Received: 20 February 2022 | Accepted: 7 December 2022

DOI: 10.1111/bjso.12619

ARTICLE

the british psychological society
promoting excellence in psychology

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

Emre Gurbuz | Andrea Paulus | Dirk Wentura

Department of Psychology, Saarland University, Saarbruecken, Germany

Correspondence
Emre Gurbuz and Dirk Wentura, Department of Psychology, Saarland University, Campus A2.4, Saarbruecken D-66123, Germany.
Email: emre.gurbuz@uni-saarland.de and wentura@mx.uni-saarland.de

Funding information
German Research Foundation, Grant/Award Number: WE 2284/14–2

Abstract

The social message account (SMA) hypothesizes that the evaluation of emotional facial expressions depends on the ethnicity of the expressers. For example, according to SMA, a happy face of a member of a prejudiced ethnicity is immediately interpreted as potentially malevolent. Evidence for this approach was found initially in evaluative priming (EP) and approach-avoidance tasks (AA) by showing an emotion × ethnicity interaction on positivity scores (EP) and approach scores (AA), respectively. Recently, attempts to replicate the EP results failed. Due to the inconclusive EP results, it was important to examine the influence of ethnicity on processing of emotional expression with another task testing involuntary evaluations. The extrinsic affective Simon task was used with stimuli varying on emotion (happy vs. fear) and ethnicity (White-Caucasian vs. Middle-Eastern men). This task was chosen because in contrast to EP (where faces are presented as task-irrelevant primes) faces are task-relevant. Experiment 1 yielded an emotion × ethnicity interaction with regard to positivity scores that fit SMA predictions. The results are also important in challenging a recent theoretical alternative to SMA, namely the processing conflict account. A generalization of the emotion × ethnicity pattern to learned arbitrary in- and out-groups (Experiment 2) failed, suggesting that involuntary processing of (task-irrelevant) group status depends on perceptual features.

KEYWORDS
emotional expression, extrinsic affective Simon task, group membership, involuntary evaluation, social message account

BACKGROUND

In our daily life, others' facial expressions help us to regulate our behaviour (Salovey & Mayer, 1990). For instance, seeing a friend's smiling or fearful face would trigger different involuntary automatic reactions like a spontaneous positive evaluation of the situation in the former or a spontaneous negative evaluation in the latter case. But how are emotional faces processed that contain a second feature that is evaluatively positive or negative, such as group status (e.g. a face might belong to a person's ethnic in-group vs. an ethnic out-group)? Recently, several studies have provided different answers to this question.

Paulus and Wentura (2014) used happy and fearful emotional expressions of White Caucasian and Middle-Eastern young men to see whether ethnicity and emotional expression interact and affect automatic approach/avoidance behaviour in White Caucasian participants. Note, there is evidence for prejudice towards Middle-Eastern men in Germany (Degner et al., 2007; Degner & Wentura, 2011; Neumann & Seibt, 2001; Wagner et al., 2003). The experiment yielded a group × emotion × response type interaction, indicating that happy in-group faces and fearful out-group faces activated approach behaviour whereas fearful in-group faces and happy out-group faces activated avoidance behaviour. The authors explained this interaction with a social message account (SMA; Weisbuch & Ambady, 2008).

According to the SMA, the same emotional expression can trigger conflicting behaviours depending on the expresser's group membership, because emotional expressions from in-groups and out-groups convey disparate social messages that affect the evaluation of these in-group and out-group emotions (Paulus & Wentura, 2014). In general, from the viewpoint of the prejudiced observer, in-group expressers tend to signal benevolent intentions whereas out-group members tend to signal malevolent intentions. More specifically, an in-group member's smile is generally taken to indicate affiliation intentions and an in-group member's fear is generally taken to indicate warning intentions. Therefore, while in-group smiles trigger positive affect, fearful expressions of in-group members trigger negative affect. The same emotional expressions would be associated with the reverse pattern of intentions when the emotional expresser is an out-group member: An out-group member's smile is generally taken to indicate dominance intentions and an out-group member's fear is generally taken to indicate submission intentions. Therefore, while out-group smiles trigger negative affect, fearful expressions of an out-group member trigger positive affect.

Thus, the results of the work of Paulus and Wentura (2014) can be explained by the SMA: in-group joy activates affiliation intentions (and therefore facilitates approach behaviour) whereas in-group fear activates warning intentions (and therefore facilitates avoidance behaviour). On the other hand, out-group joy signals dominance intentions (and therefore facilitates avoidance behaviour) whereas out-group fear signals submission intentions (and therefore facilitates approach behaviour; for further evidence, see the work of Paulus et al., 2019). The main point of interest here is the assumption that the two features of the face, that is, emotional expression and group membership, are immediately integrated into a new "social meaning" feature.

The study by Paulus and Wentura (2014) was based on earlier work by Weisbuch and Ambady (2008), who used a different approach to test the SMA. They argued that the moderation effect that ethnicity has on the affect elicited by emotional expressions should be directly assessable in an evaluative priming paradigm, which provides a measure of fast and involuntary evaluations (Fazio et al., 1986). They used in-group and out-group faces with positive and negative emotional expressions as prime stimuli that preceded positive and negative target stimuli, which had to be categorized according to valence. The prototypical effect in this paradigm is a congruence effect, that is, if primes and targets match in valence, faster and/or more accurate responses are expected (compared to mismatches). In other words: one can infer from the observed effect whether a prime is involuntarily evaluated as positive or negative.

Weisbuch and Ambady (2008) observed an interactive influence of emotion and group on evaluative responses. That is, in-group happiness and out-group fear acted as (relatively) positive primes whereas in-group fear and out-group joy acted as (relatively) negative primes. Thus, emotional expression and group membership were integrated in a way that conforms to the social message account.

In fact, the results of Weisbuch and Ambady (2008) served as the foundation of the approach/avoidance studies by Paulus and Wentura (2014): Assuming fast and involuntary stimulus evaluations in accordance with the SMA leads directly to Paulus and Wentura's hypothesis of corresponding behavioural tendencies. However, Craig et al. (2014) as well as Paulus and Wentura (2018) did not replicate the findings of Weisbuch and Ambady (2008) with the same setup (i.e. using expressions of happiness and fear from in- and out-group faces as primes). Whereas Craig et al. (2014) found only a priming effect of emotion, Paulus and Wentura (2018) reported two independent priming effects: An effect of emotional expression (i.e. happy and fearful faces acted as positive and negative primes, respectively) and a somewhat weaker effect of group membership (i.e. White Caucasian and Middle Eastern faces acted as negative and positive primes, respectively).

Thus, we are faced with diverging results: In their approach and avoidance study, Paulus and Wentura (2014) observed an interaction of group and emotion, whereas the (more recent) evaluative priming studies – which focused on involuntary evaluation (i.e. the valence of the face stimuli) – only yielded two main effects that corresponded to (a) the apparent evaluation of emotional expressions and (b) a group prejudice effect (Craig et al., 2014, as well as Paulus & Wentura, 2018). Thus, the Weisbuch and Ambady (2008) results were questioned by the later attempts to replicate them. Therefore, important support for the approach/avoidance studies of Paulus and Wentura (2014) has been removed. It might be worth to replicate the emotion × group effect with an alternative measure, that is, the Extrinsic Affective Simon Task (EAST). The choice of the EAST was not accidental because in contrast to evaluative priming and affective misattribution (Payne et al., 2005), in the EAST the stimuli of interest itself are task-relevant whereas the features of interest (here: ethnicity and expression) are task-irrelevant. Thus, finding support for the approach/avoidance result with the EAST was more likely (compared to, e.g. evaluative priming) because the approach/avoidance task as used by Paulus and Wentura (2014, Exp. 1) shares these characteristics (i.e. task-irrelevance of features, task-relevance of stimulus). Moreover, it would not be a major limitation to the validity of the SMA if the approach/avoidance result will be confirmed with the EAST only: It is plausible that social message processing of faces might be limited to task-relevant stimuli (because they mimic a kind of communication situation). Or in other words: If the stimulus is not task-relevant, a social message effect might be more fragile, as shown by the inconclusive results from the evaluative priming studies. Future research may use task-relevance of faces as a factor to examine the divergent results. Here, however, we pursue a different path: Can we establish a new foundation for the approach/avoidance study by Paulus and Wentura (2014) by providing evidence for involuntary evaluation of faces using a different paradigm, namely the EAST? To summarize, the main reason to conduct this study was to replicate the emotional expression × ethnicity interaction effect, which supports the social message account, with another "implicit" task, namely the extrinsic affective Simon task.

Another important point to mention is the group membership defining factor. The majority of studies reported here used ethnicity as the group membership defining factor. This is because the SMA addresses emotion × group interactions for groups that are the targets of (relatively) specific prejudices, that is, prejudices that classify the out-group as hostile. For instance, Weisbuch and Ambady (2008) and Craig et al. (2014) conducted their studies in the USA and Australia, respectively, where Blacks were the ethnic groups that prototypically stand for being the target of such prejudices. In Germany, similar prejudices are those towards Turkish/Middle-Eastern men (Asbrock, 2010; Degner et al., 2007; Degner & Wentura, 2011; Neumann & Seibt, 2001; Wagner et al., 2003). Therefore, SMA-related research conducted in Germany with a German sample used Turkish/Middle Eastern male faces as the out-group representation (i.e. see Kozlik & Fischer, 2020; Paulus & Wentura, 2014, 2018). To be in line with this recent research, we used ethnicity as the group membership defining factor in Experiment 1 as well. To further analyse whether the emotion × group interaction can be found in learned arbitrary in- and out-groups, we used a modified minimal group paradigm in Experiment 2.

The extrinsic affective Simon task

In the extrinsic affective Simon paradigm (De Houwer & Eelen, 1998; Degner & Wentura, 2008), there are two types of trials that are intermixed: In evaluation trials, positive and negative stimuli (e.g. words) have to be categorized according to valence by pressing one of two keys on a keyboard. As a result, the response keys are extrinsically associated with negative and positive valence. The remaining trials present the attitude-related stimuli of interest (here: faces). Participants categorize a feature of the stimuli that is varied orthogonally to their (presumed) valence (e.g. colour), using the same valence-associated keys used in the evaluation trials. If the valence of the stimulus and response key match, responses are faster compared to a mismatch.

As a proof of concept, (De Houwer, 2003) instructed participants to categorize positive and negative words based either on their meaning or their colour, using response keys P and Q. Half the words were white and required valence classification (i.e. press P for positive and Q for negative words or vice versa); the other half were coloured (i.e. either blue or green) and required colour classification (i.e. press P for blue and Q for green words or vice versa). Importantly, participants were instructed to disregard the valence of the coloured words (i.e. coloured-word valence was task-irrelevant) and to focus only on their colour. Thus, participants used the negatively and positively connoted response keys to assess a stimulus feature that was varied orthogonally to valence (i.e. colour). Results indicated that responses to positive coloured words were faster on trials in which the correct response involved the “positive” response key rather than the “negative” response key. The reverse pattern was observed for the negative coloured words (De Houwer, 2003). This result indicates that the affective meaning of the coloured words was involuntarily processed and interfered with task performance even though it was task-irrelevant. Hence, like the evaluative priming paradigm, the EAST tests for involuntary evaluations. Therefore, finding an ethnicity × emotion interaction in the EAST would provide a new foundation for the approach/avoidance task in the sense that involuntary evaluations (as assessed by the EAST) would underlie the behavioural tendencies (as assessed by the approach/avoidance task).

Notably, one aspect that the EAST and the approach/avoidance task have in common is that both require participants to respond directly to the stimuli of interest (i.e. the face stimuli are task-relevant) while the features of interest (i.e. emotional expression, ethnicity) are irrelevant for the response. By contrast, in the evaluative priming task, both the features of interest and the stimuli themselves are task-irrelevant. Thus, using the EAST as an alternative measure of unintentional evaluation was not arbitrary: If there is an ethnicity × expression interaction in the EAST, one might tentatively infer that the social message is only extracted if the critical stimulus itself is task-relevant.

Finally, use of the EAST acquired additional relevance post hoc: While planning, preparing, and conducting the present experiments, an article by Kozlik and Fischer (2020) was published that tackled the SMA by introducing an alternative account, the processing conflict account (PCA). In a nutshell, the authors argued that if a stimulus has two valent features—such as ethnicity and expression—the congruence or incongruence of the two features influences behavioural responses. The approach/avoidance pattern found by Paulus and Wentura (2014) can thus be reinterpreted as follows: the combinations in-group/fearful expression and out-group/happy expression facilitate avoidance behaviour because in both cases the two features are valence-incongruent (i.e. a processing conflict occurs); this is not the case for the reverse combinations. We will postpone a discussion of how EAST results might contribute to the debate of whether SMA or PCA is the better theory to our *General Discussion*, given the post hoc nature of this interpretation (also see Wentura & Paulus, 2022).

EXPERIMENT 1

As explained earlier, the EAST used in Experiment 1 comprised two different types of trials. In word-evaluation trials, participants were instructed to categorize negative and positive adjectives according to their valence. Thereby, the two response keys were extrinsically associated with positive and negative valence, respectively. These keys were also used for the second type of trials, the attitude-related face trials. In the face trials, faces of White Caucasian and Middle-Eastern men with fearful and happy emotional expressions were used as stimuli of interest. Face stimuli were slightly blurred on the right or the left side (following Paulus & Wentura, 2014); participants were instructed to categorize the face stimuli based on this arbitrary feature, which was varied orthogonally to the critical variables (i.e. emotional expression and ethnicity).

Degner and Wentura (2008) found that EAST effects were more pronounced in task-switch trials, that is, in attitude-related trials that followed an evaluation trial rather than another attitude-related trial. The authors provided two possible interpretations of this phenomenon: First, the association between response keys and valence may be stronger immediately after an evaluation trial. Performing an attitude-related trial (i.e. a non-valence-related categorization) might reprogram the key assignment such that the valence association is weaker in the subsequent trial. Second, the evaluation task might carry some inertia such that EAST effects result from carry-over effects of the evaluation task set to attitude-related trials immediately after task switches. Because of this inertia, the valence of stimuli may still be processed and affect responses on attitude-related trials. Whatever the correct interpretation, the results of (Degner & Wentura, 2008) suggest that trial sequence (switch versus repetition) should be taken into account in any EAST study.

Method

Participants

To detect an effect of d_Z = .30 (see Paulus & Wentura, 2014; Experiment 1), a sample of 90 participants was needed. Since data collection was online, we assumed a potentially greater number of outliers and therefore recruited a larger number of participants. The effective sample was 117 non-psychology students in the experiment (65 females, 48 males, 4 undisclosed gender; 105 right-handed, 9 left-handed, 3 ambidextrous; age *Md* = 24.57 years, range: 18–35). One subsample was recruited by research assistants at our university using an online electronic sign-up system (*n* = 60). Participants who signed up for the study were sent links by research assistants to complete the study. Another subsample accessed the study via the online recruitment platform Prolific Academic (www.prolific.co; *n* = 57). The data recruitment at the university targeted White Caucasian participants whose mother tongue was German. However, potential participants were not informed about these criteria to avoid highlighting the intergroup aspect of the experiment. Thus, 11 non-native German speakers and/or non-White Caucasian participants completed the experiment; their data were discarded before analysis.[1] In addition, although recruitment targeted participants between the ages of 18 and 35, one participant aged 61 was inadvertently included; their data were also discarded. Recruitment of Prolific Academic subsample involved the same criteria, implemented via custom prescreening on Prolific (i.e. White Caucasian participants aged 18–35, who are native speakers of German and currently tertiary students). The final sample size was *N* = 117. The experiment took approximately 25 min, and participants received €4 for their participation.

Design

We employed a 2 (group membership: White Caucasian vs. Middle Eastern) × 2 (emotional expression: happy vs. fearful) × 2 (response: positive vs. negative) design, with all factors varied within participants. In addition, the factor task-switch versus repetition was included in analyses (Degner & Wentura, 2008).

Materials

For word trials, 10 negative and 10 positive adjectives were taken from Paulus and Wentura (2018). All adjectives were of other-relevant positivity or negativity (Paulus & Wentura, 2018).[2] For face trials, fearful and happy emotional expressions from 10 White Caucasian and 10 Middle-Eastern men were selected. Most stimuli were identical to those used by Paulus and Wentura (2018). Images of the same individuals with neutral facial expressions were used in a practice block. Stimuli came 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. All face stimuli were headshots with a straight head and frontal gaze; they were edited to show only the face and the top of the neck, and were shown on a grey background. The image size was ca. 16 × 12 cm.

Procedure

The study was conducted online. Participants were asked to close all software or applications that could deliver notifications and to turn off or mute their mobile phones. 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) by using the arrow buttons on their keyboard.

In the extrinsic affective Simon task (EAST), participants were informed that faces of young men (with a slight blur on one side of the face) and adjectives 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 of the presented face, or whether the presented adjective was negative or positive in valence. The responses "negative" and "positive" as well as "left" and "right" were assigned to the "T" and the "V" key, respectively. The response assignment for the valence task was counterbalanced between participants. The response assignment for the blur-detection task was constant for all participants (with T assigned to the right and V assigned to the left).

The beginning of a trial was marked by a centrally displayed fixation cross that remained on the screen for 500 ms. It was replaced by the target adjective or target image, which remained on the screen for 1000 ms. Participants were instructed to respond as quickly and accurately as possible. If the given response was incorrect, a red X was displayed below the word/image for 1000 ms to indicate that the response was erroneous. In addition, if the response time was above 2000 ms, a warning message (i.e. "Too slow! Please respond faster!") was displayed until one of the response keys was pressed. A new trial started after an inter-trial interval of 500 ms.

The main part of the EAST consisted of 480 trials, divided into six blocks that comprised 80 trials each. Each block presented 40 target words and 40 target face images. In each block, all adjectives (i.e. 10 positive and 10 negative adjectives) were presented twice. Adjectives were drawn randomly from the list without replacement. This procedure was repeated after all adjectives had been presented once. In each block, all individuals (10 White Caucasians and 10 Middle-Eastern) were shown with both happiness and fearful expression.[3] Right-/left blurring of a given face was balanced across blocks (i.e. if a given individual was shown in the first block with the left-blurred happiness expression, the same individual's right-blurred happiness expression was presented in the second block, and so on).

Throughout the experimental session, to prevent erroneous answers due to forgetting of response-key mappings, mappings were displayed at the top of the screen for the T-key and at the bottom of the screen for the V-key (e.g. "T – Positive/Right and "V – Negative/Left"). Participants could take short breaks after every two blocks.

[2] The final selection included the following words: gierig (greedy), grausam (cruel), boshaft (malicious), gemein (mean), geizig (stingy), aggressiv (aggressive), kriminell (criminal), autoritär (authoritarian), brutal (brutal), treulos (disloyal) for the negative words and human (humane), ehrlich (honest), gütig (kind), gerecht (just), geduldig (patient), sanft (gentle), humorvoll (humorous), tolerant (tolerant), friedlich (peaceful), aufrichtig (sincere) for the positive words.

[3] Due to a programming error two Middle-Eastern face stimuli were not presented. Instead, two other Middle-Eastern face stimuli were presented twice within a block. However, each emotional expression was presented an equal number of times with a White Caucasian and a Middle-Eastern individual within each block.

Before the main part, participants practiced the adjective task with one block of 40 trials followed by a block of 20 face trials. In the face trials, each individual was shown once with a neutral expression, either with left- or right-blurring. The assignment of left- and right-blurring for each stimulus was random with the constraint that each ethnic group had an equal number of left−/right blurring trials.

Demographics were obtained via an online questionnaire at the end of the EAST. Finally, participants were fully debriefed and thanked for their participation.

Results

Only image trials (i.e. trials with emotional expressions) were included in the analysis. Trials with incorrect responses (2.73%) and trials preceded by an erroneous trial[4] (5.21% of the remaining trials) were excluded from analysis, as were trials with RTs below 200 ms or greater than 1.5 interquartile ranges above the third quantile with respect to individual distribution (3.6% of the remaining trials; Tukey, 1977). Table 1 shows mean RTs and error rates for the conditions of interest.

The results of a 2 (emotional expression: happiness, fear) × 2 (ethnicity: White Caucasian, Middle Eastern) × 2 (valence of keys: positive, negative) × 2 (trial sequence: switch, repetition) repeated measures ANOVAs with mean RTs as dependent variable are reported in Table 2, respectively. As can be seen, the analysis yielded only one significant interaction effect involving valence for the reaction time

TABLE 1 Mean RTs (in ms; error rates in parentheses) as a function of task switch, ethnicity; emotional expression, and response-key emotion (Experiment 1)

	Switch trials		Repetition trials	
	Negative key	**Positive key**	**Negative key**	**Positive key**
White Caucasian				
Happiness	575 (2.99)	575 (3.13)	538 (2.02)	538 (1.88)
Fearful	567 (2.36)	574 (4.18)	532 (2.49)	535 (2.87)
Middle Eastern				
Happiness	566 (2.89)	575 (4.37)	538 (1.99)	543 (1.68)
Fearful	563 (1.94)	571 (4.47)	531 (2.07)	530 (2.23)

TABLE 2 Results of the ANOVA for RTs (Experiment 1)

				× Task switch		
	$F(1, 116)$	p	η_p^2	$F(1, 116)$	p	η_p^2
Emotion (Emo)	41.21	<.001	.262	2.62	.108	.022
Ethnicity (Eth)	3.83	.053	.032	3.97	.049	.033
Valence of keys (Val)	3.56	.062	.030	3.53	.063	.030
Emo × Eth	1.63	.204	.014	3.72	.056	.031
Emo × Val	<1			1.36	.245	.012
Eth × Val	1.88	.173	.016	<1		
Emo × Eth × Val	7.84	.006	.063	<1		

[4] This is standard practice in task switching experiments (e.g., Kopp et al., 2020; Mayr & Keele, 2000; Meiran, 2000; Rogers & Monsell, 1995) because the type of the preceding trial is part of the design and should therefore be unambiguous.

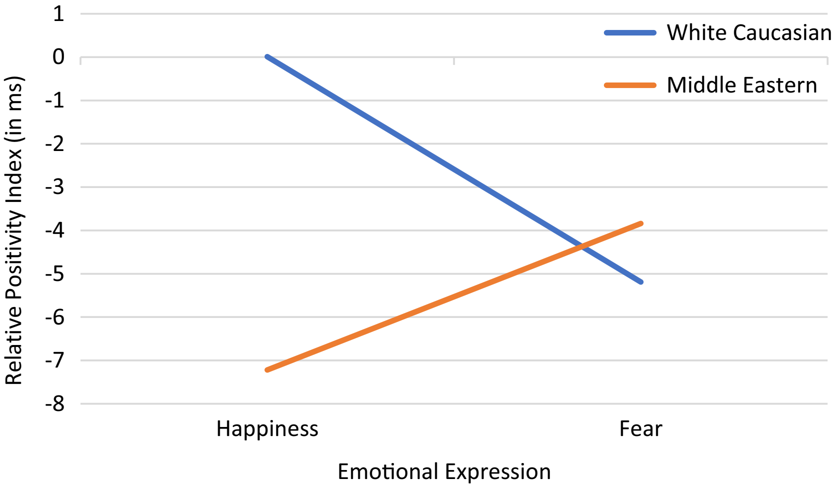

FIGURE 1 Relative positivity scores across emotional expression and ethnicity.

TABLE 3 Results of the ANOVA for error rates (Experiment 1)

				× Task switch		
	F(1, 116)	*p*	η_p^2	*F*(1, 116)	*p*	η_p^2
Emotion (Emo)	1.24	.268	.011	2.70	.103	.023
Ethnicity (Eth)	<1			2.41	.123	.020
Valence of Keys (Val)	9.85	.002	.078	7.99	.006	.064
Emo × Eth	1.97	.163	.017	<1		
Emo × Val	6.69	.011	.055	1.01	.318	.009
Eth × Val	<1			2.39	.125	.020
Emo × Eth × Val	<1			<1		

analyses,[5] that is, the expected emotion × ethnicity × valence interaction effect (see Table 2). The interaction pattern is shown in Figure 1. Depicted are relative positivity scores, that is, difference scores that were obtained by subtracting RTs of the positive key responses from RTs of the negative key responses. The significant interaction supported our hypothesis, indicating that White Caucasian happiness resulted in a higher positivity score ($M = 0$ ms, $SD = 27$ ms) compared to White Caucasian fear ($M = -5$ ms, $SD = 30$ ms; $t(116) = 2.16$, $p = .033$). For Middle-Eastern stimuli, the pattern was numerically reversed: Fearful facial expression resulted in a higher positivity score ($M = -4$ ms, $SD = 29$ ms) compared to happy facial expression ($M = -7$ ms, $SD = 30$ ms); however, the difference did not reach statistical significance, $t(116) = 1.42$, $p = .158$. As an aside, in contrast to Degner and Wentura (2008), the observed pattern was not more pronounced in task-switch trials.

In the analysis with mean error rates as the dependent variable, the three-way interaction of emotional expression × ethnicity × valence was not significant (see Table 3). Thus, we have no reason to suspect a speed-accuracy trade-off behind the crucial triple interaction effect for RTs.

[5] Since the interactions with valence are at the heart of the paradigm and the focus of our hypothesis, we refrain from discussing the significant main effect of emotion.

Discussion

The results of Experiment 1 suggested that the evaluation of emotional expression was modulated by ethnicity (i.e. White Caucasian vs. Middle Eastern). As hypothesized, the affective meaning of the face stimuli was involuntarily processed and interfered with task performance even though it was task-irrelevant. Responses using the key associated with “positive” valence were (relatively) faster for White Caucasian happiness and Middle-Eastern fear than responses using the key associated with “negative” valence. By contrast, “positive” key presses were (relatively) slower for White Caucasian fear and Middle Eastern happiness than “negative” key presses.

Experiment 1 provides evidence for involuntary processing of face valence, with a results pattern that matches the pattern found in the approach/avoidance paradigm used by Paulus and Wentura (2014). Given that the evaluative priming results of Weisbuch and Ambady (2008) failed to replicate (Craig et al., 2014; Paulus & Wentura, 2018), this is an important result because the approach/avoidance results would be puzzling without supporting evidence from a paradigm devoted to assessing involuntary evaluations.

Hitherto, all experiments that we have discussed in detail (including the present Experiment 1) employed ethnicity × emotional expression designs. In their Experiment 2, however, Paulus and Wentura (2014) used a different approach, which involved the on-line creation of participant in- and out-groups. Thus, this experiment manipulated the in−/out-group status of emotional face stimuli using what is known as a (modified) minimal group paradigm (for details, see below). Results obtained with this paradigm were comparable to Experiment 1 of Paulus and Wentura (2014), which used an ethnicity × emotional expression design (albeit with group status as the task-relevant feature instead of the blurring of one side of the face). In Experiment 2, we applied this method and tested for involuntary evaluation effects in the EAST paradigm using a modified minimal group intervention.

EXPERIMENT 2

In Experiment 2, we tested whether the emotion × ethnicity interaction pattern found in Experiment 1 would be observed in an EAST using a modified minimal group paradigm, which created in-groups and out-groups by randomly assigning participants and face stimuli to (fictional) personality styles. This follows the approach of Paulus and Wentura (2014; Experiment 2), however, with one important difference. In Experiment 2 of Paulus and Wentura (2014), group membership was a task-relevant feature. Participants categorized faces into in-group versus out-group by giving approach versus avoidance responses (in a block-wise counter-balanced assignment). In the EAST, by definition, we cannot make the group feature task-relevant (without violating the basic rationale of the EAST).[6]

Method

The experiment was preregistered (see https://aspredicted.org/L87_JR2). There were some (rather minor) deviations from the preregistration. We listed and justified them in Table A1 in Appendix 1.

Participants

The online recruitment platform Prolific Academic (www.prolific.co) was used for data collection. We determined sample size based on the following rationale (see preregistration): The test for a three-way

[6] Making the group feature task-relevant would transform the EAST into a version of the Implicit Association Test (Greenwald et al., 1998): In one block of trials, in-group/positive and out-group/negative would share a key, in another block the assignment would be reversed. This IAT would only test for the valence of the groups at the category level. It cannot be expected that variations of the exemplars of the group (i.e. whether they show a happy or fearful expression) would make a difference (see, e.g. De Houwer, 2001).

interaction in a 2 (group) × 2 (emotional expression) × 2 (response valence) within-subjects design is equivalent to a *t*-test for dependent variables comparing the difference score of (RThappy, neg – RThappy, pos) – (RTfear, neg – RTfear, pos) between in-group and out-group. We expected this score to be larger for the in-group compared to the out-group; due to this directed hypothesis, we planned a one-tailed test. The effect size of the interaction in Experiment 1 was $\eta_p^2 = .063$; this corresponds to $d_Z = .26$. While planning and preregistering the experiment we proceeded (due to preliminary analyses of Experiment 1) from a somewhat smaller effect of $d_Z = .16$. A power analysis indicated that to detect this effect with power $1-\beta = .80$ ($a = .05$; one-tailed) requires a minimum sample size of $N = 243$. The effective sample was 250 participants (141 females, 105 males, 4 participants of undisclosed gender; age $Md = 26.35$ years, range: 18–40). Note that the power to detect an effect of $d_Z = .26$ with $N = 250$ ($a = .05$; one-tailed) was $1-\beta = .99$.

To achieve this sample size, we recruited a total of 315 participants because several of the pre-registered outlier criteria applied: Data were excluded for participants who took more than 120 min to complete the study ($n = 9$), who were not convinced by the (group-defining) personality-style manipulation (see *Procedure*; $n = 10$), who could not remember the in-group label that was assigned to them ($n = 18$), who made more than 4 errors (out of 18 decisions) in the group classification of the face stimuli at the end of the study ($n = 8$), or who had an error rate greater than 20% in the main task ($n = 16$). Finally, data from participants with far-out values with regard to mean RTs (Tukey, 1977; $n = 4$) were discarded. Some participants met two or more exclusion criteria. Participants were compensated with a payment of €5. All participants were White Caucasian (in line with Experiment 1) who were native speakers of English.

An additional 87 participants started the experiment but were unable to successfully complete the face learning task (see *Procedure*); therefore, their experimental session was terminated before the main task (i.e. the EAST) and they received only €2.50.

Design

We employed a 2 (group: in-group vs. out-group) × 2 (emotional expression: happy vs. fearful) × 2 (response: positive vs. negative) repeated-measures design. In line with Experiment 1, the factor task-switch versus repetition was also included in analyses (also see Degner & Wentura, 2008).

Materials

Fear and happiness expressions from eight men and eight women were used, resulting in a pool of 32 images; these were taken from Paulus and Wentura (2014; Experiment 2). Only White Caucasian faces were used. Two stimulus sets comprising the happiness and fear expressions of four women and four men were created. The sets were comparable in terms of emotion recognition rates as well as expression intensity and attractiveness ratings (all *t*s < 1, n.s.). Eight positive and eight negative adjectives were taken from the same word list that was used in Experiment 1.

Procedure

The experiment had three stages: The modified minimal group manipulation, a learning phase, and the EAST.

Minimal group manipulation

The modified[7] minimal group manipulation procedure was identical to the one used by Paulus and Wentura (2014; Experiment 2). Participants were informed that the first part of the study involved the assessment of their personality style. Each participant rated the extent to which 20 statements (e.g. I am often in a bad mood) applied to them personally, on a 7-point scale. Then, mock feedback was given, stating that their personality style was *basal*. Finally, they were provided information about the (fictitious) basal and focal personality styles: "People with a basal personality style are characterized as sociable, agreeable, socially minded and balanced, as well as occasionally imprecise and forgetful. People with the focal style are described as egotistic, reckless, sometimes aggressive, technically skillful, and intelligent." The personality styles were created in a way that most of the students would associate with the basal personality style. In addition, the former clearly had more positive features and the latter had more negative features. A mix of positive and negative features was used for both personality styles to ensure plausibility.

Learning phase

After the minimal group manipulation, participants were informed that the goal of the next part of the study was to examine the influence of personality style on performance in a face-learning task. The face stimuli were presented with neutral expressions, and a (fictional) first name and a group-membership label (i.e. assigned personality style; basal or focal) were displayed below the image. The assignment of face sets to in-group and out-group, respectively, was counterbalanced across participants. After having been presented with each face once, participants completed a challenging learning phase that required categorization of all faces based on personality style. To strengthen the manipulation, two silhouettes were added to the sets, one representing the participant (as a member of the basal group) and—for reasons of symmetry—and one representing another anonymous participant (as a member of the focal group). The learning phase terminated when participants correctly classified all faces (i.e. 18 consecutive trials) or when they completed 27 consecutive trials with no more than one error in three consecutive blocks. If they were unable to achieve this performance level, the experiment was terminated and participants were thanked and received partial payment.

Extrinsic affective Simon task

The procedure of the EAST was the same as in Experiment 1 with the exception of the number of trials. The task now comprised 512 trials, separated into eight blocks of 64 trials. Each block featured 32 target words and 32 target face images. In each block, all adjectives (i.e. 8 positive and 8 negative) were presented twice. The procedure of presenting adjectives and emotional expressions was identical to Experiment 1.

At the end of the experiment, participants answered five questions relating to their identification with the in-group.[8] We also asked participants about their belief in our cover story ("I fully believed the story"; "I had some doubts about the story"; "I did not believe the story from the outset"). Demographics were obtained via an online questionnaire. Finally, participants were fully debriefed and thanked for their participation.

Results

As in Experiment 1, only image trials were included in the analysis. Trials with incorrect responses (7.41%) and trials preceded by an erroneous trial (7.13% of the remaining trials) were excluded from analysis, as were trials with RTs below 200 ms or greater than 1.5 interquartile ranges above the third quantile with

[7] The "modified" refers to the fact that the out-group is explicitly associated with negative attributes (see below).

[8] Items were: (1) "To what extent does the result of the personality test correspond to your own observations about your personality?" (2) "To what extent do individual characteristics of the other personality style also apply to you (i.e. the focal one if you have a basal personality style or the basal one if you have a focal personality style)?", (3) "To what extent do individual characteristics of your own personality style apply to you?" (4) "If you could choose, which personality style would you choose for yourself?" (5) "How much do you like having this personality style?" Scales ranged from 1 ("not at all") to 7 ("fully"), except 1 ("focal") to 7 ("basal") for Item 4.

TABLE 4 Mean RTs (in ms, error rates in parentheses) as a function of task switch, ethnicity, emotional expression, and response-key emotion (Experiment 2)

	Switch trials		Repetition trials	
	Negative key	**Positive key**	**Negative key**	**Positive key**
In-group				
Happiness	636 (9.75)	634 (7.96)	603 (6.01)	601 (4.73)
Fearful	632 (9.25)	629 (9.25)	594 (5.35)	595 (5.73)
Out-group				
Happiness	637 (9.11)	630 (9.94)	604 (5.84)	601 (5.31)
Fearful	628 (8.27)	629 (10.11)	594 (4.72)	594 (5.80)

TABLE 5 Results of the ANOVA for RTs (Experiment 2)

				× Task switch		
	$F(1, 249)$	p	η_p^2	$F(1, 249)$	p	η_p^2
Emotion (Emo)	51.88	<.001	.172	3.88	.049	.015
Group (Gr)	<1			1.23	.268	.005
Valence of Keys (Val)	<1			<1		
Emo × Gr	<1			<1		
Emo × Val	3.41	.07	.014	<1		
Gr × Val	<1			<1		
Emo × Gr × Val	1.31	.254	.005	<1		

Note: As noted in the main text, the inclusion of the task-switching factor was not mentioned in the preregistration. The F-values (p-values) of an analysis with aggregate variables that were created without considering task-switching were 48.63 (<.001) for Emotion, 1.43 (.233) for Group, 1.20 (.275) for Val, 3.05 (.082) for Emo × Val, Fs < 1 for the remaining effects.

respect to individual distribution (3.74% of the remaining trials; Tukey, 1977). Table 4 shows mean RTs and error rates for the conditions of interest.

A 2 (emotional expression: happiness, fear) × 2 (group: out-group, in-group) × 2 (valence: positive, negative) × 2 (trial sequence: switch, repetition) repeated measures ANOVAs with mean RTs as the dependent variable is reported in Table 5. There were no significant effects beyond a main effect of emotion and an emotion × task switch interaction that are not of direct interest here (see Table 5). The main hypothesized effect, that is, the emotion × group × valence interaction, was clearly nonsignificant.

Moreover, there was no hint of a group × valence interaction (which tests the hypothesis that the groups produce an EAST effect irrespective of emotional expression): The means of the relative positivity scores (i.e. the difference scores obtained by subtracting RTs of the positive key responses from RTs of the negative key responses) were $M = 2$ ms for out-group and $M = 1$ for in-group.

In contrast, the emotion × valence interaction can be considered significant in a one-tailed test: As noted earlier, any test in a repeated measures design involving only two-condition factors is equivalent to a one-sample t-test (with $t = \sqrt{F}$) that tests the deviance of the mean of an appropriate difference variable from zero. Thus, the difference in positivity for happy faces minus positivity for fearful faces was significantly greater than zero, $t(249) = 1.85$, $p = .033$ (one-tailed). The t-test assumes normality, which was not given for the difference variable that corresponds to the emotion × valence interaction because of outliers at both tails of the distribution ($p = .004$ according to a Shapiro-Wilks test). Therefore, we conducted a robust one-sample t-test (function *yuen.t.test* from the R package *PairedData*; Champely, 2018; see Wilcox, 2013, with regard to robust testing) with a trimming of $\gamma = .2$, which yielded $t(149) = 2.33$,

TABLE 6 Results of the ANOVA for error rates (Experiment 2)

				× Task-switch		
	$F(1, 249)$	p	η_p^2	$F(1, 249)$	p	η_p^2
Emotion (Emo)	<1			<1		
Group (Gr)	<1			<1		
Valence of Keys (Val)	<1			<1		
Emo × Gr	2.37	.125	.009	<1		
Emo × Val	12.76	<.001	.049	<1		
Gr × Val	12.64	<.001	.048	3.74	.054	.015
Emo × Gr × Val	<1			<1		

Note: As noted in the main text, the inclusion of the task-switching factor was not mentioned in the preregistration. The F-values (p-values) of an analysis with aggregate variables that were created without considering task-switching were 2.04 (.155) for Emo × Group, 14.54 (<.001) for Emo × Val, 11.52 (<.001) for Group × Val, Fs < 1 for the remaining effects.

$p = .021$, $d_Z' = .16$ (see Algina et al., 2005).[9] The trimmed mean was $M_t = 4$ ms (i.e. the positivity score for happy faces was 4 ms larger than the positivity score for fearful faces).

Table 6 shows the results of an ANOVA with error rates as the dependent variable. In our preregistration, we declared reaction time a priori as the dominant dependent variable, whereas accuracy will be checked with regard to possible speed-accuracy-trade-offs. In this regard, the only significant result of interest in the RT analysis – that is, the emotion × valence interaction – was mirrored by a significant emotion × valence interaction for error rates. Since it indicated that happy faces resulted in a higher positivity score ($M = .69$, $SD = 6.06$) than fearful faces ($M = -.082$, $SD = 6.89$), this effect corroborates the RT-based effect.

The most interesting effect, namely the emotion × group × valence interaction, was also not significant in the error analysis. However, there was another significant two-way interaction: the significant group × valence effect indicated that in-group faces led to a higher positivity score in terms of error rates for the in-group ($M = .68$, $SD = 6.58$) than the out-group ($M = -.81$, $SD = 6.34$).

Finally, an identification score was calculated by aggregating the items relating to participants' identification with the in-group (Cronbach's $a = .83$). To examine whether the identification score modulated the main results, identification was correlated with difference scores corresponding to the "emotion × group × valence" and "group × valence" interactions (calculated for RTs and error rates separately). Correlations were all non-significant (r values were between .026 and .097, all p values > .125).

Discussion

Experiment 2 examined whether group membership and emotional expression interact if newly formed in- and out-groups are used. First of all, the hypothesized emotion × group × valence interaction effect was not present (the Bayes factor in favour of the null was $BF_0 = 4.26$; "substantial evidence" according to Jeffreys, 1961).

So did the EAST fail completely in Experiment 2? No, it did not. There was evidence of a significant interaction between emotion and valence. That is, the positivity score of RTs for happy expressions was significantly greater than that for fearful faces. This result was further corroborated by the analysis of error rates. In the absence of group membership processing, this interaction can be taken as a validity check of the EAST procedure.

Is there no evidence for group membership processing? The analysis of RTs (which was our dominant dependent variable) clearly shows this because the group factor was not involved in any significant

[9] This analysis was not explicitly mentioned in our preregistration and might therefore be considered "exploratory".

effect. On an exploratory note, however, we found some evidence for group processing in the error rates, that is, the group × valence interaction was significant; the positivity score of error rates was larger for the in-group than the out-group.

Why did Paulus and Wentura (2014) find converging evidence from ethnicity and minimal-group variations? We need to keep in mind the potentially important procedural difference between Experiments 1 and 2 of Paulus and Wentura (2014) and therefore between their Experiment 2 and the present one. Whereas in Experiment 1 of Paulus and Wentura (2014) as well as in the present two experiments, left/right blur was the task-relevant feature, in Experiment 2 of Paulus and Wentura (2014), the task was to classify faces by group membership, that is, one of the features (i.e. group membership) that was relevant for the hypothesis was also the task-relevant feature. It is unclear whether the same result would have emerged if left/right blur had been the task-relevant feature in the approach/avoidance paradigm. It might be—and the present experiment speaks for this assumption—that (minimal) group membership is not involuntarily processed in this case.

The main difference between ethnicity-based groups and arbitrary groups is that ethnicity tends to be immediately perceptually obvious (and is therefore potentially processed faster than the person's identity), whereas arbitrary group membership must be retrieved as a feature of the concrete person, that is, subsequent to the processing of identity.

GENERAL DISCUSSION

The results of Experiment 1 suggested that the evaluation of emotional expression was modulated by ethnic group membership (i.e. White Caucasian, Middle-Eastern). As hypothesized, the affective meaning of the face stimuli was involuntarily processed and interfered with task performance even though the meaning of the face stimuli was task-irrelevant. Extrinsically positive key presses were (relatively) faster for White Caucasian happiness and Middle-Eastern fear than extrinsically negative key presses. On the other hand, extrinsically positive key presses were (relatively) slower for White Caucasian fear and Middle-Eastern happiness than extrinsically negative key presses.

This result was in line with the pattern found in the approach/avoidance paradigm of (Paulus & Wentura, 2014) and in the evaluative priming studies by (Weisbuch & Ambady, 2008), and it is therefore also in line with the social message account. In contrast, our result does not confirm the more recent evaluative priming studies by Paulus and Wentura (2018). To recap, confirmation of those results would have involved finding two 2-way interactions, namely emotion × valence and ethnicity × valence interactions; however, both were non-significant in Experiment 1.

At the outset of our present research we had to concede that one part of a mutually corroborating pair of studies (i.e. Paulus & Wentura, 2014 and Weisbuch & Ambady, 2008) was called into question by later replication failures (Craig et al., 2014; Paulus & Wentura, 2018). Our present Experiment 1 shows that corroborating evidence for the approach/avoidance results of Paulus and Wentura (2014) can be found with an alternative indirect measure of valence connotations, that is, the EAST.

In our EAST study, the two decisive features of the face stimuli – that is, ethnicity and emotional expression – were task-irrelevant, just as they were in the approach/avoidance study and in the evaluative priming studies. However, in the approach/avoidance task as well as in the EAST, face stimuli required a direct response (i.e. the faces themselves were task-relevant), whereas in the evaluative priming paradigm, face stimuli were only presented as primes and did not require any direct response (i.e. they were entirely task-irrelevant). This could be one potential answer to why Craig et al. (2014) and Paulus and Wentura (2018) did not observe the interactive effect of group membership and emotional expression. In other words, it is conceivable that the face stimuli must be directly targeted in order for the two features of the face (group membership and emotional expression) to be integrated and for a “social meaning” to emerge. However, this is speculation and it does not resolve the discrepancy between Craig et al. (2014) and Paulus and Wentura (2018) on the one hand and Weisbuch and Ambady (2008) on the other hand (which, however, was not the aim of the present research).

To corroborate the results from the approach/avoidance study (Paulus & Wentura, 2014), we aimed to obtain converging evidence from an indirect measure of valence connotations. This attempt was successful. However, we additionally found a clear boundary restriction: The ethnicity × emotion × valence interaction found with the EAST did not generalize to newly established in- and out-groups.

This, however, should not be considered a failure to replicate the corresponding approach/avoidance experiment of Paulus and Wentura (2014; Experiment 2). Paulus and Wentura (2014) found the group × emotion interaction (with approach scores as the dependent variable) in an experiment with group as the task-relevant feature: Participants were instructed to categorize the group; thus, they had to identify the person and retrieve the learned group assignment. Given this context, the approach/avoidance associations of emotional expressions were moderated by the group (in accordance with the SMA). The present Experiment 2 suggests that this intentional processing of group membership is needed in case of arbitrary groups to find the moderation of the emotional effect.

A recent critique of the social message account

As already briefly noted in the *Introduction*, the current Experiment 1 is of importance for a further reason. Recently, Kozlik and Fischer (2020) put forward an alternative account for the ethnicity × emotion interaction found by Paulus and Wentura (2014) and Weisbuch and Ambady (2008). They argued that (a) emotional expression and ethnicity are two features of a face that are both affectively connoted and both automatically processed, and that (b) the valence of the two features can therefore be congruent or in conflict. Depending on the task, the presence or absence of conflict affects performance. For example, according to this processing conflict account, a mismatch of features might trigger avoidance reactions (i.e. White Caucasian—fear; Middle-Eastern—happiness) whereas compatible pairs trigger approach behaviour (i.e. White Caucasian—happiness; Middle-Eastern—fear). Indeed, such an explanation is compatible with the effect found by Paulus and Wentura (2014). Although Paulus and Wentura (2014) already discussed this conflict explanation and provided arguments against this alternative explanation of their results, it nevertheless remains a viable alternative (but see Wentura & Paulus, 2022, for recent counter-evidence found with the approach/avoidance task).

It should be noted that our EAST experiment was planned before the publication of Kozlik and Fischer (2020). Therefore, the present experiments did not aim to inform the debate on whether one of the accounts (the social message account or the processing conflict account) yields a better explanation for the ethnicity × emotional expression interaction observed in recent studies. However, our study can contribute to this debate because the processing conflict account makes three possible predictions for the EAST, with only the least plausible one conforming to the results we obtained:

First, if we ignore for a moment the fact that the sequence of face trials in the EAST was interspersed with evaluative decision trials, the present experiment was an almost one-to-one replication of Experiment 2 of Kozlik and Fischer (2020)[10]: In this experiment, the faces varying in emotional expression and ethnicity were presented with right- versus left-side blurring; participants' task was to categorize faces based on the side of the blurring. The authors predicted and found that conflicting stimuli (i.e. Middle-Eastern faces expressing positive emotion; White Caucasian faces expressing negative emotion) were associated with performance decrements (i.e. slower responses) compared to non-conflicting stimuli (i.e. Middle-Eastern faces with negative expression; White Caucasian faces with positive expression). Obviously, the face trials in our EAST experiment are structurally the same as the trials in Experiment 2 of Kozlik and Fischer (2020). Thus, if we simply disregard for a moment the "EAST-rationale", that is, the potential carry-over influence of the evaluation trials on face trials, one would expect the same result as Kozlik and Fischer (2020) in their Experiment 2: Collapsed across "positive" and "negative" keys, responses in conflict trials should be slower than responses on non-conflict trials; thus,

[10] Kozlik and Fischer (2020) used anger instead of fear as the negative emotion. However, their prediction is the same for anger and fear.

an emotion × ethnicity interaction should have emerged. However, we did not find evidence for an emotion × ethnicity interaction in our data (see Table 2).[11]

Second, an implicit assumption of the conflict account is that the two valent features of a face (i.e. emotional expression and ethnicity) are initially processed independently of each other (so that conflict can then arise). We know that there are conditions under which this assumption seems to hold, for example, if faces are presented as primes in an evaluative priming task (Craig et al., 2014; Paulus & Wentura, 2018). The pattern of two independent priming effects found by Paulus and Wentura (2018)—that is, one for emotional expression and one for ethnicity—meshes well with the processing conflict account of Kozlik and Fischer (2020) if we assume that the brief presentation of primes leads to a high probability of extraction of the nominal valence of emotional expression (causing a priming effect due to emotion) and a somewhat lower probability of extraction of ethnicity-related prejudice (causing a somewhat weaker priming effect due to ethnicity). However, in the EAST this result was clearly not corroborated, since we found neither an emotion × valence nor an ethnicity × valence interaction in Experiment 1.

Third, one might in principle argue that feature match versus mismatch (i.e. valence congruency vs. incongruency) determines the effective valence of the stimulus in the EAST. This would mean, for example, that the combination of two negative features—that is, member of out-group and negative expression—constitutes a positive stimulus. In this case, the prediction of the processing conflict account matches the prediction of the social message account, that is, the triple interaction that we actually found (in Experiment 1). However, we consider this option not very plausible because it would mean that the extraction of two negative features does not bias towards the negative response, whereas the "second-order valence" (i.e. the positivity derived from the two negative features) does bias towards the positive response.

Does our Experiment 2 contribute to this debate? Only in passing. First, the processing conflict account can in principle be applied to the minimal group situation, but in fact all of Kozlik and Fischer (2020) experiments focus on the ethnicity × emotion interaction. Second, our Experiment 2 shows, except for one detail, that in the minimal group EAST only the emotional expression of faces is processed (despite task-irrelevance) and involuntarily influences responses according to their nominal valence (i.e. joy as a positive stimulus, fear as a negative one). For reaction times as the dependent variable, there is no evidence for a corresponding group membership processing. Only in the error rates, an EAST effect was found, which corresponds to the presumed evaluation of the groups. We do not want to emphasize this finding too much, as it is of rather exploratory in nature. The result could possibly be due to the fact that the learning procedure resulted in an inhomogeneous pattern of access to the group status of the stimulus persons: While it ensured explicit (i.e. intentional) retrieval of group status for (almost) all stimulus persons, immediate involuntary stimulus-driven activation of group status may have been achieved only for a small subset of stimulus persons per participant, which then triggered the effect in error rates.

Boundary conditions

Most experiments discussed in this paper employed images varying in emotional expression and ethnicity. Only Experiment 2 of Paulus and Wentura (2014) and the present Experiment 2 experimentally manipulated the in-/out-group status of emotional faces, using a modified minimal group paradigm. While Paulus and Wentura (2014) found a results pattern that supports the SMA with manipulated in-/out-group status (see also Paulus et al., 2019, for evidence with the startle paradigm), the emotion × group × valence interaction effect that would have supported the SMA was not observed in the present Experiment 2. However, it should be noted that Experiment 2 of Paulus and Wentura (2014) deviated from the other experiments by making group membership the task-relevant feature. In Experiment 2, we could not make group membership a task-relevant feature without violat-

[11] Experiment 2 of Kozlik and Fischer (2020) yielded an emotion × ethnicity interaction of η_p^2 = .28. Even when conservatively halving this effect (i.e. η_p^2 = .14) for a *post-hoc* power analysis, our experiment had a power of $1-\beta$ = .99 to find it (α = .05).

ing the basic rationale of the EAST. No significant effects involving the group factor (especially the emotion × group × valence interaction) were found for RTs (see Table 5). Thus, we can conclude that the learned group membership was not involuntarily processed to a level to influence the responses in the EAST. Of course, this null result does not invalidate Experiment 1 of the present study. The main difference between ethnicity-based groups and arbitrary groups is that the ethnicity feature is perceptually processed immediately, whereas an arbitrary group membership is potentially only retrieved as a feature of the concrete person identity node (in terms of the face recognition model by Bruce & Young, 1986). Moreover, the null finding in present Experiment 2 does not invalidate the results found by Paulus and Wentura (2014) with their Experiment 2 since they made group membership the task-relevant feature. Finally, Experiment 2 does not contribute to the social message or processing conflict debate due to the absence of (almost) all group-based effects. However, since Experiment 2 has not yielded the hypothesized interaction that would have supported the SMA, it could be the aim of future studies to replicate the emotion and group membership interaction with other non-minimal groups (such as sports fans, and religious groups). However, it cannot be expected that all negatively rated out-groups would result in the emotion × group interaction effect (which supports the SMA in the case of ethnicity). For example, Wentura and Paulus (2022) conducted a study to compare two different group contrasts in the approach/avoidance task. One condition was a replication of Paulus and Wentura (2014; Experiment 1). That is, ethnicity (i.e. Middle-Eastern vs. German) defined the group contrast. The emotion × group interaction (with approach scores as the dependent variable) found in the earlier experiment was replicated. In the other condition, age (i.e. old vs. young stimulus persons) defined the group contrast. Although ageism (i.e. a prejudice against age and older people) is well documented (e.g. Degner & Wentura, 2011; Nelson, 2005), the nature of the negativity is different from that attributed to young Middle Eastern men. While the latter is seen as hostile and threatening, old people are seen as weak and worthless. Thus, the specific predictions of the SMA in the ethnicity case (e.g. a happy expression as malevolent) do not hold here. As hypothesized, the emotion × group interaction (with approach scores as the dependent variable) was not found for age as the group-defining feature (see for further discussion Wentura & Paulus, 2022).

In conclusion, the present Experiment 1 provides further support for the social message account of processing emotional faces (Weisbuch & Ambady, 2008). Most importantly, it provides evidence from a task that measures involuntarily evoked valence—that is, the EAST—in line with the initial findings by Weisbuch and Ambady (2008) using the evaluative priming paradigm. Since Weisbuch and Ambady's (2008) initial findings were not corroborated by two later direct replication attempts (Craig et al., 2014; Paulus & Wentura, 2018), our conceptual replication is of utmost importance. It indicates that explanations of the replication failures should focus on the peculiarities of the evaluative priming paradigm, especially the fact that the stimuli of interest—that is, the primes—are entirely task-irrelevant. The EAST is suitable for assessing the valence of faces *indirectly*, while at the same time requiring a *direct* response to the facial stimuli. It produced a pattern of results that is fully compatible with the social message account.

AUTHOR CONTRIBUTIONS

Emre Gurbuz: Data curation; formal analysis; investigation; software; writing – original draft. **Andrea Paulus:** Conceptualization; funding acquisition; methodology; supervision; writing – review and editing. **Dirk Wentura:** Conceptualization; funding acquisition; methodology; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

We thank Ullrich Ecker for his helpful comments on an earlier version of this article. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This research was supported by a grant from the German Research Foundation to Dirk Wentura and Andrea Paulus (WE 2284/14–2).

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

OPEN RESEARCH BADGES

This article has earned Open Data, Open Materials and Preregistered Research Design badges. Data, materials and the preregistered design and analysis plan are available at https://osf.io/pjy2f/?view_only=de0885eaa36146528d913d34629d7096; https://aspredicted.org/L87_JR2.

DATA AVAILABILITY STATEMENT

All the data, as well as the material, is openly accessible at https://osf.io/pjy2f/?view_only=de0885eaa36146528d913d34629d7096. We report all measures, manipulations, and exclusions for our studies.

REFERENCES

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). Effect sizes and their intervals: The twolevelrepeated measures case. *Educational and Psychological Measurement*, *65*, 241–258. https://doi.org/10.1177/0013164404268675

Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology*, *41*(2), 76–81. https://doi.org/10.1027/1864-9335/a000011

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Champely, S. (2018). *PairedData: Paired data analysis*. R package version 1.1.1. https://CRAN.Rproject.org/package=PairedData

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. (2001). A structural and process analysis of the implicit association test. *Journal of Experimental Social Psychology*, *37*, 443–451. https://doi.org/10.1006/jesp.2000.1464

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

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion*, *12*(1), 45–62. https://doi.org/10.1080/026999398379772

Degner, J., & Wentura, D. (2008). The extrinsic affective Simon task as an instrument for indirect assessment of prejudice. *European Journal of Social Psychology*, *38*, 1033–1043. https://doi.org/10.1002/ejsp.536

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

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. https://doi.org/10.1037/0022-3514.50.2.229

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. https://doi.org/10.1037//0022-3514.74.6.1464

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Kopp, B., Steinke, A., & Visalli, A. (2020). Cognitive flexibility and N2/P3 event-related brain potentials. *Scientific Reports*, *10*(1), 9859. https://doi.org/10.1038/s41598-020-66781-5

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

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 and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, *129*(1), 4–26. https://doi.org/10.1037/0096-3445.129.1.4

Meiran, N. (2000). Modeling cognitive control in task-switching. *Psychological Research*, *63*, 234–249. https://doi.org/10.1007/s004269900004

Nelson, T. D. (2005). Ageism: Prejudice against our feared future self. *Journal of Social Issues*, *61*(2), 207–221. https://doi.org/10.1111/j.1540-4560.2005.00402.x

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., Renn, K., & Wentura, D. (2019). One plus one is more than two: The interactive influence of group membership and emotional facial expressions on the modulation of the affective startle reflex. *Biological Psychology*, *142*(June 2018), 140–146. https://doi.org/10.1016/j.biopsycho.2018.12.009

Paulus, A., & Wentura, D. (2014). Threatening joy: Approach and avoidance reactions to emotions are influenced by the group membership of the expresser. *Cognition and 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*, 143–154. https://doi.org/10.1016/j.jesp.2018.04.004

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, *9*(3), 185–211. https://doi.org/10.2190/DUGG-P24E-52WK-6CDG

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

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*(2), 286–298. https://doi.org/10.1037/a0022582

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., & 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*, 1–14. https://doi.org/10.3389/fpsyg.2022.885668

Wilcox, R. R. (2013). *Introduction to robust estimation and hypothesis testing*. Academic Press.

How to cite this article: 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*, 1056–1075. https://doi.org/10.1111/bjso.12619

APPENDIX 1

TABLE A1 Deviations from the preregistration of experiment 2 and their reasons

Preregistration	Deviation	Reason
"… a sample of $N = 243$ is needed … We will recruit $N = 260$ participants to account for some outliers."	We finally recruited 315 participants to achieve a valid sample of $N = 250$.	It turned out that our pre-registered outlier criteria resulted in a higher outlier percentage than 6.5% (=[260–243]/260). Therefore we had to start a further round of recruitment in Prolific (based on the outlier percentage of the first round) to have at least a minimum $N = 243$. (See also next point.)
"… we will discard participants who do not complete the experiment within 90 min."	We discarded participants who did not complete the experiment within 120 min.	The 90-min criterion turned out to be too strict. It would have resulted in discarding $n = 27$ instead of $n = 9$ (120-min criterion) participants which appeared to be inappropriate, given that the 90-min criterion was chosen a bit arbitrary. We decided for this change while we were confronted during recruitment by the much larger outlier rate than initially expected (see above). Additionally excluding participants with completion time between 90 and 120 min does not change the essential results.
[not mentioned]	An additional 87 participants started the experiment but were unable to successfully complete the face learning task (see Procedure); therefore, their experimental session was terminated before the main task.	It was clear after pilot tests that the learning task was rather hard so that we should prevent participants from entering our main task who were not able (or not motivated enough) to associate individuals with groups. We missed to mention this detail in the preregistration. However, the a priori character of this criterion is evident: The experimental session was terminated if the learning criterion was not fulfilled.
[not mentioned]	Inclusion of the task-switching factor in the analyses.	We did not preregister the inclusion of the task-switching factor in the analyses, which we decided to include to stay consistent with Experiment 1 and because it is more appropriate (see Degner & Wentura, 2008). An analysis with aggregate variables that did not take task switching into account yielded essentially the same results as those reported in the Results section (see the notes in Tables 5 and 6).
"At the end of the experiment we ask participants four questions with regard to their identification with their ingroup. We will check whether identification moderates the main result."	We had in fact five question.	Initially we had four questions (and this information entered into the preregistration). At a late stage of planning, we added a fifth question but forgot to change the information in the preregistration text.