011). Also, the exact same stimuli as in the present experiments were used before in a study with approach-avoidance tasks, which yielded results that were analogous to Weisbuch and Ambady (2008) findings (Paulus & Wentura, 2014).

The differences between experiments can be characterized in three

[10] With regard to the interaction, the Bayes factor in favor of the null hypothesis for the three experiments were $BF_{01} = 8.18$, $BF_{01} = 1.99$, $BF_{01} = 7.82$ in Experiments 1, 2, and 3, respectively.
With regard to the group effect, the Bayes factor in favor of the presence of this effect for the three experiments were $BF_{10} = 7.1 \times 10^3$, $BF_{10} = 91.75$, $BF_{10} = 0.54$ in Experiments 1, 2, and 3, respectively. For the sake of completeness, the Bayes factors in favor of the presence of an emotion effect were $BF_{10} = 1.4 \times 10^3$ (Exp. 1), 1.4×10^{10} (Exp. 2), 6.0×10^5 (Exp. 3), and 4.0×10^{16} (total sample).

ways: First, in Experiment 1 we used a group membership task that preceded the priming task, in order to make the group differences more salient to participants. This task was discarded in Experiments 2 and 3. Second, the evaluative decision task was entirely non-social in Experiment 1 (i.e., the IAPS pictures were of non-social character), it was implicitly social in Experiment 2 (i.e., all target words related to person attributes; the variation of possessor-relevant versus other-relevant attributes fostered a social evaluative stance), and it was explicitly social in Experiment 3 (i.e., participants were instructed to adopt an other-relevant perspective). Third, we used IAPS pictures (as did Weisbuch & Ambady, 2008) in Experiment 1 and adjectives (as did Craig et al., 2014) in Experiments 2 and 3.

Our results were clear-cut: In all three experiments, both group membership and emotional expression influenced evaluative responses (albeit the effect for group membership was only significant in a one-tailed test in Experiment 3). However, this influence emerged in the form of two independent main effects (with priming differences as the dependent variable): images of in-group members elicited more positive reactions than images of out-group members (irrespective of emotional expression) and individuals displaying happy faces elicited more positive reactions than individuals showing fearful faces (irrespective of group-membership). There was no evidence for an interaction between these two factors.

Regarding the size of the two main effects, emotional expression exerted a stronger influence on evaluative responses than group membership. Across Experiments 2 and 3 (i.e., the experiments without emphasis on group features), the emotion-based priming effect was "large" (according to Cohen, 1988; i.e., $d_Z = 0.79$), whereas the group-based effect was "small" to "medium" (i.e., $d_Z = 0.30$). Moreover, by comparing the (non-overlapping) 95% confidence intervals, it can be stated that the emotion-based effect is larger than the group-based effect. Thus, our results support the conclusion drawn by Craig and colleagues (p. 875): "Emotional expressions are more influential than group membership cues of race […] in the formation of implicit evaluations of faces varying on both of these dimensions." We believe that one potential reason for this finding is that emotional expressions (at least the full-fledged expressions used in the present study) are defined by clearly discriminable, qualitatively different feature configurations (e.g., the corners of the mouth either point "up" or "not up"), whereas group membership is defined by more quantitative shifts on feature continua (e.g., hair color varies between different shades of brown, with out-group members on average having darker hair). Therefore, on the individual picture, emotional expressions were more clearly distinguishable than groups. Future studies should examine if emotional expressions still yield stronger effects on evaluative responses than group membership if the discriminability of dimensions is equated (i.e., with more gradual feature variations between emotional expressions).

Despite the similarity between our results and those of Craig et al. (2014), the findings also differ in an important aspect: our studies yielded strong evidence that group membership (i.e., race) was involuntarily processed and evaluated, with a group priming effect (i.e., an in-group favoritism/out-group derogation effect) emerging in all three experiments. Most notably, this effect occurred not only in Experiment 1, where group membership was made salient by a group categorization task preceding the priming task, but also in Experiments 2 and 3, where group membership was not mentioned at all (albeit the effect for group membership was only significant in a one-tailed test in Experiment 3). This finding of a group effect replicates earlier results with emotionally neutral facial expressions (e.g., Degner & Wentura, 2011).

We believe there are (at least) two potential reasons for the difference between our findings and those reported by Craig et al. (2014): One mundane reason for the divergence might lie in test power. If we estimate the group effect size based on the pooled results from our Experiment 2 and 3 (i.e., $d_Z = 0.29$), Craig et al.'s power in their Experiment 1 (happy vs. angry; $N = 29$) and Experiment 2 (happy vs. fearful; $N = 32$) was $1 - \beta = 0.33$ and $1 - \beta = 0.36$, respectively (with $\alpha = 0.05$).[11] Moreover, Craig et al. (2014) themselves discussed the possibility that the particular out-group they used (African-American individuals) may have not been very salient for their White-Caucasian Australian sample. Their results suggest that this out-group is only evaluated negatively when attention is directed to the out-group status. The out-group we employed in our experiments (Middle-Eastern young men), however, might be more relevant for our sample and may have therefore elicited a negativity bias even if group membership was not made salient.

Our findings also differ from those observed by Weisbuch and Ambady (2008). Specifically, the disordinal interaction pattern found in all studies reported by Weisbuch and Ambady (2008) did not emerge in any of our experiments. Weisbuch and Ambady found that happy expressions shown by in-group members and fearful expressions shown by out-group members elicited (relatively) positive reactions, whereas fearful expressions shown by in-group members and happy expressions shown by out-group members elicited (relatively) negative reactions. In our experiments, only Experiment 2 produced a (marginally) significant interaction, and in contrast to Weisbuch and Ambady, the pattern of the interaction was ordinal, not disordinal. Moreover, since neither Experiment 1 nor Experiment 3 yielded an interaction, it is arguably more reasonable to trust the overall null result, as supported by the across-experiments Bayes factor analysis. Thus, based on our results and those by Craig et al. (2014), we have to conclude that emotional expression and group membership (i.e., race) do not interactively influence results in the evaluative priming paradigm.

However, the interaction hypothesis has been supported by different means (Weisbuch & Ambady, 2008, but also others, e.g., Paulus & Wentura, 2014), with evaluative priming being only one of those means. Thus, we need to discuss what the differences are between evaluative priming and, for example, the approach-avoidance task used by Paulus and Wentura (2014). We will elaborate on this point in the Theoretical implications section below. First, however, as a side note, we briefly discuss the results of Experiment 2 concerning the distinction of possessor-relevance and other-relevance.

In Experiment 2, target relevance did not moderate the group membership priming effect. Such a moderation, however, emerged in Degner and Wentura (2011; see also Degner et al., 2007). In that study, the prejudice-related priming effect regarding Middle-Eastern young men was constrained to other-relevant targets, suggesting that the automatic prejudice was hostile in nature. However, facial expressions in that study were emotionally neutral. In our Experiment 2, target relevance did not moderate the group membership effect but it did moderate the emotion priming effect: the emotion priming effect was larger for other-relevant words compared to possessor-relevant words. This seems in line with the notion that emotional expressions are evaluated in a socially relevant way. In other words, happy faces may not act predominantly as positive possessor-relevant primes (i.e., the expresser experiencing joy) but as positive other-relevant primes (i.e., the expresser signaling a benevolent environment); fearful faces may not act predominantly as negative possessor-relevant primes (i.e., the expresser experiencing fear) but as negative other-relevant primes (i.e., the expresser signaling a malevolent environment). Regarding our research question, this interpretation corroborates our assumption that the use of possessor-relevant and other-relevant adjectives can create a social-interactive context.

[11] We do not mean to accuse Craig et al. (2014) of conducting underpowered experiments. Their sample sizes allowed detection of "medium"-sized effects (Cohen, 1988) of $d_Z = 0.54$ and 0.51, respectively, with $1 - \beta = 0.80$ ($\alpha = 0.05$), which is a reasonable assumption in the absence of additional constraints.

7. Theoretical implications for further research

Based on the accumulated evidence regarding the influence of emotional expression and group membership on evaluative processes, which conclusion can be drawn? Are emotional expression and group membership evaluated independently or do they interact? How can the results observed in our experiments and by Craig et al. (2014) be reconciled with those found by Weisbuch and Ambady (2008) as well as Paulus and Wentura (2014)?

Craig et al. (2014) proposed that top down influences might serve to explain the differences between studies: They argued that factors such as task context or instructions can put emphasis on group membership (i.e., race) and thus boost the influence of this factor. However, while this argument might explain the strength of a group-based priming effect (see Craig et al., 2014, and the comparison of the present Experiment 1 with Experiments 2/3), neither Craig et al. nor we found evidence for an *interaction* between the two factors even in a context that placed clear emphasis on group membership (as indicated by the size of the group effect in our Experiment 1) or social context (Experiment 3). Therefore, we argue that the answer might instead lie in the time course of the contributing processes: while group membership and emotional expression are both processed promptly, the interactive representation of both dimensions might take time to establish. The assumption of an early processing of group membership and emotional expression is in accordance with Kubota and Ito (2007), who recorded event-related potentials and showed that race and facial expression cues are processed within the first 200 ms post stimulus onset, independently and in parallel. The role of time for person perception is also emphasized in Freeman and Ambady (2011) "dynamic interactive model of person construal," which states that person construal fluctuates over time with a stable representation needing more time to establish. Therefore, it is plausible that (relatively) fast evaluative responses, like those assessed in the evaluative priming paradigm, are independently influenced by cues indicating group membership and emotional expression, whereas slower processes, like approach and avoidance reactions and explicit evaluations, are influenced by an interactive representation of the two dimensions.

The assumption that the interactive representation of emotional expression and group membership takes time to establish has several possible implications: First, it suggests that when comparing the results found with different paradigms, the time factor should be taken into account. If the different paradigms assess the influence of evaluative factors on evaluative reactions at different time points, the results might be combined to gain a better understanding of the processes underlying the processing of stimuli defined by several evaluative dimensions. For example, the results found with the evaluative priming paradigm (relatively fast responses) and the approach and avoidance paradigm (relatively slow responses) might be considered conjointly instead of competitively to understand how the processing of faces differing in group membership and emotional expression evolves. A second possible implication is that in the case of stimuli defined by several evaluative features, *relatively* fast evaluative responses might be able to predict open behavior better than very fast evaluative responses. This should be the case because open behavior typically constitute relatively slow evaluative responses. If the interactive representation of two dimensions takes time to evolve, it is plausible that paradigms assessing the evaluative response very early in time capture a different representation than open behavior whereas the representation captured at later time points more closely resembles the one triggering the open behavior. Obviously, these possible implications need empirical examination.

8. Conclusion

In three well-powered experiments, we examined the research question how faces varying on more than one evaluative aspect influence evaluative reactions. The results of all three experiments were clear-cut: group membership as well as emotional expression influenced reactions in an evaluative priming paradigm. However, no interaction between these two features emerged. Our results therefore show that fast and (potentially) involuntary reactions are influenced by various aspects of a face.

Open practices

The experiments presented in this article earned Open Materials and Open Data badges for transparent practices.

Materials and data are available at https://osf.io/9p6t5.

Appendix

Table A.1
Mean error rates (in %) and mean reaction times (in milliseconds, in parentheses) as a function of prime group, prime emotion, and target emotion (Experiments 1 and 3) or target relevance (Experiment 2).

Target	Prime			
	Caucasian		Middle-Eastern	
	Happy	Fearful	Happy	Fearful
Experiment 1				
Positive	15.6 (387)	17.9 (391)	20.7 (389)	22.7 (397)
Negative	19.2 (407)	17.2 (398)	16.8 (400)	14.2 (396)
Experiment 2				
Possessor-relevant				
Positive	20.5 (408)	28.8 (418)	24.2 (413)	27.3 (417)
Negative	38.4 (433)	35.1 (423)	35.8 (432)	31.3 (427)
Other-relevant				
Positive	26.2 (422)	35.2 (426)	29.4 (424)	36.1 (428)
Negative	35.5 (427)	26.5 (421)	32.5 (426)	24.4 (422)
Experiment 3				
Positive	26.1 (413)	31.9 (418)	27.4 (416)	34.5 (421)
Negative	34.1 (424)	30.5 (420)	33.6 (424)	30.6 (418)

Table A.2
Target adjectives and their pleasantness norm values (Experiments 2 and 3).

	Possessor-relevant		Other-relevant	
Negative	leblos (lifeless)	−52	gierig (greedy)	−60
	unfähig (incompetent)	−50	grausam (cruel)	−84
	träge (lethargic)	−50	boshaft (malicious)	−72
	lustlos (listless)	−52	gemein (mean)	−66
	einsam (lonely)	−60	geizig (stingy)	−61
			aggressiv (aggressive)	−52
			kriminell (criminal)	−56
			autoritär (authoritarian)	−73
			brutal (brutal)	−87
			treulos (faithless)	−56
Positive	begabt (talented)	49	human (humane)	57
	schön (beautiful)	77	ehrlich (honest)	74
	gesund (healthy)	79	gütig (kind)	51
	aktiv (active)	60	gerecht (just)	60
	geschickt (skilled)	55	geduldig (patient)	68
			sanft (gentle)	56
			humorvoll (humorous)	54
			tolerant (tolerant)	68
			friedlich (peaceful)	71
			aufrichtig (sincere)	74

Note: The first 5 rows of negative and positive words, respectively, refer to the target words of Experiment 2 and were taken from Degner and Wentura (2011), except that *geschickt* replaced *heiter* (*serene*) and *geduldig* replaced *zärtlich* (*affectionate*). The complete list of other-relevant words was the target words of Experiment 3. Pleasantness norms are taken from Hager, Mecklenbräuker, Möller, and Westermann (1985), and Möller and Hager (1991); the scale ranges from −100 (very negative) to +100 (very positive).

References

Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology, 41*, 76–81. http://dx.doi.org/10.1027/1864-9335/a000011.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46*, 668–688. http://dx.doi.org/10.3758/s13428-013-0410-6.

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological Psychology, 77*, 343–352.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review, 16*, 330–350.

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *Quarterly Journal of Experimental Psychology: Section A, 58*, 1173–1197.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Craig, B. M., Lipp, O. V., & Mallan, K. M. (2014). Emotional expressions preferentially elicit implicit evaluations of faces also varying in race or age. *Emotion, 14*, 865–877. http://dx.doi.org/10.1037/a0037270.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 121*, 163–170.

Degner, J. (2009). On the (un)controllability of affective priming: Strategic manipulation is feasible but can possibly be prevented. *Cognition & Emotion, 23*, 327–354.

Degner, J., & Wentura, D. (2009). Not everybody likes the thin and despises the fat: One's weight matters in the automatic activation of weight-related social evaluation. *Social Cognition, 27*, 202–220.

Degner, J., & Wentura, D. (2010). Automatic prejudice in childhood and early adolescence. *Journal of Personality and Social Psychology, 98*, 356–374.

Degner, J., & Wentura, D. (2011). Types of automatically activated prejudice: Assessing possessor- versus other-relevant valence in the evaluative priming task. *Social Cognition, 29*, 182–209.

Degner, J., Wentura, D., Gniewosz, B., & Noack, P. (2007). Hostility-related prejudice against Turks in adolescents: Masked affective priming allows for a differentiation of automatic prejudice. *Basic and Applied Social Psychology, 29*, 245–256.

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127*, 286–303.

Epstude, K., & Mussweiler, T. (2009). What you feel is how you compare: How comparisons influence the social induction of affect. *Emotion, 9*, 1–14. http://dx.doi.org/10.1037/a0014148.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*, 247–279. http://dx.doi.org/10.1037/a0022327.

Frings, C., & Wentura, D. (2003). Who is watching "Big Brother"? TV consumption predicted by masked affective priming. *European Journal of Social Psychology, 33*, 779–791.

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorisation influence spontaneous evaluations of multiply categorisable objects? *Cognition & Emotion, 24*, 1008–1025. http://dx.doi.org/10.1080/02699930903112712.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science, 273*, 1699–1702.

Hager, W., Mecklenbräuker, S., Möller, H., & Westermann, R. (1985). Emotionsgehalt, Bildhaftigkeit, Konkretheit und Bedeutungshaltigkeit von 580 Adjektiven: Ein Beitrag zur Normierung und zur Prüfung einiger Zusammenhangshypothesen [Emotionality, imageability, concreteness, and meaningfulness of 580 adjectives: A contribution to norming and test of correlations]. *Archiv für Psychologie, 137*, 75–97.

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., ... L, S. (2013). On the automatic activation of attitudes: A quarter century of evaluative priming research. *Psychological Bulletin, 139*, 1062–1089. http://dx.doi.org/10.1037/a0031309.

Hugenberg, K. (2005). Social categorization ant he perception of facial affect: Target race moderates the response latency advantage for happy faces. *Emotion, 5*, 267–276.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640–643. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1478.x.

Imhoff, R., Schmidt, A. F., Bernhardt, J., Dierksmeier, A., & Banse, R. (2011). An inkblot for sexual preference: A semantic variant of the Affect Misattribution Procedure. *Cognition & Emotion, 25*, 676–690. http://dx.doi.org/10.1080/02699931.2010.508260.

Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.

Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods, 45*, 905–927. http://dx.doi.org/10.3758/s13428-013-0330-5.

Klauer, K. C., Mierke, J., & Musch, J. (2003). The positivity proportion effect: A list context effect in masked affective priming. *Memory & Cognition, 31*, 953–967.

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch, & K. C. Klauer (Eds.). *Psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ: Erlbaum.

Klauer, K. C., Musch, J., & Eder, A. B. (2005). Priming of semantic classifications: Late and response related, or earlier and more central? *Psychonomic Bulletin & Review, 12*, 897–903.

Kubota, J. T., & Ito, T. A. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental Social Psychology, 43*, 738–752. http://dx.doi.org/10.1016/j.jesp.2006.10.023.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Retrieved from.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24*, 1377–1388. http://dx.doi.org/10.1080/02699930903485076.

Möller, H., & Hager, W. (1991). Angenehmheit, Bedeutungshaltigkeit, Bildhaftigkeit und Konkretheit-Abstraktheit von 452 Adjektiven [Pleasantness, meaningfulness, imageability and concreteness of 452 adjectives]. *Sprache und Kognition, 10*, 39–51.

Neumann, R., & Seibt, B. (2001). The structure of prejudice: Associative strength as a determinant of stereotype endorsement. *European Journal of Social Psychology, 31*, 609–620.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*, 101–115. http://dx.doi.org/10.1037/1089-2699.6.1.101.

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology, 29*, 1049–1071.

de Paula Couto, M. C. P., & Wentura, D. (2017). The go/no-go priming task: Automatic evaluation and categorization beyond response interference. *Cognition & Emotion, 31*, 892–911.

Paulus, A., Musial, E., & Renn, K. (2014). Gender of the expresser moderates the effect of emotional faces on the startle reflex. *Cognition & Emotion, 28*, 1493–1501. http://dx.doi.org/10.1080/02699931.2014.886557.

Paulus, A., Renn, K., & Wentura, D. (2018). *Ingoup/outgoup status the expresser moderates the effect of emotional faces on the startle reflex.* Manuscript in preparation.

Paulus, A., Rohr, M., Neuschwander, M., Seewald, M., & Wentura, D. (2012). *Validation data of the Saarbrücken emotional faces database.* Unpublished data Saarbrücken, Germany: Saarland University.

Paulus, A., & Wentura, D. (2014). Threatening joy: Approach and avoidance reactions to emotions are influenced by the group membership of the expresser. *Cognition & Emotion, 28*, 656–677. http://dx.doi.org/10.1080/02699931.2013.849659.

Peeters, G. (1983). Relational and informational patterns in social cognition. In W. Doise, & S. Moscovici (Vol. Eds.), *Current issues in European social psychology: 1*, (pp. 201–237). Cambridge: Cambridge University Press.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology, 1*, 33–60.

Rohr, M., Degner, J., & Wentura, D. (2012). Masked emotional priming beyond global valence activations. *Cognition & Emotion, 26*, 224–244.

Sachisthal, M. S. M., Sauter, D. A., & Fischer, A. H. (2016). Mimicry of ingroup and outgroup emotional expressions. *Comprehensive Results in Social Psychology, 1*, 86–105. http://dx.doi.org/10.1080/23743603.2017.1298355.

van der Schalk, J., Fischer, A., Doosje, B., Wigboldus, D., Hawk, S., Rotteveel, M., & Hess, U. (2011). Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion, 11*, 286–298.

Teige-Mocigemba, S., & Klauer, K. C. (2013). On the controllability of evaluative-priming effects: Some limits that are none. *Cognition & Emotion, 27*, 632–657.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*, 426–432.

Wagner, U., van Dick, R., Pettigrew, T. F., & Christ, O. (2003). Ethnic prejudice in East and West Germany: The explanatory power of intergroup contact. *Group Processes & Intergroup Relations, 6*, 22–36. http://dx.doi.org/10.1177/1368430203006001010.

Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology, 95*, 1063–1079.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & B. K. Payne (Eds.). *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York: Guilford.

Wentura, D., Kulfanek, M., & Greve, W. (2005). Masked affective priming by name letters: Evidence for a correspondence of explicit and implicit self-esteem. *Journal of Experimental Social Psychology, 41*, 654–663.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology, 78*, 1024–1037.