The attentional bias for happy faces in the dot-probe task is moderated by the activation of a social processing mode, if faces are masked

Dirk Wentura[1], Siddhima Gupta[1,] and Benedikt Emanuel Wirth[2]

[1]Saarland University, Saarbrücken, Germany

[2]German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany

Dirk Wentura: https://orcid.org/ 0000-0001-9907-498X

Siddhima Gupta https://orcid.org/xxxxxxxxxxxx

Benedikt Emanuel Wirth: https://orcid.org/0000-0001-8039-7316

Author Note

This research was supported by a grant from the German Research Foundation to Dirk Wentura and Benedikt Wirth (WE 2284/16-1; WI 5758/1-1). Correspondence concerning this article should be addressed to Dirk Wentura, Saarland University, Department of Psychology, Campus A2.4, D-66123 Saarbrücken. E-mail: wentura@mx.uni-saarland.de

Abstract

Previous research found that an attentional bias for angry faces, assessed with the dot-probe task, depended on the social character of the target-related task ("social processing mode hypothesis"); surprisingly, a bias for happy faces did not. Since a social processing mode might already be triggered by the clearly visible happy face cues, we tested whether masked (i.e., only marginally perceptible) happy faces would produce a result comparable to that for angry faces, i.e., a moderation by the social character of the target-related task. This hypothesis was confirmed in Experiment 1: there was no bias in the non-social condition, but there was a bias in the social condition. However, whereas men showed the expected effect of a bias towards happy faces, women showed the opposite effect (i.e., a possible bias towards neutral faces). Experiment 2 confirmed that the surprising effect for women was indeed due to masking: A standard unmasked presentation condition replicated the previously published results of a bias towards happy faces, whereas the masked condition again yielded a numerically reversed effect that was significantly different from the effect found with unmasked faces. Taken together, the basic rationale that led to the masking experiments was confirmed. The pattern found for women suggests that further research is needed to clarify the dissociation between masked and unmasked cueing.

Keywords: Attentional bias; dot-probe task; spatial attention; happy faces; angry faces

The attentional bias for happy faces in the dot-probe task is moderated by the activation of a social processing mode, if faces are masked

In a series of experiments, Wirth and Wentura (2020) found consistently an attentional bias towards happy faces, using the dot-probe task. The dot-probe task can be considered as a variant of the exogenous cueing task of basic attention research (Jonides, 1981; Posner et al., 1980): A target display contains a target stimulus that appeared at one of at least two possible locations. The target has to be categorized according to a binary feature that varies orthogonal to location (e.g., whether it is a *p* or a *q; discrimination task*). The target display is preceded by a cue display which creates a *difference* with regard to the potential target locations. In the seminal research on exogenous cueing the difference was that at one location a simple stimulus (e.g., a rectangle) was very briefly presented that is missing at the alternative location; thus, an "abrupt onset" of a stimulus marks the difference between locations.

In a typical dot-probe experiment the difference is that an emotional face is presented at one location whereas a neutral face is presented at the alternative location; thus, the emotional connotation marks the difference between locations. If the difference influences the orientation of attention, the response time to targets should be a function of location match: If the target is presented at the location of momentary attention (whether it is there because of the abrupt onset or because of the emotional connotation), it will be processed faster than if it is presented at the alternative location. This facilitation will be observed with short cue-target asynchronies (CTOA; e.g., 100 ms). At least from basic attention research it is known that long CTOAs (i.e., above 300 ms) caused a reversed effect ("inhibition of return"; Posner & Cohen, 1984; Klein, 2000).

Using a 100-ms-CTOA, Wirth and Wentura (2020) as well as Wentura et al. (2024) found evidence for a bias towards happy faces whereas Wirth and Wentura (2018, 2019, 2023) found evidence for a bias towards angry faces. Both of these findings were important because it had previously been questioned whether these biases existed in unselected samples, that is, samples that were not selected for, for example, high trait anxiety.

The bias towards angry faces, however, showed an interesting characteristic: It was dependent on what the authors called social processing mode (Wirth & Wentura, 2019, 2023). The mode variation was realized by the target screen and the target task (see Figure 1).

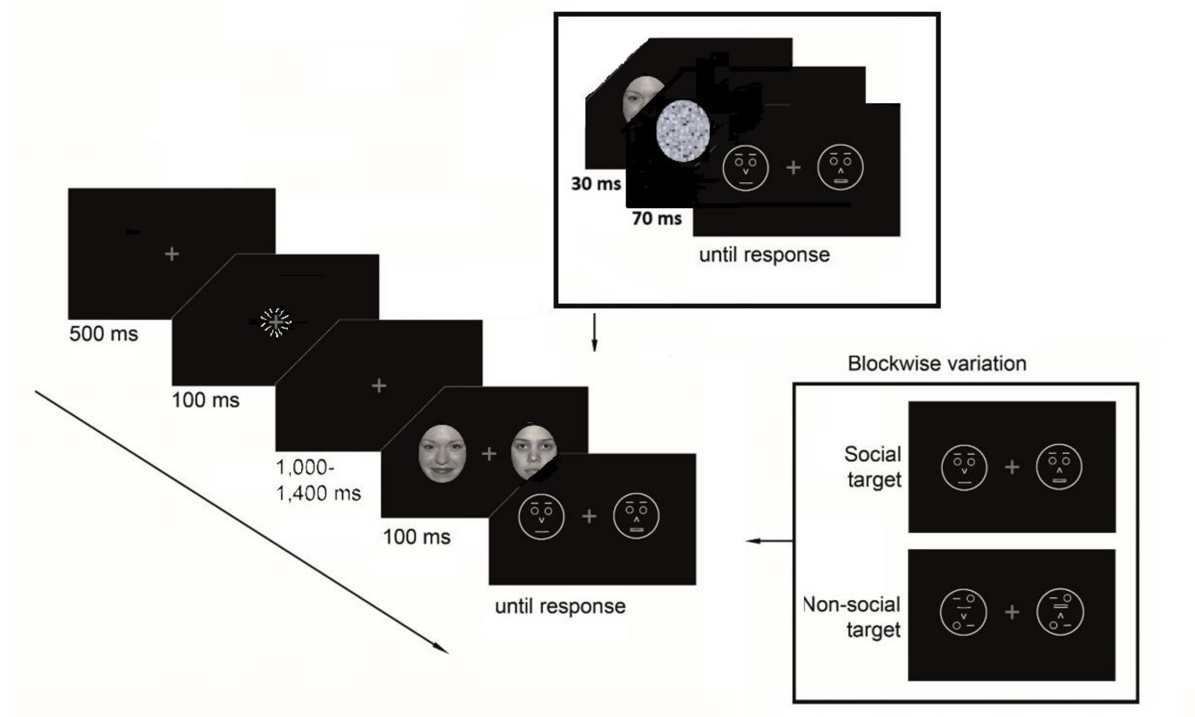

Figure 1. Schematic illustration of a typical trial of Wirth and Wentura (2019, 2020) and the present studies. Participants had to find the stimulus with the horizontal double line (here: the right one) and to categorize the direction of its nose (i.e., arrow up or down). An trial is depicted here that has the target not in the same position as the happy face. Wirth and Wentura (2020, Experiment 3) used the cue presentation of the main figure and the variation of the target (social vs. non-social; see box on the right). Wirth and Wentura (2019, Experiment 1) used angry faces instead of happy faces. The present Experiment 1 used the same target variation but the masked cue presentation as shown in the top box. The present Experiment 2 used only the social target condition but varied unmasked (i.e., the cue presentation of the main figure) versus masked cued presentation. For the sake of visibility, proportions are not true to scale.

In the social processing mode, two schematic faces were shown. The target was defined by the double-lined mouth; the nose (an arrow-like component) of the target had to be categorized as pointing up or down. In the non-social condition, scrambled versions of the schematic faces were used that had no longer a face character. The task remains to find the stimulus with the double line and to categorize the direction of the arrow. For angry faces, the attentional bias was found in the social mode but not in the asocial mode.

In contrast to the bias for angry faces, the bias for happy faces was not dependent on the target type, that is, it showed up with social targets and non-social targets (Wirth & Wentura, 2020). Of note, the bias was not found with inverted faces, that is, it cannot plausibly be attributed to perceptual characteristics.

Although face cues were presented very briefly (100 ms) in our previous studies, they were clearly noticeable. Thus, one might argue that the repeated presentation of noticeable happy faces may activate a social processing mode by itself. This might especially be the case since we used happy faces with closed mouths and thus low emotional intensity. These faces rather conveyed the impression of an affiliative smile than of an outburst of joy. (Ambadar et al., 2009, showed how subtle variations can change the perceived social message of a smile.)

Thus, the repeated presentation of happy faces might create a context that somewhat mimics a default social context in real life. This may be a difference to the experiments with angry faces. Note that several studies claim attentional avoidance of negative stimuli (e.g., Calvo & Avero, 2005; Koster et al., 2006). Due to the small time-span of a dot-probe trial, we do not believe that avoidance of angry faces can cause negative cueing effects (especially with a CTOA of 100 ms). However, avoidance might be observable on a large time scale: If targets do not require social processing, irrelevant angry face cues might effectively be ignored.

This side effect of happy faces might be prevented by masked presentation of the face cues; this is the starting point for the research presented here. Note, in their meta-analysis on dot-probe studies, Bar-Haim et al. (2007) showed that dot-probe studies with masked cues yielded similar results as dot-probe studies with unmasked cues. Moreover, Mulckhuyse and Theeuwes (2010) reviewed exogenous cueing studies of basic attention research that used masked cues. They concluded that standard results (i.e., especially facilitatory cueing) can be expected with masked cues as well. Thus, as a working hypothesis we can assume that the attentional effects of happy faces will be found in a social processing mode. In the case of masked cues, however, this mode is only triggered when social targets are used.

Thus, Experiment 1 will directly investigate the influence of social processing on attentional bias towards masked happy faces. If attentional bias towards masked happy faces is contingent on the activation of a social processing mode, we should find significant cueing scores for masked happy faces in the social target condition, but not in the non-social target condition. To anticipate, we indeed found a moderation by target type, however, not in the exact hypothesized form: As expected, there was no bias in the non-social target condition. In the social condition, the bias depended on gender of participants (not a far-fetched moderator; Campbell & Muncer, 2017). Men showed the expected bias towards happy faces, whereas women showed a reversed effect (tentatively a bias towards neutral faces). Therefore, Experiment 2 compared masked and unmasked cue presentation (with the social target type) in an all-female sample to show that masked and unmasked presentation really dissociate.

Pilot Study

For the masked dot-probe task, it was our intention to implement a combination of brief cue presentation plus masking that would prevent participants from establishing a mindset that the experiment was about processing faces, that is, social stimuli, which could easily be the case in the standard unmasked condition. It was not our intention to realize a masking that prevents any conscious processing of cue faces. That is, operationally a direct test of cue awareness should not result in a mean cue sensitivity (d') that does not significantly deviate from zero. Thus, we are not concerned with the notorious question of unconscious processing. We therefore use the term *marginally perceptible* (see also Greenwald et al., 1995; Wentura & Rohr, 2018). We piloted several versions. For estimating the degree of perceptibility of our final version, we recruited a small independent sample of participants for a direct test of cue perceptibility.

Method

Participants

A sample of 20 university students (12 women, 8 men) was recruited to assess the degree of perceptibility of the faces; age ranged from 18 to 26 ($M = 21.1$ years, $SD = 1.9$). All participants reported normal or corrected-to-normal vision and provided informed consent prior to testing. They were compensated for their participation with € 7,5.

Materials

The materials were identical to the one of Wirth and Wentura (2020). That is, as cues we used photographs of eight female and eight male individuals showing happy and neutral expressions that were taken from the NimStim set of facial expressions (Tottenham et al., 2009). Since exposed teeth are a strong perceptual confound of happy expressions that can potentially distort dot-probe effects (Wirth & Wentura, 2018b), we only employed happy faces with closed mouths. Thus, the intensity of the emotional expression is rather moderate in these faces. Using Adobe Photoshop (Adobe Systems Inc., San Jose, CA, USA), all stimuli were cropped into a standard oval shape concealing hair and external features and were converted to grayscale (see Fig. 1). Masks were created by converting four frontal views of additional faces with neutral expressions (two men and two women) into spatially quantized images (Bachmann et al., 2005; see *Appendix*).

Procedure

The procedure was identical to the one of Wirth and Wentura (2020, Experiment 3), except the masked presentation of cues and the task. The study was conducted on five PCs equipped with 17-in. CRT monitors using a resolution of 1,024 × 768 Pixels, a refresh rate of 100 Hz, and a color depth of 32 bit. The experimental routine was programmed using Psychtoolbox-3 (Kleiner, Brainard, & Pelli, 2007) for Matlab 2014a (Mathworks, Natick, MA, USA). At the beginning of the session, participants were seated in an individual testing booth, approximately 65 cm from the monitor. After completion of the consent form, they were presented with an instruction screen explaining the experimental procedure. Throughout the procedure, a gray fixation cross was presented on a black background to maintain participants' focus at the center of the screen. To indicate the beginning of a trial, the fixation cross blinked for 100 ms. The fixation cross then remained on-screen for a variable interval (chosen randomly from the set 1,000, 1,100, 1,200, 1,300, or 1,400 ms) to avoid any anticipatory effects. Subsequently, two photographic face cues, one happy and one neutral, were presented to the left and right of the fixation cross for 30 ms. The two faces were always of same gender but never of same identity. Face cues were immediately replaced by masks. The two masks were always of the same gender (see *Materials*) as the cue faces. Assignment of the two masks to locations (right/left) was random. The masks stayed on the screen for 70 ms. Each face (and mask) had a size of 4.5 × 6.2 cm (4.0 × 5.5°); the center-to-center distance between the faces was 11.1 cm (9.8°). Immediately after the offset of the cues, a white oval shape of the same size as the cue faces was presented on the left or right location until a response was given. Importantly, on half the trials, the target stimulus appeared at the location of the happy face cue and on the remaining trials it appeared at the location of the neutral face cue. Participants were informed about the presence of the cue faces; they were instructed to categorize (without time pressure) the face that was replaced by the white oval as either happy or neutral. Participants were asked to respond by pressing the "t" key for "happy," or the "v" key for "neutral," on a standard German QWERTZ keyboard. Each response was followed by a 500-ms inter-trial interval. The experiment comprised 224 trials and lasted approximately xx min. A self-paced break was included after 112 trials. At the beginning, participants were presented with 32 training trials that were not included in data analysis. These training trials used face cues of individuals that were not presented during the main trials.

Results and Discussion

Hits were defined as a "happy" response if the happy face was replaced by the white oval; false alarms were defined as a "happy" response if the neutral face was replaced by the white oval. Mean *d’* was $M = 0.67$ ($SD = 0.85$), $t(19) = 3.49$, $p = .002$, $d_Z = .78$. Thus, categorization performance was, as planned and expected, above random responding, but rather low.

Experiment 1

In order to investigate whether attentional bias towards masked happy faces is contingent on a social processing mode, we conducted Experiment 1 in close correspondence to Experiment 3 of Wirth and Wentura (2020). That is, participants performed a dot-probe task either with socially meaningful targets (schematic faces) or with meaningless targets (scrambled schematic faces; see Figure 1). The cue faces, however, were presented briefly (30 ms) and were directly overwritten by a mask (for 70 ms); thus, the cue-target asynchrony of 100 ms (used in the experiments by Wirth & Wentura, 2020) was preserved. If the social-processing-mode hypothesis is true, we would expect to find an attentional bias towards happy face cues only when participants had to classify socially meaningful targets.

Method

Participants

Eighty university students (60 women, 24 men) were compensated for their participation with € 7,5; age ranged from 18 to 29 (Md = 21.0 years). All participants reported normal or corrected-to-normal vision and provided informed consent prior to testing. The data of two further participants was excluded from all analyses, because they had an accuracy of below 60 %.

Across all experiments of our previous study (Wirth & Wentura, 2020), the attentional bias towards happy faces was of size $d_z = 0.33$. To achieve a power of $1-\beta = .80$ ($\alpha = .05$) given such an effect size, we needed to test $N = 75$ participants. (Power calculation was done with G*Power, Faul et al., 2007). We slightly oversampled to compensate for possible exclusion of participants. Note, in Experiment 1 the aim is to establish a masked attentional bias effect (in the social target condition) and to test whether the corresponding effect in the non-social target condition is absent. That is, we did no plan at this point of time for finding a significant *cue type* × *target type (social vs. non-social)* interaction.[1]

Design

We employed a 2 (*target type:* social target vs. non-social target) × 2 (*cue type: happy cue vs. neutral cue precedes the target*) design with *cue type* as a trial-by-trial within-subjects factor and *target type* as a block-wise within-subjects factor, counterbalanced for sequence.

Materials and Procedure

The materials were the same as in the *Pilot Study* and therefore identical to the one of Wirth and Wentura (2020). The procedure was identical to the one of Wirth and Wentura (2020, Experiment 3), except the masked presentation of cues. Therefore, it was identical to the *Procedure* of the *Pilot Study*, except the target screen and the task instructions.

Immediately after the offset of the cues, the target display was presented until a response was given. Thus the cue target onset asynchrony (CTOA) was 100 ms as in Wirth and Wentura (2020). The factor *target type* comprised the factor levels social target and non-social target. Figure 1 illustrates this experimental variation. In the *social target* condition, two schematic faces (with neutral expressions) were presented during the target display, one *target* face with an open mouth (symbolized by a horizontal double line) and a schematic *distractor* face with a closed mouth (symbolized by a single horizontal line). Participants had to indicate the direction of the nose of the target face while ignoring the distractor face.[2] Thus

[1] If there is an effect in the social condition and a null effect in the non-social condition, the contrast in the attentional bias between social and non-social target can be estimated as $(0.33 - 0.00)/\sqrt{2} = .23$ (if we assume equal variances of the cueing scores for the two target conditions and a null correlation between them). To achieve a power of $1-\beta = .80$ ($\alpha = .05$) with such an effect size, we would have to test N = 151 participants, which is too great an investment of resources as long as we do not know whether a masked attention bias effect can be found in the social target condition at all.

[2] Note that the direction of the distractor's nose varied orthogonally to the direction of the target's nose, that is, all combinations of directions (up/up, up/down, down/up, down/down) were equally probable.

the target had to be selected based on a socially relevant dimension (i.e., an open mouth) in this condition before it could be classified.

For the *non-social target* condition, scrambled schematic faces were presented. These scrambled faces comprised the same basic features as the schematic faces, but the spatial configuration of those features was altered (i.e., the mouth was located above the nose, one eye and one eyebrow were located above the mouth). Thus, the scrambled schematic faces conveyed the impression of a complex, meaningless pattern inside a circle. Participants' task was to find the target pattern (i.e., the one with the horizontal double line) and indicate whether the arrow in this pattern (corresponding to the nose in the social target condition) was pointing up or down. Moreover, participants were told to ignore the arrow in the distractor pattern.

The schematic faces had a size of 2.8 × 2.8 cm (2.5 × 2.5°) and the center-to-center distance between them was 11.1 cm (9.8°). Participants were asked to respond as fast as possible by pressing the "t" key for "up," or the "v" key for "down," on a standard German QWERTZ keyboard. Importantly, on half the trials, the target stimulus appeared at the location of the happy face cue and on the remaining trials it appeared at the location of the neutral face cue. Each response was followed by a 500-ms inter-trial interval. If participants made an error or took longer than 1,500 ms to submit a response, they received a 1,000-Hz warning tone of 500-ms duration via headphones. The experiment comprised 448 trials and lasted approximately 35 min. Trials were presented in two blocks consisting of 224 trials each – one with social targets and one with non-social targets, in a counterbalanced order. Within each block, a self-paced break was included after 112 trials. At the beginning of each block, participants were presented with 32 training trials that were not included in data analysis. These training trials used face cues of individuals that were not presented during the main trials.

Results

Average classification accuracy was *M* = 94.5 % (*SD* = 6.0). For the RT analysis, RTs below 150 ms were excluded, as were RTs more than 1.5 interquartile ranges above the third quartile of the individual participant's distribution (Tukey, 1977). This led to the exclusion of 2.3 % of all trials with correct responses. Table 1 shows average RTs as a function of the experimental factors.

Table 1

Mean RTs and Cueing Scores (in ms; SE in brackets) of Experiment 1 as a Function of Target Type, and Cue Type.

		Cue Type			
		Happy Face	Neutral Face		
Sample	Target Type	… preceded the Target		Cueing-Score	
Social Target	Full Sample	724 (94.8)	720 (94.8)	-4	[2]
	Women	739 (94.4)	730 (94.3)	-7	[3]
	Men	686 (95.7)	694 (96.2)	7	[4]
Non-Soc. Target	Full Sample	719 (93.9)	720 (94.4)	0	[3]
	Women	732 (93.3)	733 (93.8)	1	[3]
	Men	687 (95.3)	685 (95.9)	-2	[5]

Note. Women: n = 60; men: n = 24; accuracy rates (in %) are given in parentheses, standard errors are given in brackets, cueing score = $RT_{Neutral} – RT_{Happy}$. Deviations between the differences of mean RTs and the cueing scores are due to rounding.

A priori planned analyses

As can be seen in Table 1 ("Full sample"), surprisingly the social mode revealed a tendency towards a reversed effect, which is, however, not significant, $t(83) = 1.76$, $p = .082$, $d_Z = 0.19$. As expected, the non-social mode yielded a clear null effect, $t(83) = 0.87$, $p = .931$, $d_Z = 0.01$. (For the sake of completeness: The test for the difference yielded $t(83) = 1.18$, $p =$

.240, $d_Z = 0.13$; note again, that – as described in the *Participants* section – our experiment was a priori not powered for this comparison.)

Exploratory analyses

We routinely check for effects of participant gender (see, e.g., Campbell & Muncer, 2017; Stevens & Hamann, 2012). In the present case the result is noteworthy (see Table 1). Here, a 2 (*participant gender*) × 2 (*target type*) × 2 (*cue type*) mixed ANOVA with (correct) RTs as the dependent variable yielded a significant triple interaction, $F(1, 82) = 5.46$, $p = .022$, $\eta_p^2 = .062$, all other effects were non-significant, with Fs < 1 except *gender,* $F(1, 82) = 3.27$, $p = .074$, $\eta_p^2 = .038$ and *Cue type* × *Gender,* $F(1, 82) = 3.41$, $p = .068$, $\eta_p^2 = .040$.[3]

For the social target condition, a follow-up 2 (*participant gender*) × 2 (*cue type*) mixed ANOVA yielded a significant interaction, $F(1, 82) = 10.36$, $p = .002$, $\eta_p^2 = .112$; $F < 1$ for *cue type* and $F(1, 82) = 2.94$, $p = .090$, $\eta_p^2 = .035$ for *gender*. The numerically negative effect of the overall analysis is completely driven by the female sample (n=60), $t(59) = 3.25$, $p = .002$, $d_Z = 0.42$. Men (n=24) showed a positive effect that is associated with $t(23) = 1.79$, $p = .087$, $d_Z = 0.37$. Given the expectation of a positive effect on the basis of Wirth and Wentura (2020), one can consider this effect to be a significant replication with $p = .043$ (one-tailed).

For the non-social target condition, a follow-up 2 (*participant gender*) × 2 (*cue type*) mixed ANOVA, effects with regard to the cue type factor were non-existent, Fs < 1 for the cue type main effect and the interaction; $F(1, 82) = 2.98$, $p = .088$, $\eta_p^2 = .035$ for gender.

Discussion

As expected, there was no attentional bias in the *non-social* target condition. Surprisingly, the attentional bias in the *social* target condition tended to be of negative sign (although, admittedly, it was not significant for the full sample). However, the exploratory

---

[3] Additionally introducing *cue* gender as a further factor did not yield significant results involving this factor.

finding of a moderation by participants gender is noteworthy for three reasons. First, in detail the moderation only concerns the social mode; the non-social mode yields a null result for the full sample and for both genders. Thus, we found evidence for a dissociation between social and non-social mode, which was a central assumption of our study. Second, for one gender (men) we found evidence for our full hypothesis, that is, a cueing effect that signals attentional capture by happy faces which is restricted to the social target mode. Third, for the social mode we found for women a significant *negatively* signed cueing effect. Tentatively, it seems as if here the neutral faces were an attention attractor. This is surprising given that we consistently found a positive cueing effect for happy faces in our previous work (Wirth & Wentura, 2020). Fourth, the moderation by gender suggests that there are individual differences in the social cueing effect that are more than gradual, that is, the effect does not simply range from zero to some positive value, but from negative to positive values.

What is important now is to secure that the negatively signed effect for women is really an outcome of the masked presentation of the cue faces. That is, we have to replicate the masked social condition with a female sample, followed by the standard unmasked presentation to see whether we really find a dissociation between masked and unmasked presentation conditions.

Experiment 2

In the Experiment 2, we focus on a potential dissociation between masked and unmasked presentation conditions. Thus, firstly we aim to replicate the attentional bias towards unmasked happy face cues that was found in the Experiments of Wirth and Wentura (2020). Secondly, we aim to find evidence for a dissociation of masked and unmasked presentation conditions. Therefore, we will constrain recruitment to women because women showed the surprising negatively signed cueing effect. Experiment 2 was preregistered (https://aspredicted.org/vj4j-gw8w.pdf)

Method

Participants

Fifty-six women (students) participated in Experiment; age ranged from 18 to 29 (Md = 21.0 years). They received a compensation of € 7,5. The data of six further participants were excluded from all further analyses because their error rates were more than three interquartile ranges above the third quartile of the distribution of all participants (Tukey, 1977; see *Preregistration*). Two further participants participated; however, because of a technical failures the data were incomplete.

We preregistered the experiment with the following power planning: "Across all experiments of our previous unmasked studies (Wirth & Wentura, 2020), the attentional bias towards happy faces was of size $d_Z = 0.33$. The negative effect in the social condition of the present Experiment 1 (females) was $d_Z = 0.32$. The contrast between mask and unmasked can then be estimated as $.65/\sqrt{2} = .46$ (if we assume equal variances of the cueing scores for the two conditions and a null correlation between them). We focus on this dissociation as our main goal. According to G*Power (Faul et al., 2007), to detect an effect of size $d_Z = 0.46$ given an error of $\alpha = .05$ (one-tailed) with a power of $1\text{-}\beta = .95$, we need a sample size of 53 participants. As a secondary goal, to replicate the negative effect of size $d_Z = 0.32$ with a power of $1\text{-}\beta = .80$, we need a slightly larger sample size of 62 participants. We will take this as our planned sample size." Note, in fact the negative effect in the social condition of the present Experiment 1 (females) was $d_Z = 0.42$ (see *Results* section of the Experiment 1); the smaller value in the preregistration was based on an initial analysis that later had to be corrected. Of course, it is recommended to reduce a single effect size for the planning of replications; thus, in general we kept to our pre-registered planning, especially with regard to our main goal (which needed N=53 participants, see above). We slightly deviated from the preregistration by not recruiting further participants to replace those (n=6) who had to be discarded according to the preregistered rules. With N=56, we were still able to find a negatively signed effect of $d_Z = 0.34$ with power $1\text{-}\beta = .80$ ($\alpha = .05$, one-tailed).

Design

We employed a 2 (*cue emotion:* happy vs. angry) × 2 (*cue type: happy cue vs. neutral cue precedes the target*) design with both *cue emotion* and *cue type* as trial-by-trial within-subjects factors.

Materials and Procedure

The procedure of Experiment 2 was identical to the procedure of Experiment 1 apart from the following exceptions. First, *target type* was no longer varied; we only used the social target type. Second, the masked presentation condition was followed by a standard unmasked condition, as introduced by Wirth and Wentura (2020). That is, the presentation of a mask was omitted and the cue display was presented for 100 ms. We decided that the unmasked block always followed the masked block. The masked block should be processed without explicit knowledge of the structure of the cue presentation parameters (e.g., that there are always a happy and a neutral face). The risk that the unmasked block will yield a biased result due to being always last was seen as rather low. The experiments of Wirth and Wentura (2018, 2019, 2020, 2023) always followed a balanced design, that is, the within-participants variation of social target vs. non-social target was counter-balanced with regard to sequence. There were no indications that the second block had a less or more pronounced effect than the first one.

Results

Average classification accuracy was $M = 94.7$ % ($SD = 4.9$). For the response time (RT) analysis, the same exclusion criteria were used as in Experiment 1. This led to the exclusion of 2.8 % of all trials with correct responses. Table 2 shows average RTs as a function of the experimental factors.

A priori planned analyses

For the standard unmasked condition, we replicated the cueing effect for happy faces found by Wirth and Wentura (2020), $t(55) = 2.71$, $p = .005$ (one-tailed), $d_Z = .36$. For the masked condition, we again found a numerically negative cueing effect; however, it was very small and failed the criterion of significance, $t(55) = -0.35$, $p = .365$ (one-tailed), $d_Z = .05$. Most important, as hypothesized, the two effects significantly differ, $t(55) = 1.99$, $p = .026$ (one-tailed), $d_Z = .27$.

Table 2

Mean RTs and Cueing Scores (in ms; SE in brackets) of Experiment 2 as a Function of Presentation Condition and Cue Type.

	Cue Type		
	Happy Face	Neutral Face	
Presentation Condition	… preceded the Target		Cueing-Score
Unmasked presentation	746 (96.1)	755 (95.4)	9 [3]
Masked presentation	763 (93.8)	761 (93.7)	-1 [3]

Note. Accuracy rates (in %) are given in parentheses, standard errors are given in brackets, cueing score = $RT_{Neutral} - RT_{Happy}$. Deviations between the differences of mean RTs and the cueing scores are due to rounding.

Exploratory analyses

We additionally conducted a cross-experiments analysis comparing the masked data of the female sample of Experiment 1 with the masked data of Experiment 2. A 2 (*experiment:* Experiment 1 [women] vs. Experiment 2) × 2 (*cue type: happy cue vs. neutral cue precedes the target*) mixed ANOVA with (correct) RTs as the dependent variable yielded a significant cueing effect of $M = -5$ ms ($SE = 2$ ms), $F(1, 114) = 5.35$, $p = .023$, $\eta_p^2 = .045$, but no significant moderation by experiment, $F(1, 114) = 3.12$, $p = .080$, $\eta_p^2 = .027$; $F(1, 114) = 1.68$, $p = .198$, $\eta_p^2 = .014$ for the main effect of experiment.

Discussion

First of all, we replicated again the focal result of Wirth and Wentura (2020), that is, an attentional bias for happy faces in the standard unmasked presentation condition. This is important insofar the negatively signed effect found in Experiment 1 (for women) should not be considered a failure to replicate our earlier results.

Second, this effect significantly deviates from the numerically negative effect in the masked condition.[4] Third, masked and unmasked cueing scores correlate negatively. Thus, we should proceed from the assumption that masked and unmasked conditions dissociate with regard to the cueing effect.

The negatively signed effect in the masked condition that was found in Experiment 1 (for women) did not significantly replicate in Experiment 2. However, a cross-experiments analysis (a kind of mini meta-analysis) still indicated the negative effect.

General Discussion

With our experiments, we wanted to shed light on the following puzzle: Wirth and Wentura (2019, 2023) found a bias towards angry faces that was moderated by a social processing mode. That is, if the target task required to categorize a characteristic of a face (i.e., a social object) the bias was observed; if the target task required to categorize a characteristic of a non-social object the bias vanished. Corresponding experiments with happy faces, however, always yielded a bias towards the emotional face (Wirth & Wentura, 2020); that is, all target conditions produced an attentional bias towards happy faces. We assumed that the repeated, clearly visible presentation of friendly faces as cues evoked a social processing mode, independent of the target task. If so, a follow-up hypothesis is that the masking of the faces may interfere with this process, so that the attentional bias will now only appear in the social target condition.

[4] At first sight, it might appear trivial that a mean (i.e., the mean of the unmasked cueing) that deviates significantly from zero will also deviate from -1 (i.e., the mean of the masked cueing). That is not true because the standard error that enters into the calculation of the t-value for the first test is based solely on the variance of the unmasked cueing score whereas the standard error that enters into the calculation of the t-value for the second test is based on the variance of the difference variable, which is $\text{Variance}_{\text{unmasked}} + \text{Variance}_{\text{masked}} - 2 \times$ Covariance and therefore especially large in case of a negative correlation.

The results suggest that there may be some truth to this assumption: We found cueing effects in the social target condition whereas there were clear null effects in the non-social target condition. However, it is also part of the truth that the cueing effects in the social target condition revealed themselves only in an exploratory post-hoc analysis: Participant gender moderated the sign of the cueing effects. For men, we found the hypothesized result in all relevant aspects: the clear null effect for the non-social condition is contrasted by a positive cueing effect (that is, presumably an effect of attentional capture by happy faces) for the social condition. For women, we found a null effect for the non-social condition as well; however, the cueing effect in the social condition was reversed in comparison to men. Tentatively, it seems as if for women the neutral faces were an attention attractor. (We elaborate on this below.)

With Experiment 2, we were able to ensure that our previous results (Wirth & Wentura, 2020) – that is, a positive cueing effect for unmasked cues – also apply to an all-female sample and that a dissociation of masked and unmasked cueing can be found (in women).

How can we explain the reversed effect? We already tentatively assumed that neutral faces were an attention attractor in this case. In principle, however, we can focus on either the happy faces or the neutral faces as the trigger of the reversed effect. We will discuss both possibilities.

Happy faces as the cause of masked and unmasked effects

Of course, we started our earlier research as well as the present experiments by the assumption that the happy faces define the difference in the cue screen that causes attentional effects. However, in the present case to focus on the happy faces is to assume a truly inverse effect, that is, that attention is potentially drawn *away* from the happy faces, not toward them, when they are masked. In basic attention research, such reversed effects are known as *inhibition of return* (Posner & Cohen, 1984; Klein, 2000). However, these effects are only observed for longer CTOAs (i.e., above 300 ms, Samuel & Kat, 2003); for a CTOA of 100 ms they are rather implausible.

However, there is one study (Cooper & Langton, 2006) that already reported a negatively signed cueing effect for happy faces in a dot-probe task with 100 ms CTOA (i.e., conditions that resemble our conditions, except that they did not use masked presentation) and the authors argued in favor of a genuinely reversed effect.[5] They did so on the basis of a comparison with a baseline condition consisting of neutral/neutral face pairs as cues. Indeed, targets following neutral/neutral face pairs produced roughly the same mean response time than targets that followed the neutral face in happy/neutral cue pairs, whereas targets that followed the happy face caused a slower response. Thus, the authors interpret the finding as a kind of inhibition caused by the happy face, however, without relating it to the *inhibition of return* literature (see above). Somewhat unsatisfactorily, they left open how this effect fits into the broader literature on cueing effects.

Neutral faces as the cause of masked effects

Cooper and Langton (2006) already discussed the possibility that the reversed effect might alternatively reflect attentional capture for neutral faces, supposedly because neutral faces are in fact processed as somewhat negative or at least ambiguous, especially in the context of happy faces. The authors finally rejected this explanation because of the results for neutral/neutral control pairs (see above). However, the validity of the neutral/neutral cue pairs as a baseline should not be regarded as unquestionable. To give a potential argument: Cooper and Langton (2006) always presented face cue pairs with faces of the same identity. Thus, the neutral/neutral pairs for the assessment of the baseline comprised of two identical images whereas happy/neutral pairs comprised of two different images. Potentially, non-specific

[5] Interestingly, the participants were primarily female (82% in the whole experiment that, however, had CTOA as a between-participants variable; thus, we do not know the exact percentage for the 100 ms CTOA condition).

filtering costs (Kahneman et al., 1983) are lower for the more redundant neutral/neutral pairs. This would clearly alter the cost/benefit analyses. Therefore, it remains a viable perspective to focus on attentional capture qualities of neutral faces.

Devue and Grimshaw (2017) reported results from an oculomotor capture experiment. That is, participants had to saccade towards a target in a visual search display; erroneous first saccades towards a distractor stimulus are taken as an indicator of attentional capture processes. Interestingly, neutral distractor faces were associated with a larger probability of erroneous first saccades than angry faces. However, the difference was also obtained with inverted faces, thereby indicating that low level visual features played a role in this context.

This rather sobering interpretation can of course in principle be applied to the present negative effect as well. This would mean that the standard moderate positive cueing effect (Wirth & Wentura, 2020) is the sum of a small negative effect (due to perceptual characteristics of the neutral faces) and a larger positive effect (due to capture by happy faces). For women, masking prevents the positive part (for whatever cause). However, an attentional effect based on perceptual confounds should not be affected by the variation of the targets.

Recently, we (Wentura & Gupta, 2025) conducted an experiment that suggest a further explanation. We using the *Attentional Response to Distal versus Proximal Emotional Information* (ARDPEI) task (Grafton & MacLeod, 2014), a variant of the dot-probe task that allows to disentangle engagement processes (i.e., capture) from disengagement processes, separately for types of faces (in that case: angry versus neutral faces). Unexpectedly, we found a result that indicates a problem to disengage from neutral faces compared to – in case of that experiment – angry faces.

If we assume that dwelling on neutral faces is present in the dot-probe experiments whenever attention is accidentally on the neutral face, we can reconstruct our actual results as follows. In the unmasked condition there is a capture bias for happy faces, meaning the probability of attention being captured by a happy face is greater than the probability of attention being captured by a neutral face. Each trial with attention on the angry face contributes to a positive dot-probe effect; each trial with attention on the neutral face runs counter to it. Hence, the surplus is decisive to find a positive dot-probe effect. However, an attended face contributes to the dot-probe effect in two ways: targets that appear at the attended location can be quickly processed, but attention must move to the opposite location if the target appears there. The latter process now depends on how easily attention can disengage from the attended face. If attention dwells longer on neutral faces than on happy faces, the proportion of trials with attention on the neutral face may dampen the positive dot-probe effect, which is primarily due to the capture bias for happy faces. In the present case, for women the masked presentation seem to significantly dampen the capture bias for happy faces and then the dwelling on neutral faces comes more to the foreground. It is worth to explore this possibility in more detail.

Conclusion

We were able to solve the puzzle that attentional biases to angry faces depend on a social processing mode (Wirth & Wentura, 2019, 2023) whereas attentional biases to happy faces do not (Wirth & Wentura, 2020). Using masked cues in a dot-probe task, we found corroborating evidence for the hypothesis that variation in social context, as introduced by Wirth and Wentura (2019), also influences attentional processes with happy/neutral face cues. Unexpected was the finding of an interaction by gender. For women, a dissociation of masked and unmasked cueing effects was found, which deserves further investigation. Finally, we provided a replication of the attentional bias to happy faces found by Wirth and Wentura (2020).

Open practices statement

The data and the program code for all experiments are available on the Open Science Framework (OSF). These files can be accessed via the following link: https://osf.io/xxxx

Experiment 2 of the present study was preregistered at aspredicted.org. The documentations of these preregistrations can be accessed via the following link: https://aspredicted.org/vj4j-gw8w.pdf

The stimulus materials of the present study were taken from the NimStim set of facial expressions (Tottenham et al., 2009). Therefore, due to copyright issues, the materials cannot be made publicly available.

Acknowledgments

Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

References

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, *33*, 17-34.

Bachmann, T., Luiga, l., & Põder, E. (2005). Variations in backward masking with different masking stimuli: II. The effects of spatially quantised masks in the light of local contour interaction, interchannel inhibition, perceptual retouch, and substitution theories. *Perception*, *34*(2), 139-154. https://doi.org/http://dx.doi.org/10.1068/p5344b

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1-24. https://doi.org/10.1037/0033-2909.133.1.1

Calvo, M. G., & Avero, P. (2005). Time course of attentional bias to emotional scenes in anxiety: Gaze direction and duration. *Cognition & Emotion*, *19*(3), 433–451. https://doi.org/10.1080/02699930441000157

Campbell, A., & Muncer, S. (2017). Sex difference in awareness of threat: A meta-analysis of sex differences in attentional orienting in the dot probe task. *Personality and Individual Differences*, *119*, 181-184. https://doi.org/https://doi.org/10.1016/j.paid.2017.07.014

Cooper, R. M., & Langton, S. R. H. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour Research and Therapy*, *44*(9), 1321-1329. https://doi.org/10.1016/j.brat.2005.10.004

Devue, C., & Grimshaw, G. M. (2017). Faces are special, but facial expressions aren't: Insights from an oculomotor capture paradigm. *Attention, Perception, & Psychophysics*, *79*, 1438-1452.

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

Grafton, B., & MacLeod, C. (2014). Enhanced probing of attentional bias: The independence of anxiety-linked selectivity in attentional engagement with and disengagement from negative information. *Cognition and Emotion*, *28*(7), 1287-1302. https://doi.org/10.1080/02699931.2014.881326

Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible ("subliminal") stimuli: Dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology: General*, *124*(1), 22-42. https://doi.org/http://dx.doi.org/10.1037/0096-3445.124.1.22

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Erlbaum.

Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(4), 510-522. https://doi.org/10.1037/0096-1523.9.4.510

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*, 138-147.

Koster, E. H. W., Crombez, G., Verschuere, B., van Damme, S., & Wiersema, J. R. (2006). Components of attentional bias to threat in high trait anxiety: Facilitated engagement, impaired disengagement, and attentional avoidance. *Behaviour Research and Therapy*, *44*(12), 1757–1771. https://doi.org/10.1016/j.brat.2005.12.011

Mulckhuyse, M., & Theeuwes, J. (2010). Unconscious attentional orienting to exogenous cues: A review of the literature [Review]. *Acta Psychologica*, *134*(3), 299-309. https://doi.org/10.1016/j.actpsy.2010.03.002

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531-556). Erlbaum.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160-174.

Samuel, A. G., & Kat, D. (2003). Inhibition of return: A graphical meta-analysis of its time course and an empirical test of its temporal and spatial properties. *Psychonomic Bulletin & Review*, *10*, 897-906.

Stevens, J. S., & Hamann, S. (2012). Sex differences in brain activation to emotional stimuli: a meta-analysis of neuroimaging studies. *Neuropsychologia*, *50*(7), 1578-1593.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242-249. https://doi.org/10.1016/j.psychres.2008.05.006

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

Wentura, D., & Gupta, S. (2025). The Attentional Response to Distal versus Proximal Emotional Information (ARDPEI) task and the social processing mode hypothesis. Evidence for a moderation of disengagement processes. *Manuscript (draft)*.

Wentura, D., Messeh, L., & Wirth, B. E. (2024). Attentional bias towards happy faces in the dot-probe paradigm: it depends on which task is used. *Cognition & Emotion*, *38*(2), 217-231. https://doi.org/10.1080/02699931.2023.2283014

Wentura, D., & Rohr, M. (2018). Emotion-specific priming effects with marginally perceptible facial expression primes: Evidence from the “leave-one-out” paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(12), 1946-1969. https://doi.org/http://dx.doi.org/10.1037/xhp0000581

Wirth, B. E., & Wentura, D. (2018). Attentional bias to threat in the general population is contingent on target competition, not on attentional control settings. *Quarterly Journal of Experimental Psychology*, *71*(4), 975-988.

Wirth, B. E., & Wentura, D. (2019). Attentional bias towards angry faces is moderated by the activation of a social processing mode in the general population [Article]. *Cognition & Emotion*, *33*(7), 1317-1329. https://doi.org/10.1080/02699931.2018.1561423

Wirth, B. E., & Wentura, D. (2020). It occurs after all: Attentional bias towards happy faces in the dot-probe task. *Attention Perception & Psychophysics*, *82*(5), 2463-2481. https://doi.org/10.3758/s13414-020-02017-y

Wirth, B. E., & Wentura, D. (2023). Social processing modulates the initial allocation of attention towards angry faces: Evidence from the N2pc component. *Social Cognitive and Affective Neuroscience*, *18*(1). https://doi.org/10.1093/scan/nsad070

Appendix

The four masks used in Experiments 1 and 2