Running head: EMOTIONAL EXPRESSION AND GROUP MEMBERSHIP

Emotional face expressions and group membership:

Does affective mismatch induce conflict?

Dirk Wentura, Emre Gurbuz, Andrea Paulus, & Michaela Rohr,

Saarland University

----- Journal of Experimental Psychology: Human Perception and Performance, in press ------

Author Note

Dirk Wentura, wentura@mx.uni-saarland.de

Emre Gurbuz, emre.gurbuz@uni-saarland.de

Andrea Paulus, andrea.paulus@gmail.com

Michaela Rohr, m.rohr@mx.uni-saarland.de

This research was supported by a grant from the German Research Foundation to Dirk Wentura and Andrea Paulus (WE 2284/14-2). We would like to thank Ullrich Ecker for his helpful comments on an earlier version of this article.
Correspondence concerning this article should be addressed to Dirk Wentura, Saarland University, Department of Psychology, Campus A2.4, D-66123 Saarbruecken. E-mail: wentura@mx.uni-saarland.de

Abstract

When asked to judge or react to a facial emotional display of a person, people do not only take the emotion into account, but also other socially important features of the face, such as, for example, ethnicity (Kozlik & Fischer, 2020; Paulus & Wentura, 2014). Importantly, the emotion-related and non-emotion-related features are seemingly not (or not always) processed in a simple, additive manner, but are—in a more functional manner—integrated to provide an "amalgamated signal" on which individuals base their judgment and responses. Whereas Paulus and Wentura (2014) put forward a social-message account of this amalgamated signal, Kozlik and Fischer (2020) recently proposed a processing-conflict explanation. The empirical evidence regarding this issue is, however, mixed. In three experiments, we aimed at replicating and extending Kozlik and Fischer's central experiment in order to gain further insight into the validity of the social-message versus the processing-conflict account. However, we failed to replicate their findings. The implications of the new evidence for the two accounts are discussed.

Keywords: face processing, emotional expression, processing conflict, social message, prejudice

Public significance Statement

The study contributes to a debate about how people react to facial expressions based on both emotion and (prejudice-related) ethnicity. One account assumes that both features are processed independently (thereby potentially producing cognitive conflict); the other account assume that they are integrated (i.e., the ethnicity feature modifies the evaluation of the expression). The study attempted to replicate a previous study that was (at first sight) in favor of the first account but failed to do so.

Emotional face expressions and group membership:

Does affective mismatch induce conflict?

In recent studies, an interactive influence of emotional expression and group membership on implicit evaluation of emotional faces has been observed (Kozlik & Fischer, 2020; Paulus & Wentura, 2014; Weisbuch & Ambady, 2008). Specifically, positive reactions are facilitated by positive facial expressions of in-group members (e.g., faces of White people for White participants) and negative facial expressions of out-group members (e.g., faces of Black persons in the US, see Weisbuch & Ambady, 2008, or faces of Turkish/Middle-Eastern ethnicity in Germany, see Kozlik & Fischer, 2020, Paulus & Wentura, 2014);[1] correspondingly, negative facial expressions of in-group and positive facial expressions of out-group members facilitate negative reactions. Initially, this interactive influence of emotional expression and group membership was explained by the social-message account (SMA; Weisbuch & Ambady, 2008; see also Paulus & Wentura, 2014). The SMA suggests that a specific facial expression is interpreted differently depending on whether a person is perceived as a member of a benevolent ("friend") or malevolent ("enemy") group. Therefore, the same expression can be seen as conveying a different social message. Specifically, a positive expression (e.g., happiness) of an in-group member is interpreted as an affiliation intention, whereas the positive expression of an out-group member is seen as signaling a negative intention (e.g., being mischievous or experiencing *Schadenfreude*). In the same vein, a negative expression (e.g., fear) of an in-group member is expected to convey a negative signal (e.g., in the case of fear, a warning of danger); a negative facial expression of an out-group member is expected to convey a positive signal (e.g., submission). An important

[1] In both cases (i.e., Black persons in the US; Turkish/Middle-Eastern persons in Germany) the existence of racial prejudices can be taken for granted (see, e.g., Degner et al., 2007; Degner & Wentura, 2011; Fazio et al., 1995; Meissner & Rothermund, 2013; Neumann & Seibt, 2001). In the remainder of the paper, where we use the terms in-group and out-group when discussing our own experiments or those of Kozlik & Fischer, 2020, "in-group" refers to White people and out-group to Turkish/Middle-Eastern people (samples were recruited accordingly).

assumption here is that the two features – emotional expression and ethnicity – are immediately integrated.

However, an alternative explanation for the interactive influence of emotional expression and group membership was recently proposed: The *processing-conflict account* (PCA; Kozlik & Fischer, 2020). The PCA suggests that there are two independent features of the face (i.e., emotional expression and ethnicity) that can be positive or negative; the emotional expression is evaluated in line with the valence displayed (i.e., happy is positive, fear is negative), and ethnicity is evaluated in line with the prevailing prejudice or group-membership, respectively. If one feature is positive and the other one is negative (e.g., White/fearful or Middle-Eastern/happy, for a prejudiced White observer), the account postulates a processing conflict that – depending on task-context – promotes avoidance behavior or distraction from ongoing goal-oriented behavior. Such evaluative conflicts are well-known sources of effects in experimental paradigms used to measure (non-intentional) evaluations. If the PCA is correct, the observed effects would not be especially social in nature. In the spirit of Occam's razor (i.e., if two theories are equally powerful, choose the simpler one), PCA seems to have an advantage. Therefore, we deemed it important to compare and test the two accounts' predictions.

This is not an easy task, as SMA and PCA make similar predictions for some tasks and paradigms. For instance, the interaction of emotional expression and ethnicity found in Paulus and Wentura (2014) with an approach/avoidance task can be explained by both accounts (but see Wentura & Paulus, 2022). The SMA would predict that in-group happiness and out-group fear convey positive social signals (see above), thus promoting more approach behavior (and less avoidance behavior) compared to in-group fear and out-group happiness, respectively. Similarly, the PCA would suggest that in-group happiness and out-group fear are congruent in valence, and this congruency creates positive affect that promotes an

approach reaction. On the other hand, in-group fear and out-group happiness result in valence incongruency, which creates negative affect that promotes avoidance responses.

Indeed, Kozlik and Fischer (2020) tested their account using positive and negative emotional expressions and ethnicities in different tasks. They presented happy and angry or fearful (Exp. 1) in-group and out-group faces to participants, but instead of requiring an approach/avoidance response, participants had to either categorize the displayed emotion (Exp. 1, 3, 4_{Block2}) or decide which side of the presented face was blurred (Exp. 2, 4_{Block1}). The authors predict for both versions slower responses in case of conflict (i.e., happy/out-group and angry/in-group) and argue that these effects are not easily explainable by SMA.

While the two tasks might seem quite similar at first glance, there is an important methodological difference between them: in the first case, one of the evaluative stimulus features is task relevant (i.e., emotional expression); in the second case, the response-relevant feature is orthogonal to the two evaluative features of interest. In our view, this difference completely changes the hypothesized mechanisms, as we will elaborate below. Thus, methodologically the empirical part of Kozlik and Fischer (2020) has to be separated into two parts, as we will outline in the following.

The Stroop-like paradigm: A different type of processing conflict

In three of the five Kozlik and Fischer (2020) experiments (Exp. 1, 3, 4_{B2}), emotional expression was the task-relevant feature (i.e., “Is the facial expression positive or negative?”) whereas group status (in-group vs. out-group) was a task-irrelevant feature; this created a Stroop-like paradigm (Stroop, 1935). In an abstract sense, stimuli in Stroop-like paradigms have two features: one task-relevant and one that is task-irrelevant but maps onto the response categories as well (in the classical task, the ink color of the word – i.e., the task-relevant feature – has to be named whereas the meaning of the color word – i.e., the task-irrelevant feature – has to be ignored). Even though participants are instructed to ignore the

task-irrelevant feature, it is nevertheless processed and either facilitates (in case of congruency) or hinders (in case of incongruency) the correct response, resulting in congruency effects on reaction times and/or error rates. Kozlik and Fischer found evidence for such a Stroop-like effect: responses were relatively slower if the task-irrelevant ethnicity valence did not match the response category. That is, positive responses to happy out-group faces as well as negative responses to angry or fearful in-group faces were relatively slower than their counterparts. Moreover, Kozlik and Fischer found evidence for other signature effects of Stroop-like paradigms, that is, a congruency sequence effect (CSE; i.e., a smaller congruency effect following incongruent trials; Exp. 1 and Exp. 4_{B2}; Gratton et al., 1992; for a review see Duthoo et al., 2014) and a proportion effect (i.e., a smaller congruency effect if the congruency proportion is low; Exp. 3; Bugg & Crump, 2012).

These experiments do not show, however, that the two features (i.e., emotional expression and ethnicity) *per se* are in conflict. We argue that a conflict might only be triggered by the task demands and is then localized at the *response* level: Participants prepare a response according to the categorization of the emotional expression; if the evaluation of the group is congruent with this response, responses are facilitated; if the group evaluation matches the alternative response, responses are slowed down because the response conflict needs to be resolved. In our view, results of Kozlik and Fischer (2020) do not show that the two features are always in conflict (i.e., irrespective of task), as the authors argue. Thus, even though we agree that the results are indicative of conflict, we believe that this conflict results from task mechanisms.

Another structural sibling of the paradigm used by Kozlik and Fischer (2020) should be mentioned here: the evaluative priming paradigm (Fazio et al., 1986; Herring et al., 2013). In this paradigm, participants have to quickly categorize target stimuli as positive or negative (see Wentura & Degner, 2010, for a discussion of different variations). Targets are preceded

by briefly presented prime stimuli, which are positive or negative as well. The evaluative priming effect is reflected in faster responses in valence-congruent trials compared to incongruent trials. The typical interpretation is that the prime is involuntarily evaluated and therefore facilitates or hinders responses depending on whether it matches or mismatches the required target response. The results can thus be explained without the need to assume that prime and target stimuli are *per se* in conflict. It is sufficient to assume that the task and the evoked response tendencies matter. Thus, the Stroop-like paradigm used by Kozlik and Fischer can be thought of as an ingenious variant of the evaluative priming paradigm for assessing involuntarily evoked prejudices, in line with the seminal study by Fazio et al. (1995). Whereas Fazio and colleagues found that voluntarily generated positive and negative responses to target words were influenced by the involuntarily evoked prejudices towards black (compared to white) faces that preceded the targets, Kozlik and Fischer integrated prime information (i.e. the group feature) and target information (i.e., emotional expression) within a single stimulus. Thus, a congruency effect as found by Kozlik and Fischer reveals prejudice but is no convincing evidence for PCA, which states that group status and emotional expression are in conflict.

In conclusion, we believe that response conflict plays a role in the Stroop-like experiments of Kozlik and Fischer (2020), but we do not follow the authors' argument that conflict between the two features – as postulated by the PCA – is responsible for their results.

The "unrelated task" paradigm: Inconclusive evidence

In contrast to the experiments discussed above, in the remaining two experiments by Kozlik and Fischer (2020), both emotion expression (joy vs. anger) and group status (in-group vs. out-group) were task-irrelevant characteristics. Given our arguments above, one might say that these experiments aim at testing the central hypothesis of the PCA that the two features trigger positive affect in case of congruency (i.e., happy in-group, fearful/angry out-

group) and negative affect in case of incongruency (i.e., happy out-group, fearful/angry in-group). In Experiments 2 and 4_{B1}, again a 2 (emotion: happy vs. angry) × 2 (ethnicity: White vs. Turkish/Middle-Eastern) variation of stimulus materials was used. However, the participants' task was now completely detached from the emotional expression and the group feature. In Experiment 2, front-view faces were slightly blurred on the left or right side (see also Paulus & Wentura, 2014) and participants had to categorize items based on the side of blurring. In Experiment 4_{B1}, half-profile view faces were used; participants had to categorize items by the gaze direction (i.e., left vs. right). Thus, any presumed processing conflict can no longer be explained as task-induced but must lie in the evaluative incongruency of the two features. And indeed, the authors found evidence for such a congruency effect (i.e., slower [quicker] responses for happy [angry] Middle-Eastern and angry [happy] White faces) in Experiment 2.[2]

This is a remarkable result. To put it into context, other than the Stroop-like experiments discussed earlier, to our knowledge only two studies proceeded from a similar vantage point (Gast et al., 2014; Hermans et al., 1998). Both are from the "branch" of evaluative priming research that investigates whether priming effects can be found even when the target valence is not task relevant (for a review, see Rohr & Wentura, 2022). Within this branch, Hermans et al. (1998) proposed the *affective-motivational account* of evaluative priming. They argued that automatic evaluation triggers action tendencies of – roughly speaking – approach and avoidance. If prime and target trigger diverging action tendencies, a conflict has to be resolved, which slows down any concurrent task.

[2] Kozlik and Fischer (2020) did not expect and did not find a congruency sequence effect (CSE), a result that is emphasized by the authors. The prediction is uncontroversial: the CSE is a marker of response interference tasks (like Stroop); since Experiment 2 is not a response interference experiment, it is plausible to not expect a CSE.

Hermans et al. (1998) tested this account using a color-naming task in the evaluative priming paradigm: The color of the (positive or negative) target had to be named; a (positive or negative) prime preceded the target. The similarity to the paradigm used in Kozlik and Fischer (2020) Experiments 2 and 4_{B1} is striking; the only difference is that in Hermans et al., valence congruency or incongruency was a feature of a stimulus *pair* (and not of a single stimulus as in Kozlik and Fischer. However, the four experiments conducted by Hermans et al. yielded no support for the hypothesis of a congruency effect in color-naming latencies (see also Rothermund & Wentura, 1998).

In another attempt to test the *affective-motivational account*, Gast et al. (2014) presented two valent pictures (in fast succession), which were either affectively congruent or incongruent; participants had to categorize a letter (X or Y) that was superimposed on the second picture. The authors indeed found an affective congruency effect (letter categorization was faster in case of picture congruency). However, they found this effect only when an evaluative context was given (i.e., when trials with valence categorization of primes were interleaved in the trial sequence). Thus, given the results of Hermans et al. (1998) and Gast et al. (2014) it is astonishing that Kozlik and Fischer (2020) found an effect.

However, if we take the emotion × ethnicity interaction effect in Kozlik and Fischer (2020) Experiment 2 at face value, can the SMA account for this finding as well? We believe it can. Our reasoning is based on the results from a different paradigm, namely the emotional Stroop task (e.g., Frings et al., 2010; McKenna & Sharma, 1995; Pratto & John, 1991; Wentura et al., 2000). In this task, participants have to name the color of positive and negative words (thus, structurally the paradigm is different from the original Stroop task; Algom et al., 2004). Negative stimuli tend to slow down color-naming responses (Frings et al., 2010; McKenna & Sharma, 1995; Pratto & John, 1991; Wentura et al., 2000). Since, according to the SMA, happy out-group faces as well as fearful in-group faces are more

negative than their respective counterparts (i.e., happy in-group faces and fearful out-group faces, respectively), they might tend to tend to slow down "blurredness" responses.

Of course, we need to keep in mind here that Kozlik and Fischer (2020) used anger instead of fear stimuli in Experiment 2 and that the SMA does not predict that responses to anger stimuli are moderated by group membership. However, the congruency effect (i.e., the emotion × ethnicity interaction) found in Experiment 2 of Kozlik and Fischer is nevertheless compatible with the SMA in principle, because the SMA still predicts an emotion × ethnicity interaction (i.e., moderation of responses to happy faces by ethnicity, no moderation in case of angry faces).[3]

Thus, the congruency effect found in Experiment 2 of Kozlik and Fischer (2020) is in principle compatible with the SMA. Note, however, that given some intricacies of the emotional Stroop paradigm (Frings et al., 2010), we would not have predicted this effect with confidence. Thus, while the effect is an important pillar of defense of the PCA (Kozlik & Fischer, 2020), the SMA is neutral with regard to the results obtained with this paradigm.

Experiment 4_{B1} of Kozlik and Fischer (2020) was a conceptual replication of Experiment 2; the only substantive difference between experiments was that the task in Experiment 4_{B1} focused on the gaze direction of half-profile faces. This experiment is especially interesting, as PCA and SMA make different predictions. Specifically, it can be plausibly argued that the change to face orientation makes a difference for the SMA but not the PCA: the social messages assumed by the SMA apply if a face is directed at the participant, but not if the face looks away from the participant (i.e., an averted happy face signals a message for someone else, not for the observer). However, from a PCA perspective,

[3] Admittedly, Kozlik and Fischer (2020) found a *disordinal* interaction pattern in their Experiment 2, that is, significant congruency effects for happy *and* for angry faces. However, to decompose a 2 × 2 within-participants interaction pattern is always presuppositional (see, e.g., Wentura & Degner, 2010). We will not elaborate on this issue here because the discussion is moot in light of the to-be-presented results.

face orientation should make no difference as long as the two independent features (i.e., emotional expression and group) can be easily extracted. Thus, the PCA would predict a congruency effect in Experiment 4_{B1} similar to the one found in Experiment 2. However, Kozlik and Fischer did not report whether or not there was a congruency effect in Experiment 4_{B1}.[4] Therefore we reanalyzed the data of Experiment 4_{B1}.[5]

Reanalysis of Experiments 2 and 4_{B1} of Kozlik and Fischer (2020)

Table 1 shows the mean response times (RT) and error rates for the conditions of interest in Experiments 2 and 4_{B1} of Kozlik and Fischer (2020). Following Kozlik and Fischer, we report means for "congruent trials" that comprise happy/White and angry/Middle-Eastern stimulus configurations and means for "incongruent trials" that comprise happy/Middle-Eastern and angry/White stimulus configurations, because the data (of Experiment 4_{B1}) available to us were collapsed over emotion categories. Note, the emotion × ethnicity interaction in a 2 (emotion: happy vs. angry) × 2 (ethnicity: White vs. Middle-Eastern) design is of course equivalent to a main effect of the congruency factor.

For both experiments, we conducted a one-factorial congruent vs. incongruent repeated measures ANOVA for RTs and errors as dependent variables, respectively. For Experiment 2, our reanalysis confirmed the significant congruency effect reported by Kozlik and Fischer

[4] Kozlik and Fischer (2020) only analyzed the full Experiment 4, which included a second block (that always came last) in which emotional expression was the task-relevant feature of the averted faces. Thus, this block belongs to the Stroop-like paradigm, as Experiments 1 and 3 of Kozlik and Fischer. They conducted a congruency (congruent vs. incongruent) × $congruency_{n-1}$ (i.e., congruency of the preceding trial) × gaze direction (left vs. right) × block (valence-irrelevant vs. valence-relevant task) analysis. They found an unexpected four-way interaction that was based on a three-way interaction for right-looking faces that was not significant for left-looking faces. The three-way interaction (for right-looking faces) was based on a significant congruency sequence effect (i.e., a congruency × $congruency_{n-1}$ interaction) for the valence-relevant block, which was not significant for the valence-irrelevant block (i.e., Experiment 4_{B1}). Thus, Kozlik and Fischer reported that the valence-relevant block (i.e., Experiment 4_{B2}) produced the signature effect of a response interference paradigm (see Section "The Stroop-like paradigm") and that this signature effect was absent in the valence-irrelevant block (i.e., Experiment 4_{B1}).

[5] Thanks to Julia Kozlik for providing us with the data.

(2020; see p. 558), $F(1,34) = 12.89$, $p < .001$, $\eta_p^2 = .275$. (For errors, $F(1,34) = 1.73$, $p = .197$, $\eta_p^2 = .049$.) For Experiment $4_{B1,}$ however, there was no congruency effect, $F < 1$. Constraining the analysis to left-looking and right-looking faces, respectively, yielded null results as well: $M = 5.5$ ms ($SE = 4.2$ ms), $F(1,26) = 1.69$, $p = .205$, $\eta_p^2 = .061$, for left-looking faces; $M = -5.8$ ms ($SE = 4.1$ ms), $F(1,26) = 2.01$, $p = .168$, $\eta_p^2 = .072$, for right-looking faces. (For errors, $F < 1$.)

A 2 (congruent vs. incongruent) × 2 (Experiment 2 vs. Experiment 4_{B1}) ANOVA yielded, on the one hand, no significant interaction, $F(1,60) = 2.42$, $p = .125$, $\eta_p^2 = .039$, presumably because the congruency difference variable was quite noisy in Experiment 4_{B1} compared to Experiment 2 (see the *SE*s in Table 1). On the other hand, the main effect of congruency failed to reach significance as well, $F(1,60) = 2.16$, $p = .147$, $\eta_p^2 = .035$. Thus, we are faced with the following unsatisfactory situation:

(a) If the congruency effect (i.e., the emotion × ethnicity interaction) in Experiment 2 and the lack of such an effect in Experiment 4_{B1} are taken at face value, the balance of evidence would be more in line with the SMA than the PCA, because the SMA is compatible with finding the interaction with frontal-view faces but not with averted faces, whereas the PCA – according to Kozlik and Fischer (2020) – predicts the interaction in both conditions. However, in order to corroborate this conclusion, it is necessary to show a higher-order interaction in a replication study that has frontal versus profile view as a factor.

(b) If the congruency effect (i.e., the emotion × ethnicity interaction) in Experiment 2 is taken at face value and the lack of an effect in Experiment 4_{B1} is taken as a Type II error, the balance of evidence would favor the PCA over the SMA. Then, a replication of Experiment 4_{B1} should yield the expected congruency effect (i.e., the emotion × ethnicity interaction).

(c) If the absence of a congruency effect (i.e., the emotion × ethnicity interaction) in Experiment 4_{B1} is taken at face value and the effect in Experiment 2 is taken as a Type I

error, there is no longer any evidence for the PCA (i.e., an affective mismatch of two valent stimulus features producing a processing conflict). That is not to say that the PCA is incorrect, but positive evidence would be lacking.

Overview of Studies

We initially proceeded from the assumption that both the congruency effect (i.e., the emotion × ethnicity interaction) in Experiment 2 of Kozlik and Fischer (2020) and the failure to find this effect in Experiment 4_{B1} can be taken at face value; thus, we aimed to conduct an experiment that conceptually replicates Experiments 2 and 4_{B1} of Kozlik and Fischer in a single experiment with the factor frontal versus profile-view face. Note, however, that experiments of Kozlik and Fischer used two different tasks: participants categorized faces based on either which side was blurred (Experiment 2) or which side was averted (i.e., gaze direction; Experiment 4_{B1}). In a single experiment, it would be desirable to have a common task. Therefore, we invented the "mole task" – a task where participants decide whether there are two or three moles on a face – because this task can be applied to both frontal-view and averted faces. In a first step, we used only frontal-view faces to test whether we could conceptually replicate the congruency effect (i.e., the emotion × ethnicity interaction) found in Experiment 2 of Kozlik and Fischer with this task. To anticipate, we did not find it. Therefore, we then tried to establish whether the result of Experiment 2 of Kozlik and Fischer was replicable at all. In two experiments, we found it was not. In light of these results, it seemed more natural to deviate from the chronological order and to first report the closer replication attempts as Experiments 1 and 2, followed by the "mole task" study as Experiment 3. Finally, Experiment 4 was added to scrutinize a main effect of ethnicity that we observed in the first three experiments; Experiment 4 used inverted faces as stimuli.

As discussed earlier, Kozlik and Fischer (2020; as well as Paulus & Wentura, 2014) examined the emotion × group membership interaction in a German sample. Therefore, the

face stimuli used in their experiments were White/Caucasian for in-groups and Turkish/Middle Eastern for out-groups since a prejudice against Turkish/Middle Eastern men in Germany can be taken for granted (Degner et al., 2007; Degner & Wentura, 2011; Neumann & Seibt, 2001; Wagner et al., 2003). Here, we used the same stimuli.

The experiments were conducted online with recruitment via Prolific (www.prolific.co) in contrast to the lab experiments by Kozlik and Fischer (2020). They recruited students from their (German) university being "native Whites". To most closely match the population, the following filter criteria were used to recruit participants via Prolific: Participants (a) aged between 18 and 35, (b) who are White/Caucasians (c) who have German nationality, (c) whose first language is German (d) who are fluent in their first language (d) who were born in Germany, and (e) who are living in the country that they were born (i.e., Germany).[6]

The experiments were pre-registered (see links below) and the materials are available online https://osf.io/c7hnj/?view_only=44ba65859de1420e9477b0c9170a2b28. Our studies did receive approval of the research ethics committee of the faculty of human and business sciences of Saarland university, Germany.

Experiment 1

Experiment 1 aimed to replicate Experiment 2 by Kozlik and Fischer (2020). As in Kozlik and Fischer, participants' task was to indicate the blurred side of the presented frontal-view faces. Neither emotion nor ethnicity were task-relevant. The pre-registration can be found at https://aspredicted.org/HRM_431.

Method

[6] Criteria (d) and (e) were not yet available in Prolific for Experiment 3 which was – as already noted – chronologically first.

Participants. The effective sample was $N = 58$ participants (26 females, 32 males; 52 right-handed, 5 left-handed, 1 ambidextrous; age *Md* = 27 years, range: 18-35). To achieve this sample size, we recruited a total of 60 participants. One participant reported a migration background and their data were excluded from analysis (see *Pre-registration* for exclusion criteria). Another participant's data were not recorded due to technical problems. Participants were compensated with £3.

The sample size was determined based on the following rationale: Experiment 2 of Kozlik and Fischer (2020) reported an ethnicity × emotion interaction effect of $d_Z = .61$.[7] We lowered the expected value to $d_Z = .50$ since the experiment was conducted online (with potentially more noise). To detect an effect of $d_Z = .50$ with a power of $1\text{-}\beta = 0.95$ ($\alpha = 0.05$), a sample size of 54 was needed (G*Power; Faul et al., 2007). To account for potential outliers, we aimed to recruit 60 participants (see *Pre-registration*).

Design. We used a 2 (ethnicity: Middle-Eastern vs. White) × 2 (emotion: happiness vs. anger) within-subjects design.

Materials. We used the faces of the same people used in Experiment 4 of Kozlik and Fischer (2020),[8] that is, eight White and eight Middle-Eastern male faces with happiness and anger expressions from the Radboud Faces Database (Langner et al., 2010). The same faces with neutral expressions were selected for the practice block. All images had straight head orientation and gaze directed at the viewer. Each face stimulus was slightly blurred on the left or right side as in Paulus and Wentura (2014, Exp. 1) and Kozlik and Fischer (Exp. 2).

---

[7] Experiment 2 of Kozlik and Fischer (2020) reported $F(1, 34) = 12.89$, $p = .001$, $\eta_p^2 = .28$ for the emotion by ethnicity interaction effect; $d_Z = \sqrt{F/N} = \sqrt{12.89/35}$ (see, e.g., Lakens, 2013).

[8] As already noted above, we started this series of experiments with the ultimate goal of replicating the full pattern found by Kozlik and Fischer (2020) in their Experiments 2 and 4_{B1}. Therefore, a set of stimuli was needed with both frontal and profile-view faces, which is provided by. the Radboud Faces Database (Langner et al., 2010). Their outgroup stimuli depict in fact Moroccan males. In line with Kozlik and Fischer and our former research, we kept with the term "Middle-Eastern" to denote a phenotype.

Procedure. Participants were instructed to turn off all software that could deliver notifications and to turn off or mute mobile phones to create a silent and non-distracting environment during the experimental session. To adjust presentation parameters to the actual screen size, participants were asked to resize a credit-card image (presented on the screen) to the size of a real credit card or equivalent (e.g., an ID card or driver's license) by using left/right aKozliknd up/down arrow buttons on their keyboard before the start of the experiment.

Participants were informed that faces of young men with slight blur on the left or right side of the image would be presented on the screen in random order and that their task was to decide as quickly as possible whether the blur was on the left or the right side. Thus, critical variables (i.e., emotional expression and ethnicity) were task-irrelevant. The responses "left" and "right" were assigned to the "x" and "m" keys, respectively.

The trial sequence began with a centrally displayed fixation cross that remained on the screen for 500 ms. The fixation cross was replaced by a target face image, which remained on the screen for 1500 ms. Participants were instructed to respond as fast and accurately as possible. If the response took longer than 2000 ms (from target onset), a warning message (i.e., "Too slow! Please respond faster!") was displayed. A new trial started after an inter-trial interval of 200 ms.

The main part of the experiment consisted of 512 trials, separated into 8 blocks of 64 trials. In each block, eight White faces and eight Middle-Eastern faces (a total of 16 unique individuals) were presented with both happy and angry expressions; each image was presented once with blur on the left and once with blur on the right.

Before the experimental trials, the same faces were presented with a neutral expression, again once with blur on the left and once with blur on the right, resulting in a total of 32 trials. Half of these trials served as a practice block, in which participants received

response feedback (i.e., “correct” or “false”, presented for 700 ms). The other half served as warm-up trials at the beginning of the first experimental block, with no response feedback.

Results

Trials with incorrect responses (2.71%) were excluded from analysis, as were trials with RTs shorter than 150 ms or longer than 1.5 interquartile ranges above the third quantile with respect to the individual distribution (5.05% of the remaining trials; Tukey, 1977).[9] Table 1 shows mean RTs and error rates for the conditions of interest.

A 2 (emotional expression: happiness, anger) × 2 (ethnicity: White, Middle-Eastern) repeated measures ANOVA with mean RTs as the dependent variable yielded a main effect of ethnicity, $F(1, 57) = 35.39$, $p < .001$, $\eta_p^2 = .383$ ($BF_{10} = 82144.87$; i.e., “decisive evidence” following Jeffreys, 1961), indicating that responses to White faces were $M = 9$ ms ($SD = 12$ ms) faster than responses to Middle-Eastern faces. The main effect of emotion reached significance as well, $F(1, 57) = 4.05$, $p = .049$, $\eta_p^2 = .066$, suggesting that responses to happy faces were $M = 2$ ms ($SD = 9$ ms) faster than responses to angry faces. However, the Bayes factor of $BF_{10} = 0.94$ indicated “no evidence” in support of the hypothesis.

The emotion × ethnicity interaction, that is, the effect of most interest here, did not reach the criterion of significance, $F < 1$, with a Bayes factor of $BF_{0+} = 6.03$ providing “substantial evidence” in favor of the directed null hypothesis that the mean difference of RT(incongruent) minus RT(congruent) is not greater than zero.

An analyses of error rates yielded no significant effects. The emotion main effect, $F(1, 57) = 2.03$, $p = .160$, $\eta_p^2 = .034$, with Bayes factor $BF_{01} = 2.68$ indicating “anecdotal evidence” in favor of the null hypothesis. The main effect for ethnicity, as well as the

[9] Kozlik and Fischer (2020) excluded not only erroneous trials but also trials following an error. Their RT outlier criterion was 2.5 *SD*s from individual mean RTs per condition. Analyzing our data using these criteria do not essentially change the results. This holds for all experiments reported in this text.

interaction between emotion and ethnicity showed evidence in favor of the null hypothesis (both *Fs* < 1; BF_{01} = 6.93 for the emotion main effect and BF_{0+} = 2.93 for the interaction effect).

Discussion

The emotional expression × ethnicity interaction that would have corroborated the findings of Kozlik and Fischer (2020) was not observed. We only found an ethnicity effect and weak evidence for an emotion effect. This suggests that the emotion × ethnicity interaction predicted and found by Kozlik and Fischer is not easily replicable. Given the conflicting results of Kozlik and Fischer and our Experiment 1, we wanted to put the hypothesis of an emotional expression × ethnicity interaction effect to a further test. Additional methodological reasons to replicate Experiment 1 were: (a) The inter-trial interval was shorter (i.e., 200 ms) than that used in Kozlik and Fischer's Experiment 2 (1000 ms); (b) there were 512 experimental trials rather than 200 as in Kozlik and Fischer's Experiment 2;[10] (c) practice and experimental trials used the faces of the same people, rather than using different individuals (as in Kozlik & Fischer, 2020; Experiment 2). Finally, (sequential) Bayesian analyses were not specified a priori in Experiment 1; these were deemed more appropriate given the possible validity of null hypotheses (Schönbrodt et al., 2017).

Experiment 2

Method

The pre-registration of Experiment 2 can be found at https://aspredicted.org/2TS_ZKH.

Participants. We aimed for a Bayes factor of either BF_{0+} > 6 in favor of the null hypothesis for the interaction (i.e., mean RT[incongruent] minus mean RT[congruent] is not

[10] Note, however, that the null result in Experiment 1 with regard to the crucial interaction also held with only 192 trials (3 blocks; the trial number closest to 200 that was still fully balanced).

greater than zero), or a Bayes factor of $BF_{+0} > 6$ in favor of the alternative hypothesis (following Schönbrodt et al., 2017). The data-collection strategy was as follows: (a) Collect an initial sample of 40 participants and test if the Bayes factor is in favor of the null or the alternative hypothesis (according to the criteria described above); (b) if necessary, test additional participants in batches of 20 and reanalyze the data after each batch until one of the two critical BF thresholds is reached; (c) terminate data collection when a sample size of 100 is reached regardless of achieved BF (i.e., a stopping rule). As a critical BF threshold was reached with the initial sample of 40 participants, data collection ceased.

The effective sample was $N = 39$ participants (19 females, 20 males; 33 right-handed, 6 left-handed; age $Md = 26$ years, range: 18-34). One participant reported a migration background and their data were therefore excluded (see *Pre-registration* for exclusion criteria). Participants were compensated with £1.50.

Materials and Procedure

Experiment 2 was a replication of Experiment 1 with the following differences: (a) The inter-trial interval was extended to 1000 ms; (b) the number of trials was reduced to 192 (i.e., three rather than eight experimental blocks of 64 trials); (c) six additional neutral male faces (3 White, 3 Middle-Eastern) were used for practice trials. The changes were done to make the Experiment most similar to Experiment 2 of Kozlik and Fischer (2020).

Results

Trials with incorrect responses (3.54%) were excluded from analyses, as were trials with RTs shorter than 150 ms or longer than 1.5 interquartile ranges above the third quantile with respect to the individual distribution (4.36% of remaining trials; Tukey, 1977). Table 1 shows mean RTs and error rates for the conditions of interest.

The mean difference score (i.e., RT[incongruent] – RT[congruent]) indicating the emotion × ethnicity interaction was $M = -1$ ms ($SD = 13$ ms); the Bayes factor of $BF_{0+} = 8.05$

suggested "substantial evidence" (Jeffreys, 1961) in favor of the (directed) null hypothesis. For the sake of completeness, we additionally report results of the standard 2 (emotional expression: happiness, anger) × 2 (ethnicity: White, Middle-Eastern) repeated measures ANOVA with mean RTs as the dependent variable. This analysis yielded only a main effect of ethnicity $F(1, 38) = 12.55$, $p = .001$, $\eta_p^2 = .248$, indicating that responses to White faces were $M = 11$ ms ($SD = 20$ ms) faster than responses to Middle-Eastern faces. The associated Bayes factor of $BF_{10} = 29.03$ provided "strong evidence" in favor of the alternative hypothesis. Both the main effect of emotion and the interaction did not reach the criterion of significance (both $Fs < 1$); the emotion effect was associated with $BF_{01} = 3.76$, which reflects "substantial evidence" in favor of the null.

An analyses of error rates yielded no significant effects, $F(1, 38) = 2.64$, $p = .113$, $\eta_p^2 = .065$ for the emotion main effect ($BF_{01} = 1.74$), both Fs < 1 for the ethnicity main effect and the interaction ($BF_{01} = 5.02$ and $BF_{0+} = 5.50$, respectively).

Discussion

Experiment 2 was our second attempt to replicate the emotion × ethnicity interaction effect found by Kozlik and Fischer (2020) in their Experiment 2; this attempt also failed. Before drawing firm conclusions, we will first report the initial experiment of the series as Experiment 3. As noted in the *Overview* section, the aim of Experiment 3 was to investigate a potential emotional expression × ethnicity interaction with a different experimental task than the blur task. To this end, we placed clearly visible moles on each face, and participants' task was to decide whether the face had two or three moles. The *a priori* reason behind this change of task was to use a different set of targets (i.e., averted faces) in potential forthcoming experiments without having to change the experimental task. In the context of this paper, Experiment 3 can be seen as an additional attempt to obtain the crucial emotion × ethnicity interaction, with an alternative task.

Experiment 3

Experiment 3 was structurally equivalent to Experiments 1 and 2. However, the task was to categorize the number of moles on the face (2 vs. 3); therefore it can be considered a conceptual replication. As noted above, this was chronologically the first experiment in the series. The pre-registration can be found at https://aspredicted.org/FZ4_XHP.

Method

Participants. The sampling plan was the same as for Experiment 1. The effective sample was $N = 57$ participants (34 females, 23 males; 48 right-handed, 9 left-handed; age $Md = 23$ years, range: 18-35). To achieve this sample size, we recruited a total of 60 participants. Three participants reported a migration background; thus, their data were discarded. Participants were compensated with £3.

Materials. The stimuli were the same male faces (8 White, 8 Middle-Eastern) as in Experiments 1 and 2. We placed easily visible moles on the faces to create two different versions of each image: one with two moles and one with three. Since the skin color of the faces varied, we collected RGB color information from three different parts of each face image (i.e., left cheek, right cheek, and forehead). The darkness of the moles was adjusted to the mean RGB color information of each individual face.

Procedure. Participants were informed that each face stimulus had moles and that their task was to decide whether the presented face had two or three moles, by pressing the "v" or "t" key, respectively. Thus, as in the previous experiments, the emotional expression and ethnicity of the stimuli were task-irrelevant. Participants were instructed to respond as fast and accurately as possible. Throughout the experimental session, the response keys and their respective meanings were presented on the screen (i.e., "V – 2" was presented at the bottom of the screen, and "T – 3" was presented at the top of the screen) to prevent erroneous

answers due to forgetting of the response-key assignment. Participants could take a short break after each block.

The trial sequence began with a centrally presented fixation cross that remained on the screen for 1000 ms. It was replaced by a target face that remained on the screen for 1500 ms. If the response took longer than 2000 ms (from target onset), a warning message (i.e., "Too slow! Please respond faster!") was displayed. The inter-trial interval was 1000 ms (irrespective of response). Practice and experimental phases comprised the same number of trials and blocks as Experiment 1.

Results

As in the previous experiments, trials with incorrect responses (3.4%) were not included in the analyses. Trials with RTs shorter than 100 ms or longer than 1.5 interquartile ranges above the third quantile with respect to the individual distribution (3.96% of remaining trials; Tukey, 1977) were excluded. Table 1 shows mean RTs and error rates for the conditions of interest.

A 2 (emotional expression: happiness, anger) × 2 (ethnicity: White, Middle-Eastern) repeated measures ANOVA with mean RTs as the dependent variable yielded a main effect of emotion, $F(1, 56) = 25.50$, $p < .001$, $\eta_p^2 = .313$, indicating that responses to angry faces were $M = 6$ ms ($SD = 9$ ms) faster than responses to happy faces. The Bayes factor of $BF_{10} = 3529.56$ provided "decisive/extreme evidence" (Jeffreys, 1961; Wagenmakers et al., 2011) in favor of the alternative hypothesis. In addition, the main effect of ethnicity reached significance, $F(1, 56) = 18.76$, $p < .001$, $\eta_p^2 = .251$; responses to White faces were $M = 5$ ms ($SD = 8$ ms) faster than responses to Middle-Eastern faces. The Bayes factor of $BF_{10} = 342.54$ again provided "decisive/extreme evidence" in favor of the alternative hypothesis.

Most importantly, the emotion × ethnicity interaction did not reach the criterion of significance, $F < 1$. The associated Bayes factor analysis of the difference score

RT(incongruent) minus RT(congruent) yielded "strong evidence" in favor of the null hypothesis, $BF_{0+} = 12.87$.

An analyses of error rates yielded no significant effects, $F(1, 56) = 2.12$, $p = .151$, $\eta_p^2 = .036$ for emotion, $F(1, 56) = 1.55$, $p = .218$, $\eta_p^2 = .027$ for ethnicity, and $F(1, 56) = 1.27$, $p = .265$, $\eta_p^2 = .022$ for emotion × ethnicity.

A *post hoc* (i.e., not preregistered) exploratory analysis of RTs included number of moles (2 vs. 3) as an additional factor (see *Appendix*). This suggested that the unexpected emotion effect emerged only for the 2-moles condition, casting doubt on whether it is a true emotion effect (see *Appendix* for a discussion). The emotion × ethnicity interaction was not affected by number of moles, $F < 1$ (both Fs < 1 for emotion × ethnicity within the 2-moles and 3-moles conditions).

Discussion

Experiment 3, which was structurally equivalent to Experiments 1 and 2, again yielded no evidence for the emotion × ethnicity interaction of interest. Thus, we have to state that we failed to find the emotion × ethnicity interaction that is predicted by the PCA, according to Kozlik and Fischer (2020).

We again observed an ethnicity main effect (as in Experiment 1 and 2). Responses were slower for Middle-Eastern faces than White faces. This result was unexpected. Although it does not contribute to the SMA versus PCA debate, it is of interest for the evaluation of the paradigm whether the bias towards Middle-Eastern faces could be caused by group valence (i.e., the prejudiced evaluation) or might be simply based on the perceptual characteristics of the faces (e.g., slightly darker skin tone). To shed some light on the likelihood of the two possibilities, we replicated Experiment 1 with inverted faces (i.e., upside-down faces). Inversion disrupts the holistic processing of faces (Valentine, 1988).

Therefore, a prejudice-related group effect should be less likely. In contrast, low-level perceptual characteristics of faces are still preserved when faces are inverted.

Experiment 4

The pre-registration of Experiment 4 can be found at https://aspredicted.org/99R_QBW.

Method

Participants. Again, we relied on sequential testing with Bayes factors. We aimed for a Bayes factor of either $BF_{01} > 6$ in favor of the null hypothesis of no biased perception of out-group faces or a Bayes factor of $BF_{10} > 6$ in favor of the alternative hypothesis. Since we obtained a critical Bayes factor with the initial sample of 40 participants, we stopped data collection at that point.

The effective sample was $N = 38$ participants (21 females, 14 males, 3 non-binary; 34 right-handed, 4 left-handed; age $Md = 24$ years, range: 18-33). One further participant had more than 20% errors (across all trials), and another participant's mean reaction time was a "far-out value" with regard to the grand mean RT (see *Pre-registration* for the outlier criteria). Therefore, the data of these two participants were discarded before analysis. Participants were compensated with £3.

Materials and Procedure. The same happy and angry Middle-Eastern and White male faces as in Experiment 1 were used; the procedure and trial sequence were also identical to Experiment 1. The only difference between Experiments 1 and 4 were the stimuli: In Experiment 4, inverted face stimuli were used (i.e., faces rotated by 180°).

Results and Discussion

Trials with incorrect responses (2.42%) were excluded. The same outlier criteria as in the previous experiments were used (4.38% of remaining trials). Table 1 shows mean RTs and error rates for the conditions of interest.

The Bayes factor for the main effect of ethnicity suggested "decisive evidence" in favor of the alternative hypothesis, BF_{10} = 6177.21, with responses to White faces M = 5 ms (SD = 6 ms) faster than responses to Middle-Eastern faces. The main effect of emotion, as well as the interaction effect of emotion × ethnicity, were associated with "substantial evidence" in favor of the null hypothesis, BF_{01} = 3.32 and BF_{0+} = 6.51, respectively.

The analyses on error rates were in line with the previous experiments. None of the effects indicated a hint in favor of alternative hypothesis. The emotion main effect, the ethnicity main effect, and the interaction effect were associated with Bayes factors of BF_{01} = 2.63 , BF_{01} = 5.73 , and BF_{0+} = 4.75 , respectively.

Experiment 4 thus revealed that the main effect of ethnicity found in the foregoing experiments is most likely due to perceptual features (e.g., slightly darker skin), differentiating White and Middle-Eastern faces. We therefore refrain from a discussion of this effect as potentially reflecting prejudice.

General Discussion

The present study aimed at examining the validity of the processing-conflict account (PCA), introduced by Kozlik and Fischer (2020), in the unrelated-task paradigm. Specifically, our long-term goal was to pit the PCA and the social-message account (SMA) against each other in a single experiment that used frontal and profile-view faces. However, we already failed to replicate the critical emotion × ethnicity interaction with frontal-view faces: In contrast to Kozlik and Fischer's Experiment 2, we found clear null results in three experiments. In fact, for the full data set of Experiments 1 to 3 (N = 154), the Bayes factor in favor of the null hypothesis was BF_{0+} = 19.22 ("strong evidence"; Jeffreys, 1961). Even if the data of Kozlik and Fischer's Experiments 2 and 4_{B1} are added to this data set (i.e., total N = 216), the Bayes factor in favor of the null hypothesis is BF_{0+} = 8.11 ("substantial evidence"; BF_{0+} = 6.46 for all frontal-view experiments, i.e., Kozlik & Fischer's Exp. 2

included, Exp. 4_{B1} excluded, $N = 189$). Thus, which conclusions can be drawn based on this evidence?

To recapitulate, our starting point was to differentiate two pillars of argumentation in favor of the PCA that was proposed by Kozlik and Fischer (2020). One pillar was associated with a variant of the Stroop task, in which response compatibility can be assumed to be the mechanism underlying the observed congruency effects: In this task, stimuli (here: faces) must be categorized according to a certain feature (here: valence of the emotional expression); another feature (here: ethnicity) is either congruent or incongruent with the response to be given, if we assume that the processing of faces involves prejudiced associations. With this task, Kozlik and Fischer found clear evidence for congruency effects. This is an important result and we believe in its validity; however, it is not a surprising finding because response compatibility is a well-known mechanism in such Stroop-like paradigms (including evaluative priming). However, the result does not demonstrate that involuntary processing of *both* features triggers a processing conflict (in case of incongruency), as postulated by Kozlik and Fischer. What the result from the Stroop-like task shows is that one feature – which is processed involuntarily even though it is task-irrelevant (here: ethnicity) – potentially conflicts with the intentionally generated response to the stimulus. Of course, the basis of this response is the other feature – that is, the emotional expression – but this can be extracted from the typical perceptual features of happy versus fearful/angry faces. Therefore, in our view the result does not conflict with the SMA.

The second pillar of argumentation of Kozlik and Fischer (2020) is the more interesting one with regards the SMA. It rests on results from what we termed here the unrelated-task paradigm: In this task, the two features of interest are both task-irrelevant. Kozlik and Fischer argued that congruency effects result from a conflict of the two features despite their task-irrelevance. The conflict then slows down responses in a task that is

focused on the stimulus (i.e., the face itself is task-relevant), but which is neutral with regard to the two features (i.e., both features are task-irrelevant). However, we pointed out that the evidence provided by Kozlik and Fischer using this paradigm was inconclusive: a congruency effect was found in Experiment 2 with frontally viewed faces, whereas the effect was absent in Experiment 4_{B1} (with averted faces; according to our reanalysis presented in this article). As mentioned earlier, the PCA would not predict a difference between frontal and profile-view faces, so this evidence is somewhat problematic for the PCA if we take the results at face value.

By contrast, the SMA is compatible with the findings of Kozlik and Fischer (2020), that is, a congruency effect in Experiment 2,[11] but no congruency effect in Experiment 4_{B1}. Thus, we aimed to provide more conclusive evidence by following a step-by-step plan: *If* the result of Experiment 2 of Kozlik and Fischer had replicated, we would have conducted an experiment with the full view (i.e., frontal vs. profile-view) × emotion × ethnicity design. However, in three attempts to replicate Experiment 2 of Kozlik and Fischer, we obtained clear evidence that the congruency effect as observed by Kozlik and Fischer did not replicate.

Thus, if we take these failures to replicate Experiment 2 of Kozlik and Fischer (2020) and the null result found in their Exp. 4_{B1} seriously, what remains from Kozlik and Fischer is the data from the Stroop-like paradigm, which provided clear results (at least in Exp. 1 and Exp. 3; for Exp. 4_{B2}, see Footnote 4 above). As already argued in the *Introduction*, this paradigm can be considered an ingenious variant of the evaluative-priming paradigm for assessing involuntarily evoked prejudices, in line with the seminal study by Fazio et al.

[11] But note again that from the SMA perspective, a congruency effect in Experiment 2 would be a risky prediction because it would be based on an emotional-Stroop interpretation of the paradigm (see *Introduction*). The emotional Stroop paradigm has some intricacies (Frings et al., 2010); thus, a null result found with this paradigm would not be at odds with the SMA.

(1995); however, results from this paradigm cannot be taken as evidence in favor of the PCA and against the SMA.

We should hasten to add that the null results provided in this article do not falsify the PCA. At several points in the text we made clear that it was a risky prediction from the start on, given previous research on (a specific version of) evaluative priming, to expect that congruency (versus incongruency) of two task-irrelevant features will influence the responses with regard to a third feature (i.e., side of blurredness). That is, the paradigm might not be suited at all to test PCA. The null results (together with our reinterpretation of the Stroop-like experiments) simply mean that the PCA is yet no supported by distinct empirical evidence.

Interestingly, however, evidence for *a basic premise* of the PCA was found in a series of experiments by Paulus and Wentura (2018). The premise is that the two face features of interest (i.e., emotional expression and ethnicity) are independently and involuntarily extracted – at least in some specific circumstances. Paulus and Wentura presented faces varying in emotion and ethnicity as primes that preceded positive and negative target images, which had to be categorized accordingly. The authors found two independent priming effects for emotional expression (i.e., happy and fear as positive and negative primes, respectively) and ethnicity (i.e., White and Turkish/Middle-Eastern faces as positive and negative primes, respectively). Thus, in the context of this paradigm, it seems that the features are independently processed, differently weighted, and not integrated.[12]

[12] For the sake of completeness and transparency: Weisbuch and Ambady (2008), who introduced the SMA, first provided evidence for the SMA from the evaluative-priming paradigm. Their results were compatible with the SMA in that priming effects for in-group primes conformed to the nominal evaluation of emotions (i.e., happy vs. fearful faces acted as positive vs. negative stimuli, respectively); the effect reversed for out-group faces (i.e., in this case happy vs. fearful faces acted as negative vs. positive stimuli, respectively). However, Craig et al. (2014) did not replicate this result; they only found a priming effect for emotional expression. The study by Paulus and Wentura (2018) was an attempt to clarify the issue with three high-powered experiments.

However, we recently identified a further context (beyond the approach/avoidance task; Paulus & Wentura, 2014; Wentura & Paulus, 2022) that yielded results compatible with the SMA, although both features were task-irrelevant (Gurbuz et al., 2023). Using the Extrinsic Affective Simon Task (EAST; De Houwer, 2003), we again found evidence for an emotion × ethnicity interaction in involuntary evaluations that conform to the SMA.

It is important to report this experiment in more detail because it resembles Experiment 2 of Kozlik and Fischer (2020) and our present experiments in that participants were presented with faces (varying in emotion and group) and had to categorize side of blurredness. However, here the face trials were mixed with evaluation trials, that is, positive and negative words had to be categorized according to their valence, using the same keys as in face trials. The basic assumption of the EAST is that keys acquire a positive and negative meaning by these evaluation trials. Thus, responses to faces (based on side of blurredness) can be interpreted as giving a positive or negative evaluative response to them. Depending on match between response and evaluation of the face, responses are facilitated or slowed down. Using this task, Gurbuz et al. found an emotion × group × (key) valence interaction that fit the predictions of SMA.

Can PCA explain this result as well? Essentially not. In a nutshell (for a detailed discussion see Gurbuz et al., 2023), one can say, first, that disregarding the evaluation trials, the experiment was a replication of Experiment 2 by Kozlik and Fischer (and the present ones); however, no emotion × ethnicity interaction was found. (Note, this null result was found despite the fact that the significant emotion × group × key valence interaction indicated that emotional expression and group status – although task-irrelevant – were processed in these trials.) Second, since PCA is based on the idea that the two valent features (i.e., group and emotion) are independently extracted, one could have alternatively expected emotion × (key) valence and ethnicity × (key) valence interactions, which were not found. Third, PCA is

consistent with the results of the EAST experiment (i.e., the emotion × group × key valence interaction) only if we assume that the extraction of two negative features (i.e., fearfulness, outgroup) does not lead to a facilitation of negative responses, whereas "second-order valence" (i.e., positivity resulting from the match of two negative features) leads to facilitation of a positive response. This is not very plausible.

Interestingly, in both the approach/avoidance task and the EAST, the face stimulus itself is task-relevant (but not the critical features), which does not hold for the evaluative-priming paradigm. If task-relevance of the face stimulus constitutes a boundary condition of the SMA, it is a rather meaningful one: Task relevance could be interpreted as "imitating a communication situation", as rudimentary as this imitation might be. However, there seem to be other contexts – e.g., the evaluative priming paradigm – that support the independent extraction of the critical features (i.e., emotional expression and group), which is a basic premise of PCA. It is conceivable that a paradigm can be developed that shows that the two features produce conflict (in the sense predicted by PCA) under certain circumstances. However, the paradigms suggested in Kozlik and Fischer (2020) seems to be not well suited to show this.

References

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General, 133*(3), 323-338. https://doi.org/10.1037/0096-3445.133.3.323

Bugg, J. M., & Crump, M. J. C. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00367

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*(5), 865-877. https://doi.org/10.1037/a0037270

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*(2), 77-85. https://doi.org/10.1026//1618-3169.50.2.77

Degner, J., & Wentura, D. (2011). Types of automatically activated prejudice: assessing possessor- versus other-relevant valence in the evaluative priming task. *Social Cognition, 29*(2), 182-209. https://doi.org/10.1521/soco.2011.29.2.182

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*(3), 245-256. https://doi.org/10.1080/01973530701503150

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The heterogeneous world of congruency sequence effects: an update. *Frontiers in Psychology, 5*, 9, Article 1001. https://doi.org/10.3389/fpsyg.2014.01001

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*(2), 175-191. https://doi.org/10.3758/BF03193146

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*(6), 1013-1027. https://doi.org/10.1037/0022-3514.69.6.1013

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*(2), 229-238. https://doi.org/10.1037/0022-3514.50.2.229

Frings, C., Englert, J., Wentura, D., & Bermeitinger, C. (2010). Decomposing the emotional stroop effect. *The Quarterly Journal of Experimental Psychology, 63*(1), 42-49. https://doi.org/10.1080/17470210903156594

Gast, A., Werner, B., Heitmann, C., Spruyt, A., & Rothermund, K. (2014). Evaluative stimulus (in)congruency impacts performance in an unrelated task. Evidence for a resource-based account of evaluative priming. *Experimental Psychology, 61*(3), 187-195. https://doi.org/10.1027/1618-3169/a000238

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information - strategic control of activation of responses. *Journal of Experimental Psychology-General, 121*(4), 480-506. https://doi.org/10.1037//0096-3445.121.4.480

Gurbuz, E., Paulus, A., & Wentura, D. (2023). Involuntary evaluation of others' emotional expressions depends on the expresser's group membership. Further evidence for the social message account from the extrinsic affective Simon task. *British Journal of Social Psychology, 62*(2), 1056-1075. https://doi.org/https://doi.org/10.1111/bjso.12619

Hermans, D., Van den Broeck, A., & Eelen, P. (1998). Affective priming using a colour-naming task: A test of an affective-motivational account of affective priming effects. *Zeitschrift fur Experimentelle Psychologie, 45*(2), 136-148.

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., Taylor, J. H., & Crites, S. L. (2013). On the automatic activation of attitudes: a quarter century of evaluative priming research. *Psychological Bulletin, 139*(5), 1062-1089. https://doi.org/10.1037/a0031309

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

Kozlik, J., & Fischer, R. (2020). When a smile is a conflict: affective mismatch between facial displays and group membership induces conflict and triggers cognitive control. *Journal of Experimental Psychology-Human Perception and Performance, 46*(6), 551-568. https://doi.org/10.1037/xhp0000732

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00863

McKenna, F. P., & Sharma, D. (1995). Intrusive cognitions: An investigation of the emotional Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1595-1607. https://doi.org/10.1037/0278-7393.21.6.1595

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology, 104*(1), 45-69. https://doi.org/10.1037/a0030734

Neumann, R., & Seibt, B. (2001). The structure of prejudice: Associative strength as a determinant of stereotype endorsement. *European Journal of Social Psychology, 31*(6), 609-620. https://doi.org/10.1002/ejsp.69

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*(4), 656-677. https://doi.org/10.1080/02699931.2013.849659

Paulus, A., & Wentura, D. (2018). Implicit evaluations of faces depend on emotional expression and group membership. *Journal of Experimental Social Psychology, 77*(2), 143-154. https://doi.org/10.1016/j.jesp.2018.04.004

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*(3), 380-391. https://doi.org/10.1037/0022-3514.61.3.380

Rohr, M., & Wentura, D. (2022). How emotion relates to language and cognition, seen through the lens of evaluative priming paradigms. *Frontiers in Psychology*, *13*, Article 911068. https://doi.org/10.3389/fpsyg.2022.911068

Rothermund, K., & Wentura, D. (1998). Ein fairer Test für die Aktivationsausbreitungshypothese: affektives Priming in der Stroop-Aufgabe [An unbiased test of a spreading activation account of affective priming: Analysis of affective congruency effects in the Stroop task]. *Zeitschrift fur Experimentelle Psychologie*, *45*(2), 120-135.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662. https://doi.org/10.1037/h0054651

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471-491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

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*(3), 426-432. https://doi.org/10.1037/a0022790

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*(1), 22-36. https://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*(5), 1063-1079. https://doi.org/10.1037/a0011993

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). Guilford.

Wentura, D., & Paulus, A. (2022). Social message account or processing conflict account – which processes trigger approach/avoidance reaction to emotional expressions of in- and out-group members? *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.885668

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*(6), 1024-1037. https://doi.org/10.1037/0022-3514.78.6.1024

Tables

Table 1

Mean Reaction Times in ms (Error Rates [%] in Parentheses) as a Function of Congruency for Experiments 2 and 4_{B1} of Kozlik and Fischer (2020)

Ethnicity	*Experiment 2*	*Experiment 4_{B1}*
Congruent	527 (1.50)	420 (1.35)
Incongruent	533 (1.91)	420 (1.22)
Congruency Effect	5 (0.41)	0 (-0.13)
SE of Congr. Effect	1 (0.31)	3 (0.36)

Congruent = happy/White, angry/Middle-Eastern; Incongruent = happy/Middle-Eastern, angry/White; Congruency effect = incongruent – congruent

Table 2

Mean Reaction Times (Error Rates [in %] in Parentheses) as a Function of Emotion and Ethnicity for Experiments 1, 2, and 3

		Experiment			
Ethnicity	*Emotion*	*1*	*2*	*3*	*4*
White	Happiness	546 (2.49)	527 (3.10)	596 (3.55)	524 (2.26)
	Anger	548 (2.91)	528 (3.74)	589 (2.96)	525 (2.57)
Middle-East.	Happiness	555 (2.68)	538 (3.37)	600 (3.58)	529 (2.32)
	Anger	557 (2.77)	540 (3.95)	595 (3.52)	531 (2.51)
Congruency Effect		0 (0.17)	-1 (0.03)	-1 (-0.27)	0 (0.06)
SE of Congr. Effect		1 (0.18)	2 (0.39)	1 (0.24)	1 (0.27)

Congruency effect = mean(happy/Middle-Eastern, angry/White) – mean(happy/White, angry/Middle-Eastern)

Appendix

A *post hoc* exploratory analysis of Experiment 3

For Experiment 3, we added a *post hoc* (i.e., not preregistered) exploratory analysis of RTs with number of moles (2 vs. 3) as an additional factor to shed some light on the significant main effect of emotion. This effect was not significant in Experiments 1 and 2, and it was in an unexpected direction in Experiment 3 (i.e., typically, happy faces are associated with faster responses). There was no main effect of the number of moles, $F(1, 56) = 1.97$, $p = .166$, $\eta_p^2 = .034$, but significant interactions with emotion, $F(1, 56) = 5.11$, $p = .028$, $\eta_p^2 = .084$, and ethnicity, $F(1, 56) = 4.16$, $p = .046$, $\eta_p^2 = .069$.

For the 2-moles conditions, the emotion effect, $F(1, 56) = 35.30$, $p < .001$, $\eta_p^2 = .387$, was significant but not the ethnicity effect, $F(1, 56) = 2.47$, $p = .122$, $\eta_p^2 = .042$. The pattern reversed for the 3-moles condition, $F(1, 56) = 2.29$, $p = .136$, $\eta_p^2 = .039$ for emotion, and $F(1, 56) = 15.35$, $p < .001$, $\eta_p^2 = .215$ for ethnicity. The emotion × ethnicity interaction was not affected by number of moles, $F < 1$ (both Fs < 1 for emotion × ethnicity within the 2-moles and 3-moles conditions).

The results might reveal a weakness of the task: Our *post-hoc* exploratory analysis indicates that the effect of emotional expression is only present in the 2-moles condition. In fact, to respond with “Two moles!” means to *negate* the presence of a third mole. Thus, this resembles the situation in a “target absent” visual search condition. Response time in this case clearly depends on the complexity of the display. Although this statement does not straightforwardly explain why happy expressions produced slower responses, it casts doubts on any attempt to interpret the result as a genuine emotion effect (and not as an effect of typical perceptual features of happy and angry faces). Thus, in any future attempt to use the task, one might a priori declare the 2-moles condition as a filler condition and restrict

analyses to the 3-moles condition. *Post hoc* we can state that the 3-moles condition produced exactly the same results pattern that we found in Experiments 1 and 2.