What’s in a Name, When There is an Emotional Face: Interference From Emotional Face Distractors at High Perceptual Load

Siddhima Gupta[1] and Dirk Wentura[1]

[1]Saarland University, Saarbrücken, Germany

Author Note

Correspondence concerning this article should be addressed to Siddhima Gupta, Saarland University, Faculty of Behavioral Sciences, Department of Psychology, Campus A2 4, D-66123 Saarbrücken. Email: siddhima.gupta@uni-saarland.de

Abstract

We examined distractor effects caused by neutral versus emotional faces in a perceptual load paradigm. In a series of experiments, we asked participants to categorize female or male names (the target) into their respective genders. The target was presented with 1, 3, 5, or 7 pseudo-names (i.e., varying perceptual load) in the center of the screen, with an irrelevant face distractor (to the right or left of center). At low load, we expected flanker effects (i.e., faster responses if face gender and name gender were congruent compared with the incongruent condition) that were expected to vanish at high load, in line with the perceptual load literature. For emotional faces, however, flanker effects were expected to be present at all load levels. In Experiment 1, we presented happy versus neutral faces. In Experiment 2, we presented angry versus neutral faces. In Experiment 3, we replicated both earlier experiments, varying type of emotion in a between-participants design. All experiments show an emotional flanker effect advantage, meaning that at a load level where neutral flanker effect ceases, emotional flanker effects persist. Finally, the analysis of the full data set supported our hypothesis that flanker effects for emotional faces but not for neutral faces were found even at the highest load level. We discuss the results given the prevailing theories to explain perceptual load effects, with an emphasis on attentional slippage theory.

Keywords: perceptual load, emotional faces, flanker effects

Public significance statement

The results of this study are significant because they expand our knowledge of emotional face perception. Specifically, it was shown that angry and happy faces (compared to neutral, faces), though completely task-irrelevant, are involuntarily processed even in complex perceptual environments.

What's in a Name When There is an Emotional Face:

Interference from Emotional Face Distractors at High Perceptual Load

Attentional research has repeatedly contended with the question: To what extent do humans process stimuli that are not the focus of our attention? As per early selection accounts of attention (Broadbent, 1958; Moran & Desimone, 1985; Treisman, 1969), there is an initial processing of physical characteristics of all stimuli, and an advanced distinction is made between attended and unattended information. Early selection argues that attentional resources are "limited" and it calls upon selective attention to proceed. Ultimately, the unattended information is not fully perceived. Deutsch and Deutsch's (1963) opposing theory of late selection proposed an initial involuntary semantic parallel processing of all stimuli. This parallel processing led to equivalent and early processing of physical features and identities of all stimuli. The selection of information was then performed at a later stage.

Kahneman and Treisman (1984) attributed the shift from early to late selection models of attention to the use of different paradigms. Early selection theories often used filtering paradigms, such as dichotic listening, which loaded participants with large amounts of task-relevant and task-irrelevant information (Pashler & Johnston, 1998). On the other hand, paradigms such as spatial and semantic priming found support for late selection theories, and they tended to use single or a few irrelevant stimuli with the relevant/target stimulus (Jonides, 1981; Posner, 1980; Shiffrin & Schneider, 1977).

Based on this distinction, Lavie (1995; see also Lavie & Tsal, 1994) proposed the theory of perceptual load, in which visual perception is conceptualized as a limited capacity resource. Thus, "simple" visual displays fall within a "limited capacity" and are processed completely, as in late selection models. With more complex visual displays, the capacity "runs out" and not all the information is processed, which was more like early selection models. For example, in a flanker task (Eriksen & Eriksen, 1974) in which distinguishable target and distractor stimuli are presented in a context of homogeneous filler stimuli (*low perceptual load*), the distractor is processed, that is, target responses are slower if the distractor category does not match the response when compared with a matching distractor. However, if the target and distractor are presented in a context of heterogeneous stimuli that share features with both (*high perceptual load*), the distractor is unlikely to be processed, and no congruency effect will be found (Lavie, 1995). Thus, according to load theory, early selection happens at higher levels of load, whereas late selection happens at lower load levels. Thus, perceptual load theory can be seen as a reconciliation between early and late theories of selective attention.

Perceptual load effects are shown in a number of paradigms and scenarios, including letter discrimination paradigms (Jenkins et al., 2005; Lavie, 1995; Lavie & Cox, 1997; Lavie & Fox, 2000), flicker- or motion-based inattentional blindness (Carmel et al., 2007; Hay et al., 2006; Macdonald & Lavie, 2008; Simons & Chabris, 1999), and paradigms that use irrelevant distractors, such as recognizable faces, meaningful objects, cartoons (Forster & Lavie, 2008; Lavie et al., 2009; Pinsk et al., 2004), and places (Yi, et al., 2004). In all these experiments, when the distractors are presented at lower load levels, they are more likely to be processed and create relevant distractor effects (i.e., if the distractor is incongruent to the response needed for the target, responses are slower compared to the congruent case). At higher levels of load, they are less likely to be processed and have barely any bottom-up impact on the task.

Human faces and perceptual load

Lavie et al. (2003) demonstrated one exception to this rule: the case of (famous) faces. To do so, they used Young et al.' face-name response competition task, with varying loads (Young et al., 1986). They investigated the degree of interference from peripherally presented famous faces (e.g., Bill Clinton), while presenting a famous name (e.g., Michael Jackson) in the center of the screen, embedded in a list along with 1, 3, 5, or 7 random letter "names" (e.g., Csiprmy Qhpirt), thereby manipulating perceptual load. The faces and names that were shown referred to either politicians or pop stars, and participants were asked to categorize the words accordingly. The distractor face could be either congruent (match the name; i.e., match the response category) or incongruent (belong to the competing response category) with the target. Faces caused a congruency effect at all levels of load. As stated above, this was an exception: Replacing the faces and names of famous politicians and pop stars with images and words referring to musical instruments and fruits, respectively, and the instructions to categorize the words accordingly yielded congruency effects at the lower levels of perceptual load, but not at the highest, thereby again supporting the theory of perceptual load. Lavie et al. claimed that, due to the biological and evolutionary role that faces play, their processing was considered an exception to the rule of perceptual load. However, a specific detail of their paradigm was that they used identity-priming in the congruent condition of their celebrity name categorization task (e.g., Bill Clinton's name was accompanied by the face of Bill Clinton), a detail that might have played a role in the results. This detail was rectified in Forster and Lavie (2021), where they replicated Lavie et al.' (2003) work. This time, they administered a name-gender classification task, with distractor celebrity faces (whose gender either matched or did not match the gender of the name), and found that distractor interference was reduced at higher perceptual load. Only if identity priming was used (i.e., in the congruent condition, the name matched the image of the celebrity) were the perceptual load effects disrupted (as in Lavie et al., 2003).

Moreover, the study by Lavie et al. (2003) left open the question of whether faces in general or only famous faces "cancel out" the effect of perceptual load. There is evidence—although not in a directly comparable experimental setting—that the assertion holds for faces in general. Neumann and Schweinberger (2008) as well as Neumann et al. (2011) built on an effect known as repetition priming event-related potential (ERP; N250r), which is an occipito-temporal ERP modulation (in the time range of 200-350 ms) elicited by immediate repetitions of a face. In these studies, the first brief presentation of faces was embedded in a low versus high load letter flanker task (see above). Neumann and Schweinberger (2008) found an N250r for famous faces, that is, a kind of conceptual replication of Lavie et al. (2003). More important for our discussion, Neuman et al. yielded the same effect for unfamiliar faces, whereas the same pattern was not observed for control stimuli (houses and hands). In terms of the ERP task, irrelevant faces are processed at even higher levels of load, supporting the idea that faces are generally an exception to the moderating effect of perceptual load. However, what is missing here is the evidence that face processing comes with the behavioral consequence of distractor interference. Importantly, other behavioral work has drawn a distinction between familiar and unfamiliar faces in the perceptual load paradigm (He & Chen, 2010)

He and Chen (2010) designed a direct follow-up to Lavie et al. (2003) by varying the familiarity of faces. They conducted a study where female and male first names were presented as targets and the task was to categorize the gender of the name; faces were presented as distractors. The authors varied perceptual load as Lavie did with a varying number of pseudo-names into which the target was embedded. They hypothesized that the interference from the famous faces in the experiment by Lavie et al. (2003) was not due to the innate importance of the faces but was due to their familiarity. They presented participants with familiar[1] and unfamiliar faces as distractors. In line with perceptual load theory, they observed congruency effects (i.e., responses were faster if the gender of the name and face matched compared with a nonmatch) for unfamiliar faces at low but not high perceptual load. However, familiar faces interfered even at higher levels of load, thereby highlighting the importance of face familiarity as a moderator of perceptual load effects.

To our surprise, up to now, no study has tested an additional feature of faces that might "cancel out" the effect of perceptual load: emotional expression. In the following experiments, we wanted to examine whether the processing of emotional faces is moderated by perceptual load in the same way as it is for familiar faces.

The Case for Emotional Faces

We wanted to test whether emotional faces are processed even when they are task-irrelevant and are presented in a complex perceptual context (high perceptual load). Does the mere presence of an emotional face distractor vary behavioral responses to a target stimulus? Prior research indicates that processing emotional information is prioritized; both in terms of speed (Globisch et al., 1999) and in terms of interference with perception (Pessoa et al., 2002; Tipples & Sharma, 2000).

For instance, various attentional paradigms, such as the dot-probe task (e.g., Bar-Haim et al., 2007; MacLeod et al., 1986), have provided support for an attentional bias toward emotional faces. In a dot-probe paradigm, participants are first asked to fixate on a centrally presented cross, after which an emotional and a neutral stimulus are briefly presented (with random assignment) to the left and right of the fixation cross. A target stimulus (e.g., two dots) replaces

[1] Familiarity was implemented by having participants undergo a familiarization phase before performing the load experiment; the participants were presented with a set of faces and asked to memorize them for a later test.

one of the cue stimuli; participants are asked to categorize a feature (e.g., whether the two dots are vertically or horizontally aligned). If the response is faster when the target replaced the emotional cue as opposed to the neutral cue, it is typically assumed that the emotional cue captured the participant's attention. Wirth and Wentura (2018, 2019, 2020) conducted a series of experiments using a variation of the dot-probe paradigm to test for the conditions under which the bias to emotional faces occurs. Wirth and Wentura (2019) used the dot-probe task to test whether an attentional bias to angry faces occurs by using. angry and neutral faces as cues. There were indeed significant cueing scores indicating that spatial attention was captured by the angry face (Wirth & Wentura, 2019; see also Wirth & Wentura, 2023). The effect was found, however, only in a "social processing mode" that was defined by the fact that targets were schematic faces as opposed to scrambled schematic faces. In a similar experiment with happy and neutral faces as cues, Wirth and Wentura (2020) found that there was a bias toward happy faces, irrespective of target type. From these experiments, we can gather that there is an attentional bias toward emotional faces and that perhaps the social significance of emotions differs.

In a perceptual load experiment, as in the dot-probe paradigm, the emotional stimuli are completely task-irrelevant. Nevertheless, a perceptual load experiment is clearly different from a dot-probe task. First, a dot-probe effect is usually interpreted in terms of spatial attentional biases; the cue (e.g., the emotional face) has no influence on response generation. To the contrary, although it can be debated whether spatial attention processes contribute to effects found in the perceptual load paradigm (see Gaspelin et al., 2014; Luck et al., 2021) (we will return to this issue in the *General Discussion*), the flanker interference—that is, the "meddling" of the distractor with response generation—is in the foreground (Wentura & Rothermund, 2003). Second, dot-probe effects are obtained in a scarce perceptual environment, whereas, in perceptual load experiments, the point is whether (or to what extent) distractors still meddle with response formation in a perceptually complex environment.

We should hasten to add that some published studies have presented neutral and emotional faces while varying perceptual load; however, these studies did not test the meddling of distractors with response generation (i.e., they had no flanker design). Several studies have explored the perception of fear-inducing stimuli and fearful faces under varying levels of perceptual load (Norberg et al., 2010; Pessoa et al., 2005; Schindler et al., 2021; Wang et al., 2015). For instance, Wang et al. (2015) conducted a letter discrimination task with letters presented in the center flanked by neutral or fearful but otherwise identical faces on either side. The faces were presented very briefly and then masked. Behaviorally (i.e., in response times), they found no difference between the two face types. The authors analyzed several ERP components and found some differences between fearful and neutral faces at low but not at high load. Similarly, in a Functional Magnetic Resonance Imaging (fMRI) study, Pessoa et al. (2005) showed that differences in amygdala activity between the presentation of task-irrelevant fearful faces and neutral faces were found only at the lowest level of perceptual load. Thus, these studies indicate that the processing of emotional faces is influenced by a load manipulation.

On the other hand, Muller-Bardorff et al. (2016) showed that, overall, processing of emotional stimuli was unaffected by load manipulation in a paradigm where faces were presented in the center and were flanked by two horizontal bars on either side. Load was manipulated by varying the lengths of the bars, and participants were asked to make a length judgment. On average, an emotional distractor led to a larger N170 (i.e., an ERP associated with facial identity and expression) amplitude than neutral distractors did. There were some differences within emotions: Angry distractors led to a higher N170 amplitude than happy distractors at lower levels of load, but there was no difference between them at higher levels of load. Unlike Wang et al. (2015), the difference between emotional and neutral distractors was not qualified by an interaction with load. In support of this finding, a recent Electroencephalography (EEG) study by Schindler et al. (2021) showed that face-processing visual ERPs tend to act independently of perceptual load manipulations.

Finally, Gupta et al. (2016) presented a circular arrangement of letters on the screen, one of them being an X or an N, and participants had to categorize them accordingly. Load was manipulated by presenting either five Os as filler letters (i.e., low load) or the set H, K, W, M, and Z (i.e., high load). In their Experiment 2, in a minority of trials (in total 25%), either an angry or a happy face was presented in the center of the screen, filling the otherwise empty space of the letter circle. Of interest was whether the emotional expressions differed in prolonging responses to the targets, depending on load. Interestingly, there was no difference between angry and happy faces at low load, but they found a difference at high load: Trials with happy faces were associated with longer RTs compared with angry faces (see Gupta & Srinivas, 2015, for a comparable result). Thus, they found a happy bias, in contrast to the neuroscientific evidence of Muller-Bardorff et al. (2016), who found no difference between emotions at higher levels of load.

Taken together, a variety of experiments have been run, heterogeneous not only in their design and measures but also in results. Moreover, all studies were limited in scope, as the irrelevant distractor faces were not associated with responses (i.e., no flanker designs were used). Remington and Folk (2001) proposed an important distinction between *attention* and *selection*. Here, *attention* refers to orienting to a stimulus, whereas *selection* means activating the features of a stimulus such as to influence response processes. Under selection, a task-relevant feature of distractors can influence RTs by activating the associated response so that target-related performance is positively or negatively affected, depending on whether the task-relevant target feature is congruent or incongruent with the distractor feature (Klein et al., 2015; Lachter et al., 2009; Remington & Folk, 2001). There is a debate on whether attention without selection is possible (see, e.g., Zivony & Lamy, 2016), but that is not our point here. The dot-probe task showed only that emotional faces can attract attention; the potential for "meddling" was simply not tested. In the context of the perceptual load paradigm when the irrelevant face distractors have no feature overlap with the response-relevant features—as in the experiments cited in the preceding paragraphs—again, their potential for "meddling" could not be adequately assessed. Thus, using a flanker paradigm in which the irrelevant distractors are as easy to categorize as the targets with respect to the task at hand would lead to "selection" and allow for the generation of flanker effects.

For the sake of completeness it should be mentioned that there are two studies that used a flanker task with (a) a gender classification and (b) faces as the distractor, and (c) a variation of emotional expression of the distractor faces (Kim et al., 2017, Exp. 1; Tannert & Rothermund, 2020, Exp. 4b). However, both did not vary the perceptual load, and both experiments can be considered as representing a low perceptual load (as there is no search for the target in a noisy environment). Not surprisingly, both show gender congruency effects that are not moderated by expression variation. This is exactly what we expect for our low perceptual load condition.

That is, Lavie's seminal face paradigm with the combination of emotional flankers (both happy and angry) has yet to be tested. This paradigm allows for the exploration of potential interactions with congruency, valence, and load. Thus, in the following experiments, the faces were task-irrelevant but still acted as flankers, we manipulated congruency, and there was an incremental increase in load (four levels of load).

Overview

We used He and Chen's (2011) version of Lavie et al.'s (2003) seminal task. Participants were provided with four differently sized sets (load manipulation) composed of mostly pseudo-names with one target name. They were expected to report the gender of the target name. The faces were presented to the side of the center, peripheral to the target. The gender of the irrelevant face distractor and the gender of the target name rendered each trial congruent (matching) or incongruent (nonmatching). For neutral faces, we expected a congruency effect at low perceptual load levels but not at high perceptual load levels, that is, a pattern that would replicate He and Chen's unfamiliar condition.

For emotional flankers, a straightforward transfer from He and Chen's (2011) familiar condition and from Lavie et al.'s (2003) famous face condition led to the expectation that flanker effects would be present at all levels of load. In the remainder of the text, we will refer to this expectation as the *strong hypothesis*. A weaker version of the basic assumptions was that emotional distractors would still cause flanker effects at a load level at which flanker effects for neutral faces had already ceased.

We began by conducting a pilot experiment with only neutral faces to establish the basic result pattern, that is, that neutral faces interfere at lower but not at higher perceptual load levels, adhering to perceptual load theory. In Experiment 1, we manipulated emotion as a factor, along with load and congruency, and we compared happy and neutral faces. Experiment 2 was similar with the exception of the emotional manipulation: we compared angry and neutral faces. In Experiment 3, we conducted a replication of Experiments 1 and 2 but with emotion as a between-subjects factor (i.e., happy/neutral vs. angry/neutral).

Pilot Experiment

This experiment was preregistered (https://aspredicted.org/YT7_B1N). In this pilot experiment, we investigated whether cognitive load would moderate the degree to which participants process neutral unfamiliar distractor faces. Participants were asked to perform a name categorization task of typical American male and female names presented amongst a set of pseudo-names while ignoring faces presented to the right or left of center. If the gender of the face and name matched, a trial was congruent; if there was a mismatch, the trial was incongruent. The load levels were increased by increasing the number of pseudo-names.

Method

Participants

We recruited $N = 66$ participants of which 38 were women, 24 were men, and four identified as (diverse) nonbinary; the age range was 19-35 years ($M = 26.4$ years, $SD = 4.2$). In total, 67 participants were recruited from Prolific, an online data collection tool. One participant was excluded from the final analysis because their overall accuracy fell below 75% (Tukey, 1977). All participants provided informed consent prior to testing.

On, prolific, an online platform for data collection we pre-selected some demographic criteria that created a customized data pool. These demographics are self-reported by participants when they register themselves on prolific. We filtered participants based on mother tongue: English, fluency in English: high, age range: 18 -35, and other researcher's approval rating of participants: 100%. All participants were compensated based on the calculation of 9 £ per hour, the average time of completing the experiment was 20 minutes and thus each participant was compensated with approx. 3 £ for this experiment.

We based our power calculation for detecting an interaction between congruency and load on Lavie's (2003) study, specifically the Load × Congruency interaction effect size ($\eta_p^2 = 0.11$). With a power of $1-ß = .80$ and $\alpha = .05$, we needed $N = 65$ participants. Power analyses were implemented with GPower (Faul et al., 2007). In the study by He and Chen (2010), the experiment that found an interaction between perceptual load and congruency for unfamiliar faces had a similar effect size of $\eta_p^2 = 0.13$. We calculated conservatively and used the smaller effect size.

Design

We used a 4 (Perceptual Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Congruency: incongruent vs. congruent trials) within-participant design. Load levels were defined by the number of pseudo names plus 1 (i.e., the target).

Materials

As distractors, we used eight male and eight female Caucasian faces from the Chicago face inventory (Ma et al., 2015). We used Adobe Photoshop to crop all images to show only the face and hair. We decided to keep the hair on the faces, similar to the way He and Chen (2010) presented their distractor face. Keeping the hair helps distinguish the gender of the faces. Considering the ages of potential participants, the target names were the eight female and eight male most common U.S. census baby names from the year 2001 (see Appendix A for the list of names). The pseudo-names were created by shuffling letter strings of 25 female and 25 male Caucasian names. An initial list of 700 pseudo-names was screened for any string that was too "name-like", which resulted in a final list of 550 pseudo-names with roughly equal sets of 5, 6, or 7 letters[2]. The targets and pseudo names were presented in Arial font with a height of app. 0.5 cm and the size of distractor faces was app. 4.4 cm × 5.5 cm. A complete list of the pseudo-names can be found at https://osf.io/3nz87 (see Appendix A for the list of names).

Procedure

The study was conducted online on Prolific with Psychopy version 2022.1.3. First, participants were given the task of adjusting the representation of a credit card shown on the screen so that it matched the size of an actual card. This procedure is standard in online experiments, used to ensure that the stimuli-to-screen ratio is consistent across the sample. Then participants were assigned a unique identification code, asked for their age and gender, and given a consent form. Subsequently, they were given instructions with a schematic representation of what the task would look like. They were first presented with a fixation cross for 500 ms, and

---

[2] The length of the target names varied from 5 to 9 letters (with only one name having 5 letters and 3 names having 9 letters; see Appendix A). Due to the use of a variable-width typeface in the presentation it was appropriate to limit the length of the pseudo-names to 7 because some of them (e.g., Bamamae, Aedbwno) have roughly the same factual length as, e.g., Alexander.

then the target name and the pseudo-names horizontally replaced the positioning of the fixation cross; the location of the target name within the set of pseudo names was chosen on a random basis. To the side of the center, a distractor face was presented. The distractor face was presented randomly either to the left or right of center. The participant's task was to categorize the real name as typically male or female. The pseudo-names, target name, and the faces were presented until response, and the participant's key press led to the next trial. Participants received both positive ("Right!") and negative feedback ("Too Slow!" or "Wrong key") in the practice trials and only received negative feedback in the main trials. There was one block of 24 practice trials and two blocks of 256 main trials each, with a self-timed pause between each block. The trial number was chosen because there were 4 (perceptual load) × 2 (gender) × 2 (distractor location) × 2 (congruency) × 8 (face identity) = 256 conditions.

Transparency and Openness

All experiments have been pre-registered, and links are provided within each experiment section. All data and analysis codes are available at the Open Science Framework. We have reported how we determined our sample size, all data exclusions are reported, all manipulations, and all measures in the study. The data was collected in 2022 for the pilot experiment, and 2023 for experiments 1, 2 and 3. The files can be accessed via the following link: osf.io/3nz87. Our studies did receive approval from the research ethics committee of the faculty of human and business sciences of Saarland University, Germany.

Results

For the Response Time (RT) analysis, only trials with correct responses were considered. The average classification accuracy was 94.1 % (*SD* = 4.3 %). Moreover, we removed RTs below 150 ms or above 1.5 Interquartile Range (IQR) of each participant's distribution (separately for the load conditions)[3] above the third quartile range (Tukey, 1977). This procedure led to the exclusion of 3.1% of all trials with correct responses. Table 1 shows the mean RTs. For

[3] We forgot to add "separately for the load conditions" in our preregistration. However, it is obvious that it does not make sense to collapse data from different load conditions with mean values around 800 ms (Load Level 2) to mean values around 1,500 ms (Load Level 8) in order to identify outliers.

the sake of brevity and comprehensibility, we calculated and analyzed flanker differences as dependent variables. That is, we subtracted congruent RTs from incongruent RTs for each load condition. Larger flanker differences are thus indicative of incongruent trials leading to slower RTs or congruent trials leading to faster RTs. Figure 1 shows the mean flanker differences. Analyses with mean RTs (for all experiments) are reported in *Appendix B*. Throughout the article, we applied an α criterion of .05. For simple flanker effects, we always used one-tailed tests.

Table 1

Mean RTs (Accuracy Rate) for Congruent and Incongruent Trials at Each Level of Load (Pilot Study)

	Trial type	
Load level	Congruent	Incongruent
2	786 (.95)	812 (.94)
4	1,049 (.95)	1,076 (.94)
6	1,302 (.94)	1,316 (.94)
8	1,527 (.94)	1,542 (.94)

Response Times

We conducted a repeated-measures MANOVA (O'Brien & Kaiser, 1985) with load level as a factor and flanker differences as the dependent variable. The overall congruency effect (i.e., the intercept test of the analysis) was significant, $F(1, 65) = 13.40$, $p < .001$, $\eta p^2 = .171$; that is, averaged across load conditions, there was a flanker effect. The load effect was not significant, $F(3, 63) = .46$, $p = .714$, $\eta p^2 = .021$, $F(1, 65) = 1.02$, $p = .316$, $\eta p^2 = .015$ for the linear trend, *F*s < 1 for the remaining trends.[4] That is, the MANOVA showed no significant decrease in the

[4] We used the MANOVA approach for repeated measures (O'Brien & Kaiser, 1985). This approach involves testing a vector of means of p-1 difference variables (where p = number of within-participant conditions) against the corresponding null vector. The difference variables can be created according to different schemes (e.g., Helmert contrasts, polynomial contrasts). Although this choice does not affect the overall test, it allows for more specific results depending on the research question. In the case of an ordered factor (such as perceptual load), it is most

flanker effects as the load increased. Regardless, it was adequate to look at the simple effects, as numerically the flanker effects for higher load levels were clearly smaller and burdened by more error variance (see Figure 1).[5] In one-sample *t* tests, there were significant flanker effects for Load Level 2, $t(65) = 7.83$, $p < .001$ (one-tailed), $d_Z = .96$, and Level 4, $t(65) = 3.47$, $p < .001$ (one-tailed), $d_Z = .43$. However, the flanker effects for Load Levels 6, $t(65) = 1.13$, $p = .131$, d_Z =.14, and 8, $t(65) = 1.00$, $p = .160$, $d_Z = .12$, were not significant.

The non-significance of the load effect (with flanker differences as the dependent variable) may be due to the large error variances in the higher load conditions or due to the fact that the decrease in the flanker effects from 26-27 ms (Load Levels 2 and 4) to 14 ms (Load Levels 6 and 8) was too marginal. A simple calculation indicated that it was primarily the large error variances that were responsible: When we linearly transformed the flanker differences for Load Levels 4, 6, and 8 so that they retained their means but had the same standard deviation as the flanker difference for Level 2, the main effect for load (in a repeated measures MANOVA with flanker differences as the dependent variable) was significant, $F(3, 63) = 4.42$, $p = .007$, $\eta p^2 = .174$, $F(1, 65) = 9.72$, $p = .003$, $\eta p^2 = .130$, $F < 1$, and $F(1, 65) = 3.31$, $p = .074$, $\eta p^2 = .048$ for the linear, quadratic, and cubic trend, respectively. (Note again, this finding should not be taken as an empirical result but only as a hint about the interpretation of the non-significance of the load effect in the main analyses.)

Accuracy

Accuracy data were checked in order to detect possible speed-accuracy tradeoffs. This was not the case here. For the main ANOVA, the overall congruency effect was not significant, $F(1, 65) = 1.38$, $p = .245$, $\eta p^2 = .021$. The load effect was not significant either, $F < 1$.

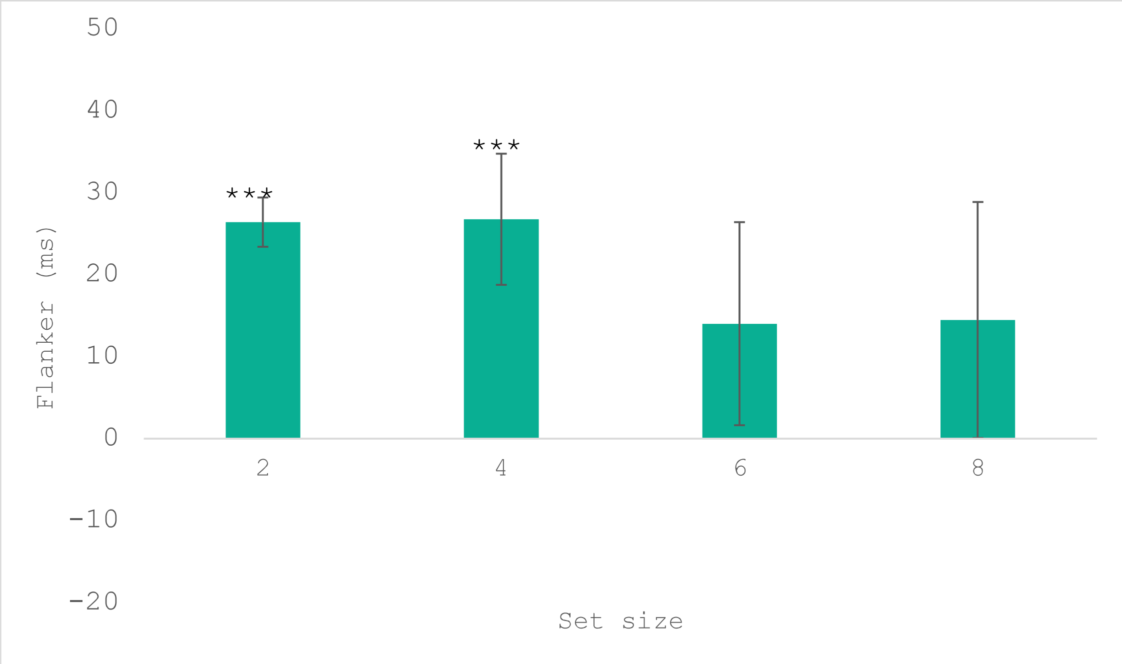
---

useful to additionally report polynomial contrasts. This has the advantage of providing more specific tests, which are potentially more powerful than the omnibus test.

[5] Note, it was *a priori* planned to report the congruency effects for each level of perceptual load (see our preregistration).

Figure 1

Graphical Representation of the Differences Between Incongruent-Congruent Trials (i.e., Flanker Effects; Pilot Study)



Note. Error bars depict ±1 standard error of the mean; asterisks (***) denote $p < .001$

Discussion

In the pilot study, we found a broadly expected pattern of results. Unfamiliar neutral faces were distracting at lower load levels but not at higher load levels. These results are in line with perceptual load theory. Whereas neutral expressions act as effective distractors at lower loads, the same is not true at higher load conditions. Admittedly, the effect for load (with flanker differences as the dependent variable) was not significant. We attribute this finding to the large error variances in the higher load conditions; a post hoc calculation showed that the decline in mean flanker differences between lower and higher levels of load should be considered valid. Thus, in terms of flanker effects, our pilot study replicated He and Chen's (2010) results for unfamiliar faces.

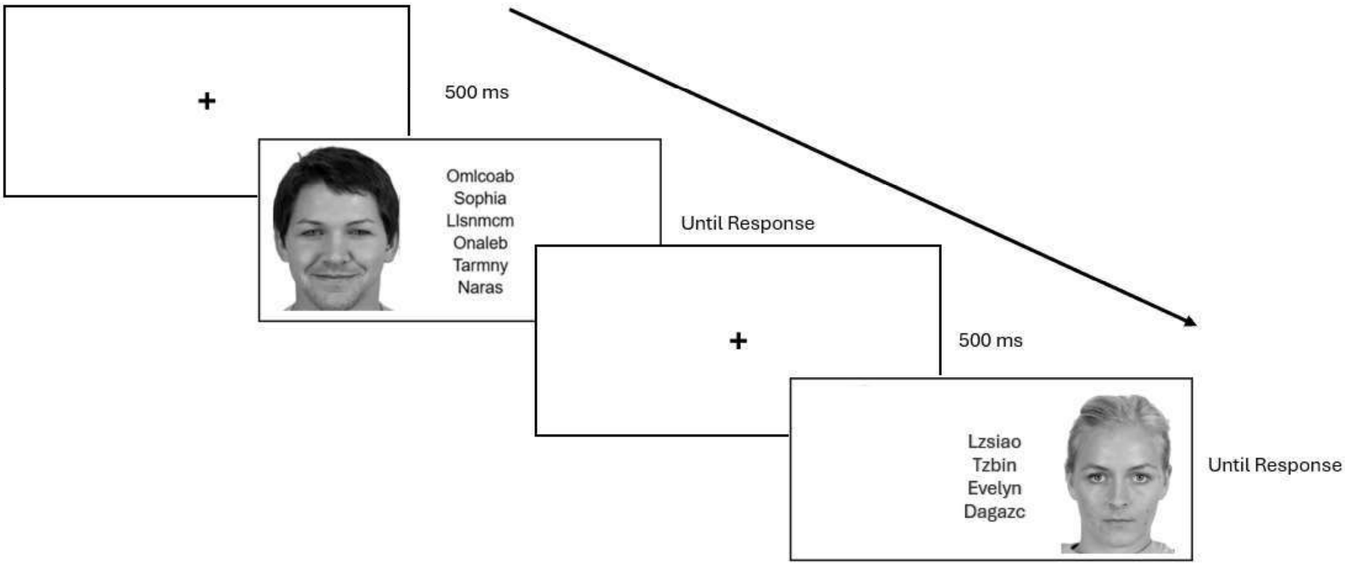
In the next three experiments, we tested whether emotional stimuli differed from neutral stimuli by leading to a flanker effect at all levels of load. This finding would be equivalent to the effects found for famous faces (Lavie et al., 2003) and familiar faces (He & Chen, 2010). Will emotional faces lead to flanker effects even at higher loads, even though they are completely task irrelevant? Are emotional faces an exception to perceptual load effects?

Our strong hypothesis was that we would find flanker effects at all load levels for emotional faces (whereas neutral faces would show the pattern found in the pilot study). A weak version of the hypothesis is that emotional faces outperform neutral faces in the sense that emotional faces still show flanker effects at a level of perceptual load at which interference by neutral flankers has already disappeared.

We report the experiments in chronological order. As a preview of the findings, Experiment 1 deviated from our expectations, in that it showed evidence only for the weak hypothesis but not for the strong one. This situation is of course preferable to the reverse (i.e., finding a hypothesis-congruent result in an early experiment that cannot be replicated in subsequent experiments). However, in terms of the narrative structure of an empirical article, this constellation is not so welcome, as the first experiment typically sets the agenda; hence, this advance organizer. Finally, we present a cross-experiment analysis, which shows that the overall picture is quite clear and that "experiment" as a factor did not moderate the main results.

Figure 2

Example of a trial sequence with (a) an incongruent trial with load level 6 (target name of "Sophia", and a happy male face as distractor) followed (b) by a congruent trial with load level 4 (target name of "Evelyn", and a neutral female face as distractor).



Experiment 1

This experiment was preregistered (https://aspredicted.org/PQM_WBJ). In this experiment, participants were equally likely to be presented with either a neutral or a happy irrelevant face distractor. The aim of this experiment was to replicate the pattern of neutral flankers and to test for whether emotionally irrelevant distractor faces would lead to flanker effects regardless of level of load (strong hypothesis) or at least to a flanker effect at a perceptual load level at which flanker effects for neutral faces had already ceased.

Method

Participants

The final analysis included $N = 144$ participants of which 63 were women, 72 were men, and 10 were diverse; their ages ranged from 18-35 years ($M = 27.7$ years, $SD = 4.8$ years). A total of 155 participants were recruited from Prolific. Eleven participants were excluded from the final analysis because their overall accuracy fell below 75%. All participants provided informed consent prior to testing. The same filtering criteria were used as in the pilot, and the participants were again compensated based on 9 £/hour. The experiment on average took 25 minutes to complete and participants were compensated with approximately 4.5 £.

The focus of power planning was on the congruency effect for Load Levels 6 and 8 combined because, according to the pilot study, we expected no differences between neutral and happy faces for the lower load levels. For neutral faces, we expected to replicate a null effect, whereas, for happy faces, we expected a congruency effect comparable to the congruency effect for perceptual Load Level 4 in the pilot study ($d_Z = 0.43$; i.e., ~ 0.40). The difference between the effects for happy and neutral faces should be significant at Load Levels 6 and 8 (combined). Our power planning (see below) focused directly on the difference in flanker effects for emotional and neutral faces for the higher load levels to keep the sample size in a manageable range. The contrast between happy and neutral faces was estimated to be 0.40/squareroot(2) = 0.28 (as we assumed equal variances for the congruency differences for happy and neutral faces and a null correlation between them). According to GPower (Faul et al, 2007), to detect an effect of size $d_Z = 0.28$, given an error of $\alpha = .05$ with a power of $1 - \beta = .95$, we needed a sample size of 140 participants.

Design

We used a 2 (Face Type: neutral vs. happy expression) × 4 (Perceptual Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Distractor: incongruent vs. congruent trials) with all factors varied within participants.

Materials

We used the same faces from the Chicago Face inventory that were used in the pilot study. We used both their neutral and closed mouth happy expressions. Dot-probe experiments have shown that exposed teeth are a strong perceptual confound in happy expressions (Wirth & Wentura, 2018b), so we employed only happy faces with closed mouths in the present study. They were edited in the same way as they were in the pilot, and the same target and pseudo-names were used as in the pilot.

Procedure

We used the same procedure that was used in the pilot study. The practice trials included both neutral and happy face expressions so that participants knew that they would be exposed to two types of distractors.

Results

For the RT analysis, only trials with correct responses were considered. The average classification accuracy was 94.0 % (*SD* = 4.8 %). For the exclusion of RT outliers, the same criteria were applied as in the pilot. This procedure led to the exclusion of 3.1% of all trials with correct responses. Table 2 shows the mean RTs. Figure 3 (top left) shows the flanker effects for the Emotion × Load conditions.

Response Times

For the sake of convenience (and as preregistered), we first report a 2 (Face Type: happy vs. neutral) × 4 (Load: 2 vs. 4 vs. 6 vs. 8) repeated-measures MANOVA (O'Brien & Kaiser, 1985) with flanker differences as the dependent variable. There was a significant overall congruency effect, $F(1, 143) = 22.85$, $p < .001$, $\eta p^2 = .138$. Moreover, there was a significant main effect of load, $F(3, 141) = 4.03$, $p = .009$, $\eta p^2 = .08$, but we did not find either a significant main effect of face type, $F < 1$, or a Face Type × Load interaction, $F(3, 141) = 1.60$, $p = .193$, $\eta p^2 = .033$, $F(1, 143) = 1.58$, $p = .211$, $\eta p^2 = .011$, $F < 1$, and $F(1, 143) = 2.03$, $p = .156$, $\eta p^2 = .014$ for linear, quadratic, and cubic trends)

Given the backdrop of the pilot experiment, we expected differences between emotional and neutral distractors at Load Levels 6 and 8 (see the preregistration and section *Participants* above). Figure 3 already indicates that the hypotheses regarding emotional distractors at these levels (i.e., flanker effects persisting at a load level where neutral flanker effects vanished) were not supported. Numerically, the flanker effects for neutral distractors even exceeded those for happy distractors. However, in a 2 (Load: 6 vs. 8) × 2 (Face Type: happy vs. neutral) ANOVA with flanker differences as the dependent variable there was no significant main effect of face type, $F(1, 143) = 1.26$, $p = .264$, $\eta p^2 = .009$ (*F*s < 1 for the main effect of load and the Face Type × Load interaction). In terms of the simple effects, only the neutral distractor flanker effect at load level 6 was significant, $t(143) = 2.23$, $p = .014$ (one-tailed), $d_Z = .19$, the distractor effect for happy flanker was not significant, $t(143) = 1.10$, $p = .136$ (one-tailed), $d_Z = .09$. For the highest level of load, neither the neutral distractor flanker effect, $t(143) = 1.37$ (one-tailed), $p = .087$, $d_Z = .11$, nor the happy distractor flanker effect $|t| < 1$, were significant.

Figure 3 also indicates that our presupposition of clear flanker effects (irrespective of face expression) at Load Levels 2 and 4 did not hold. A 2 (Load: 2 vs. 4) × 2 (Face Type) ANOVA with flanker differences as the dependent variable yielded a significant main effect of load, $F(1, 143) = 9.77$, $p = .002$, $\eta p^2 = .06$, and a main effect of face type that just missed the criterion for significance, $F(1, 143) = 3.54$, $p = .062$, $\eta p^2 = .02$. The Face Type × Load interaction was not significant, $F(1, 143) = 2.04$, $p = .155$, $\eta p^2 = .01$. Simple flanker effects were as follows. For Load Level 2, both effects were significant; happy faces: $t(143) = 7.32$, $p < .001$ (one-tailed), $d_Z = .61$; neutral faces: $t(143) = 6.62$, $p < .001$ (one-tailed), $d_Z = .55$. For Load Level 4, only the flanker effect for happy distractors was significant, $t(143) = 3.44$, $p < .001$ (one-tailed), $d_Z = .29$; but $|t| < 1$ for neutral distractors. The difference was significant, $t(143) = 1.98$, $p = .025$ (one-tailed), $d_Z = .16$.

The flanker differences for emotional and neutral flankers at Level 4 were plagued by some outliers, which decrease the power of conventional tests (Wilcox, 2013). Therefore, we repeated the tests for the flanker effects at Load Level 4 with a robust one-sample t test (yuen.t.test function from the R package PairedData; Champely 2018; see Wilcox, 2013, with regard to robust testing) with a trimming of $\gamma = .25$, which yielded $t(72) = 3.68$, $p < .001$ (one-tailed), $d_{Z'} = .34$ for emotional distractors, and $|t| < 1$ for neutral distractors. A robust t test for dependent measures (yuend function from the R package WRS2; $\gamma = .25$; Mair & Wilcox, 2020) yielded $t(72) = 2.15$, $p = .017$ (one-tailed), $dz' = 0.19$, for the difference.[6]

Table 2

Mean Response Times (Error Rates in Parentheses) for Congruent and Incongruent Trials at Each Level of Load (Experiment 1)

	Neutral		Happy	
Load level	Incongruent	Congruent	Incongruent	Congruent
2	824 (.94)	796 (.95)	828 (.94)	796 (.95)
4	1088 (.94)	1086 (.94)	1092 (.95)	1072 (.95)

[6] An attentive reader might ask whether the contrast between neutral and happy flankers at level 6 and 8 (which are numerically hypothesis-incongruent) might be significant if treated by robust testing as well. This is not the case, $t(72) = -0.83$, $p = .205$ (one-tailed), $dz' = 0.07$, for the difference at load level 6 and $t(72) = -0.07$, $p = .472$ (one-tailed), $dz' = 0.01$, for the difference at load level 8.

6	1349 (.94)	1327 (.94)	1338 (.94)	1328 (.95)
8	1574 (.93)	1555 (.94)	1568 (.92)	1561 (.93)

Accuracy

For the accuracies in a 2 (Face Type: happy vs. neutral) × 4 (Load: 2 vs. 4 vs. 6 vs. 8) repeated-measures MANOVA with flanker differences as the dependent variable, there were no significant effects except a significant overall congruency effect, $F(1, 143) = 14.84$, $p < .001$, $\eta p^2 = .094$, all other Fs < 1.37.

Discussion

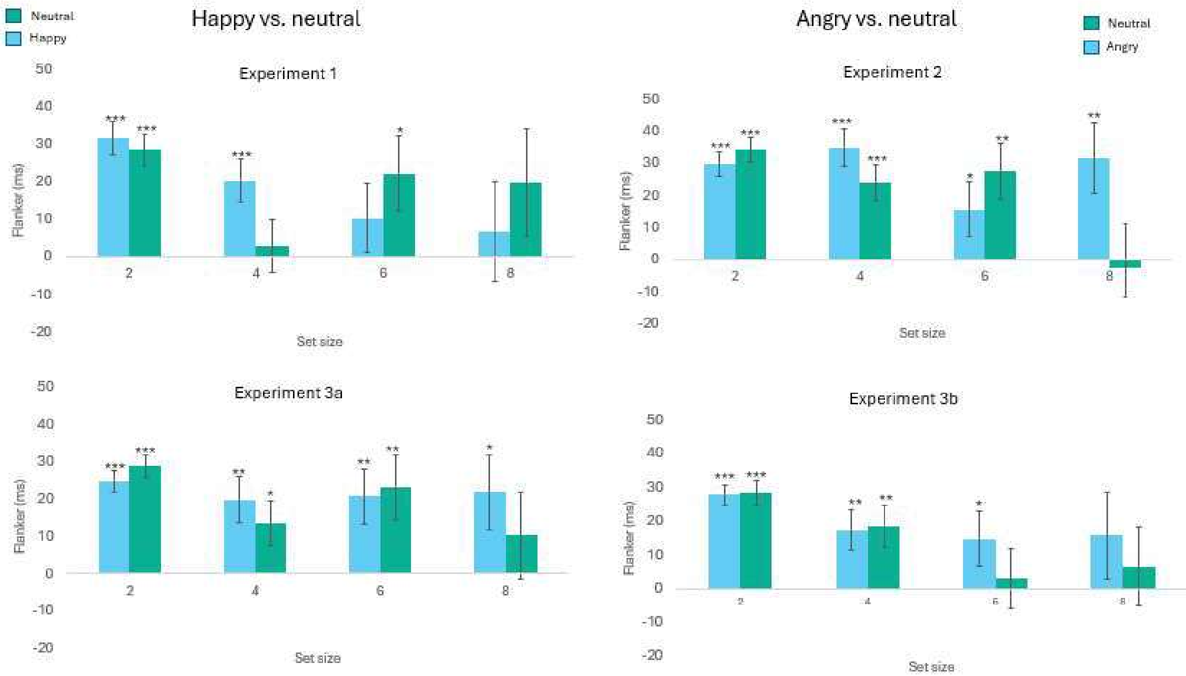
Experiment 1 provided no support for the strong hypothesis (i.e., significant flanker effects for emotional distractors at all load levels, whereas flanker effects for neutral faces were expected to vanish at higher load levels). Happy distractors showed flanker effects only for the two lower load levels.

There is some evidence for the weak hypothesis (i.e., a flanker effect for emotional distractors for a load level at which the neutral effect had already vanished). At Load Level 4, we still observed a flanker effect for happy faces, whereas there was a clear null effect for neutral faces, and the difference was significant. However, this result has to be taken with caution since in the Pilot study neutral distractors caused a flanker effect at Load Level 4 and we found a small effect for neutral faces at load level 6 in the present experiment. With Experiment 3, we will report an attempt to replicate the results.

But first we wanted to test what would happen to negative stimuli in the same paradigm. In previous studies, effects have occurred more often with negative stimuli than with positive (Moriya et al., 2014; Öhman et al., 2001; Wirth & Wentura, 2019). Thus, it might be the case that negative stimuli create distractor interference at higher levels of load than happy faces in the perceptual load paradigm. In Experiment 2, emotion was manipulated with angry and neutral irrelevant face distractors.

Figure 3

Flanker Effects (i.e., the Difference Between Gender-Incongruent and Congruent Trials) as a Function of Load Level and Emotion (Experiments 1, 2, and 3a,b)



Note. Error bars depict ±1 standard error of the mean; asterisks denote $p < .001$ (***), $p < .01$ (**), and $p < .05$ (*)

Experiment 2

The experiment was preregistered (https://aspredicted.org/BLM_JRH).

Method

Participants

The final sample included $N = 180$ participants of which 68 were women, 111 were men, and one was diverse; their ages ranged from 19-35 years ($M = 29.2$ years, $SD = 4.5$). A total of 204 participants were recruited from Prolific. Twenty-three participants were excluded from the final analysis because their overall accuracy fell below 75%. One additional participant was excluded, as their average RT was three IQR over the general RT distribution. All participants provided informed consent prior to testing. The same filtering criteria were used as in the pilot, and the participants were compensated with the same amount as in Experiment 1.

In terms of the power analysis, we expected there to be a moderating effect of emotion on congruency (i.e., flanker effect for angry faces vs. no effect for neutral faces) at one of the load levels beyond Level 2. (Given the results of the pilot study and Experiment 1, we expected a congruency effect for both expression types at the lowest load.) For example, if the congruency effect for neutral faces was already null at Load Level 4 (and did not return at higher levels), we still expected a congruency effect for angry faces at this level. (This finding would conform to the result for happy faces in Experiment 1.) If the effect for neutral faces was not null until Load Level 6 (and did not return at Level 8), we still expected a congruency effect for angry faces at this level. Finally, if the effect for neutral faces did not vanish until Level 8, we still expected a congruency effect for angry faces at this level; this would then also provide evidence for the strong hypothesis.

At whatever load level the decisive test was applied, we expected the contrast of flanker effects to be comparable to the congruency effect contrast between a happy face and a neutral face for perceptual Load Level 4 in Experiment 1. According to G*Power, to detect an effect of size $d_Z = 0.19$ given $\alpha = .05$ (one-tailed) with a power of $1 - \beta = .80$, we needed a sample size of 173 participants.

Design

We used a 2 (Face Type: angry vs. neutral expression) × 4 (Perceptual Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Congruency: incongruent vs. congruent trials) within-participant design.

Materials and Procedure

Everything was the same as in Experiment 1, except that happy faces were replaced by (closed mouth) angry faces. For two of the stimulus identities chosen for Experiment 1, no closed mouth angry face was available in the stimulus set. Thus, to keep consistent identities across neutral and angry expressions, we replaced two face identities from the pilot study and Experiment 1.

Results

The average classification accuracy was 93.9 % (*SD* = 4.2 %). To exclude RT outliers, we applied the same criteria as in Experiment 1. This led to the exclusion of 2.9 % of all trials with correct responses. See Table 3 for the mean RTs and accuracies; see Figure 3 (top right) for the flanker effect.

Response Times

Again, we first report the overall 2 (Face Type: angry vs. neutral) × 4 (Load: 2 vs. 4 vs. 6 vs. 8) repeated-measures MANOVA with flanker differences calculated from correct RTs as the dependent variable. There was a significant overall congruency effect, $F(1, 179) = 82.22$, $p < .001$, $\eta p^2 = .315$. The main effects of load, $F(3, 177) = 1.68$, $p = .173$, $\eta p^2 = .028$, and face type, $F(1, 179) = 1.56$, $p = .214$, $\eta p^2 = .009$, were not significant. Finally, the Load × Face Type interaction was significant, $F(3, 177) = 2.87$, $p = .038$, $\eta p^2 = .046$; $F(1, 179) = 2.67$, $p = .104$, $\eta p^2 = .015$, $F(1, 179) = 1.29$, $p = .258$, $\eta p^2 = .007$, and $F(1, 179) = 5.74$, $p = .018$, $\eta p^2 = .031$ for linear, quadratic, and cubic trend, respectively.

To elucidate this interaction, a repeated-measures MANOVA of the flanker effects for angry faces yielded a significant overall congruency effect, $F(1, 179) = 52.87$, $p < .001$, $\eta p^2 = .228$, but no load effect, $F(3, 177) = 1.29$, $p = .281$, $\eta p^2 = .021$ (*Fs* < 1 for the linear and quadratic trend, respectively; $F(1,179) = 3.66$, $p = .057$, $\eta p^2 = .020$ for the cubic trend). The simple flanker effects for all load levels were significant, $t(179) = 7.63$, $p < .001$ (one-tailed), $d_Z = 0.57$, $t(179) = 5.92$, $p < .001$ (one-tailed), $d_Z = 0.44$, $t(179) = 1.83$, $p = .035$ (one-tailed), $d_Z = 0.14$, $t(179) = 2.87$, $p = .002$ (one-tailed), $d_Z = 0.21$, for Load Levels 2, 4, 6, and 8, respectively. [7]

A corresponding MANOVA of the neutral flanker effects yielded a significant overall congruency effect, $F(1, 179) = 32.18$, $p < .001$, $\eta p^2 = .152$, and a significant load effect; $F(3, 177) = 3.21$, $p = .024$, $\eta p^2 = .052$, $F(1, 179) = 6.93$, $p = .009$, $\eta p^2 = .037$, $F < 1$, and $F(1,179) = 2.03$, $p = .156$, $\eta p^2 = .011$ for the linear, quadratic, and cubic trend, respectively. The simple flanker effects were significant up to Level 6, $t(179) = 9.40$, $p < .001$ (one-tailed), $d_Z = 0.70$,

[7] Some readers might expect an alpha adjustment here because we conducted eight simple flanker tests (i.e., Load × Face type) in this *Results* section. Of course, this would be appropriate if we would hypothesize "faces cause flanker interference, in *one or more (non-specified)* Load × Face Type conditions" (disjunction testing, Rubin, 2021). In that case, one can legitimately claim that keeping $\alpha = .05$ for the single tests yields a Type I error rate of $1-.95^8 = .34$, whereas adjustment keeps a Type I error rate at $\alpha = .05$. Our strong hypothesis, however, means that the flanker effect for emotional faces should in principle be present *at all* load levels (conjunction testing) whereas the flanker effect for neutral faces should be vanished (at least) at the highest level of load. Thus, we should neither decrease the power of the simple tests for emotional faces nor should we increase the probability of a Type II error for the flanker effect of neutral faces at Load Level 8 by adjusting the alpha criterion. Thus, adjustment is not appropriate here (see, e.g., Kennedy, 2024; Rubin, 2021). But for the sake of transparency, Bonferroni-Holm adjustment would only alter a single result in Exp.2: The flanker effect of angry faces at Load Level 6 (see text) would fail the adjusted criterion of $\alpha_{adj} = .05/2$.

$t(179) = 4.29$, $p < .001$ (one-tailed), $d_Z = 0.32$, $t(179) = 3.14$, $p < .001$ (one-tailed), $d_Z = 0.23$, for Load Levels 2, 4, and 6, respectively. For Load Level 8, however, the flanker effect for neutral distractors was nonexistent, $|t| < 1$.

These results are in alignment with the results of our a-priori planned (preregistered) comparison. We identified Load Level 8 as the first level at which no neutral effect existed. At that level, however, there was still an effect for angry faces (see above), as the contrast between the flanker effect for angry and neutral faces was significant, $t(179) = 1.94$, $p = .027$ (one-tailed), $d_Z = 0.14$.

Table 3

Mean RTs for Congruent and Incongruent Trials at Each Level of Load (Experiment 2)

	Neutral		Angry	
Load level	Incongruent	Congruent	Incongruent	Congruent
2	818 (.93)	784 (.96)	817 (.93)	788 (.96)
4	1075 (.94)	1051 (.94)	1088 (.94)	1053 (.95)
6	1323 (.93)	1295 (.94)	1318 (.93)	1303 (.94)
8	1527 (.92)	1528 (.93)	1543 (.92)	1512 (.94)

Accuracy

For accuracy, a 2 (Face Type: angry vs. neutral) × 4 (Load: 2 vs. 4 vs. 6 vs. 8) repeated-measures MANOVA yielded a significant overall congruency effect, $F(1, 179) = 33.54$, $p < .001$, $\eta p^2 = .158$, as well as significant main effect of load, $F(3, 177) = 8.38$, $p < .001$, $\eta p^2 = .124$. As can be seen from Table 3, flanker effects were more pronounced at the lowest load level. The main effect of emotion was not significant, $F(1, 179) = 1.76$, p = .187, $\eta p^2 = .010$, nor was the Load × Emotion interaction, $F < 1$.

Discussion

With regard to our preregistered criterion, we can state that our strong hypothesis is supported in Experiment 2. Neutral flankers produced significant congruency effects up to Load Level 6, but clearly no effect at Load Level 8. Angry flankers, however, produced significant congruency effects at all load levels; the difference in flanker effects at Load Level 8 was significant.

The conclusion is corroborated by the standard MANOVA analysis: It showed a significant Load × Face Type interaction; follow-up analyses yielded an overall congruency effect that was not moderated by load for angry flankers, but which was moderated for neutral flankers. Thus, the pattern corresponded closely to the pattern that He and Chen (2010) found for familiar versus unfamiliar faces.

For two reasons, we ran Experiment 3, which was a replication of both Experiments 1 and 2 but with happy versus angry distractors as a between-participants comparison. First, we wanted to see whether the "advantage" of angry faces over happy faces could be replicated in a direct comparison. Second, in Experiment 1, happy flankers were only significant up to Load Level 4, whereas neutral flankers did not produce a flanker effect at this level. This finding could be taken as support for the weak hypothesis. However, the results for neutral faces in Experiment 2 (i.e., significant flanker effects for neutral faces at Load Level 4 and even at Level 6) undermined this interpretation.

Experiment 3

Experiment 3 was a replication of Experiments 1 and 2, with type of emotion as a between-participants factor.[8]

Method

Participants

The final analysis included $N = 366$ participants ($n = 178$ for angry vs. neutral, $n = 188$ for happy vs. neutral) of which 200 were women, 164 were men, and 2 were diverse; their ages ranged from 18-35 years ($M = 27.5$ years, $SD = 4.7$ years). A total of 390 participants were recruited on Prolific. Twenty-three participants were excluded from the final analysis because their overall accuracy fell below 75% (Tukey, 1977). Furthermore, one more participant was excluded because their average RT was three IQR above the average of other participants. All participants provided informed consent prior to testing. The same filtering criteria were used as

[8] This experiment was again preregistered (https://aspredicted.org/pmr6-3c7w.pdf). However, the preregistration was partly based on an erroneous initial analysis of Experiment 2, which indicated a different result pattern. Therefore, this preregistration was valid only to a limited extent.

in the pilot, and the participants were compensated as in previous experiments (i.e. approx. 4.5 £).

As indicated by Footnote 8 power planning for Experiment 3 was based on an erroneous initial analysis of Experiment 2. Nevertheless, for all within-participant comparisons, the power was comparable to Experiments 1 and 2. With $N = 366$, we were able to detect within-participant effects of size $d_Z = 0.13$ and between-participants effects of size $d = 0.26$ given an error of $\alpha = .05$ (one-tailed) with a power of $1 - \beta = .80$.

Design

We used a 2 (Face Type: emotional vs. neutral) × 4 (Perceptual Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Congruency: incongruent vs. congruent) × 2 (Emotion Type: happy vs. angry) ANOVA with face type, load, and congruency as within-participant factors and emotion type as a between-participants factor.

Materials and Procedure

Everything was the same as in Experiments 1 and 2, except that the face identities from Experiment 2 were now used for the happy versus neutral sample as well. That is, in comparison with Experiment 1, two of eight identities were replaced. (Of course, the new happy face stimuli were closed-mouth exemplars as well.)[9]

Results

The average classification accuracy was 95 % (*SD* = 3.7 %). To exclude the RT outliers, we applied the same criteria as in Experiment 1, thus leading to the exclusion of 3.0 % of all trials with correct responses. Table 4 presents the mean RTs and accuracy values. Figure 3 (bottom) shows the flanker effects for the Emotion × Load conditions (left for happy faces, right for angry faces).

Response Times

The 2 (Face Type: emotional vs. neutral) × 4 (Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Emotion Type: happy vs. angry) repeated-measures MANOVA with the first two factors as within-

[9] For exploratory reasons, participants completed the trait-anxiety scale of the State-Trait Anxiety Inventory (STAI; Laux, Glanzmann, Schaffner, & Spielberger, 1981) at the end of the procedure. For the interested reader, a brief summary of (largely null) results are given in *Appendix D*.

participant factors, emotion type as a between-participants factor, and flanker differences as the dependent variable yielded a significant overall congruency effect, $F(1, 364) = 93.30$, $p < .001$, $\eta p^2 = .204$. Moreover, there was a significant main effect of load, $F(3, 362) = 5.99$, $p < .001$, $\eta p^2 = .047$, $F(1, 364) = 5.84$, $p = .016$, $\eta p^2 = .016$ for the linear trend, *F*s < 1.08 for the remaining trends), but no main effect of emotion, $F(1, 364) = 1.00$, $p = .318$, $\eta p^2 = .003$, or group, $F < 1$. The most interesting interaction, the Face Type × Load interaction, was not significant, $F < 1$ (all *F*s < 1.08 for the trends); all remaining interactions were nonsignificant as well, all *F*s < 1.

Thus, the overall analysis suggested a decrease in flanker effects from lowest to highest load without indicating any differences between emotional and neutral distractors. However, the danger of a "false positive" result lurked. A robust *t* test for dependent measures comparing the linear trends for emotional versus neutral distractors yielded $t(183) = 1.65$, $p = .05$ (one-tailed), $d_{Z'} = 0.10$ (see *Appendix C* for details). Additionally, given the result of Experiment 2, we felt we could justify conducting separate MANOVAs for emotional and neutral distractors.

Although there was no hint of differences between emotions (i.e., between angry and happy) not in the MANOVA reported above nor in a robust *t* test for independent samples regarding the linear trend for emotional distractors, $|t| < 1$—we kept the between-participants variable emotion type in these analyses to see whether there was (a) any hint of a difference between angry and happy distractors or (b) a difference between neutral distractors that are presented in the contexts of angry versus happy faces.

For emotional distractors, a 4 (Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Emotion Type: happy vs. angry) repeated-measures MANOVA with flanker differences as the dependent variable yielded a significant overall congruency effect, $F(1, 364) = 56.03$, $p < .001$, $\eta p^2 = .133$, but no load effect, $F(3, 362) = 1.43$, $p = .234$, $\eta p^2 = .012$ (all *Fs* < 1 for the trends; $|t| < 1$ for a robust test of the linear trend; see *Appendix C*), no emotion type effect, and no Load × Emotion Type effect, both *F*s < 1. All simple flanker effects were significant, $t(365) = 12.52$, $p < .001$ (one-tailed), $d_Z = 0.65$, $t(365) = 4.26$, $p < .001$ (one-tailed), $d_Z = 0.22$, $t(365) = 3.29$, $p < .001$ (one-tailed), $d_Z = 0.17$, $t(365) = 2.33$, $p = .010$ (one-tailed), $d_Z = 0.12$, for load levels 2, 4, 6, and 8, respectively.

For the sake of transparency, it should, however, be noted that the overall result (i.e., significant flanker effects for all load levels) repeats for happy faces, but not for angry faces (see Figure 3). The flanker effect for angry faces at Load Level 8 is associated with $t(177) = 1.23$, $p = .111$ (one-tailed). We will return to this issue in the *Discussion*.

For neutral distractors, a corresponding repeated measures MANOVA – that is, a 4 (Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Emotion Type: neutral in the context of happy vs. neutral in the context of angry) analysis – yielded a significant overall congruency effect, $F(1, 364) = 36.80$, $p < .001$, $\eta p^2 = .092$, and a significant load effect, $F(3, 362) = 4.72$, $p = .003$, $\eta p^2 = .038$, $F(1, 364) = 5.73$, $p = .017$, $\eta p^2 = .015$ for the linear trend; Fs < 1 for quadratic and cubic trends; $t(183) = 2.70$, $p = .007$, $d_Z = 0.18$ for a robust test of the linear trend, see *Appendix C*); there was no emotion type effect and no Load × Emotion Type interaction, both Fs < 1. Simple flanker effects were significant up to load level 6, $t(365) = 11.93$, $p < .001$ (one-tailed), $d_Z = 0.62$, $t(365) = 3.71$, $p < .001$ (one-tailed), $d_Z = 0.19$, $t(365) = 2.15$, $p = .016$ (one-tailed), $d_Z = 0.11$, for load levels 2, 4, and 6. At load level 8, however, the flanker effect for neutral distractors was not significant, $t(365) = 1.03$, $p = .153$ (one-tailed), $d_Z = 0.05$.

Again, the flanker differences for emotional and neutral flankers at Load Level 8 were burdened by outliers. We therefore repeated the tests for the flanker effects at Load Level 8 with a robust one-sample t test (see Experiment 1) with a trimming of $\gamma = .25$, which yielded $t(183) = 2.94$, $p = .002$ (one-tailed), $d_{Z'} = .17$ for emotional distractors, but $|t| < 1$ for neutral distractors. A robust t test for dependent measures ($\gamma = .25$) yielded $t(183) = 1.66$, $p = .049$ (one-tailed), $d_{Z'} = 0.09$ for the difference.

Table 4

Mean RTs (Error Rates in Parentheses) as a Function of Face Type (Emotional vs. Neutral) and Trial Type (Congruent vs. Incongruent), Load Level, and Emotion Type (Experiment 3)

	Emotional		Neutral	
Load level	Incongruent	Congruent	Incongruent	Congruent
Emotion = Angry				
2	796 (.94)	769 (.96)	794 (.95)	765 (.96)

4	1057 (.94)	1040 (.94)	1059 (.95)	1040 (.95)
6	1306 (.93)	1291 (.95)	1305 (.95)	1302 (.95)
8	1530 (.93)	1514 (.94)	1526 (.94)	1520 (.94)
Emotion = Happy				
2	817 (.94)	792 (.96)	819 (.95)	791 (.96)
4	1073 (.95)	1053 (.96)	1071 (.95)	1058 (.95)
6	1317 (.94)	1296 (.95)	1326 (.95)	1303 (.95)
8	1549 (.93)	1527 (.95)	1540 (.94)	1530 (.94)

Accuracy

The overall 4 (Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Face Type: emotional vs. neutral) × 2 (Emotion Type: happy vs. angry) repeated-measures MANOVA with flankers differences as the dependent variable yielded a significant overall congruency effect, $F(1, 364) = 16.85$, $p < .001$, $\eta p^2 = .044$. There was a main effect of load, $F(3, 362) = 6.88$, $p < .001$, $\eta p^2 = .054$. As can be seen from Table 4, flanker effects were more pronounced at the lowest load level. There were no other effects, all Fs < 1.

Discussion

With some caution, Experiment 3 again provided support for the strong hypothesis: Emotional distractors caused flanker effects up to the highest load level, whereas neutral distractors clearly showed no flanker effects at the highest load level. There were no indications of differences between angry and happy distractors.

However, there is again a caveat: If we analyze flanker effects separately for happy and angry faces – despite the absence of statistical interactions –, the overall result (i.e., significant flanker effects for all load levels) repeats for happy faces; for angry faces, the effect at Load Level 8 was not significant (see Figure 3).

Thus, we are faced by two inconsistencies across the three experiments: First, the just mentioned one for angry faces. Second, for happy faces there is a discrepancy between Experiment 1 (i.e., only evidence for the weak hypothesis) and Experiment 3, which shows evidence for the strong hypothesis.

Thus, curiously enough, we are faced with one result supporting the strong hypothesis for each of the two emotions and a second result for each emotion not supporting a strong hypothesis, which prevents us from drawing a clear conclusion. For cases like this –that is, data of different experiments show consistent trends, but provide some inconsistencies with regard to statistical inference, it is recommended to use the full power of all data at hand. For example, Goh and colleagues (2016) suggest conducting a "mini meta-analysis" on the experiments that are reported in a manuscript. However, if the raw data of all experiments are available it is self-suggestive to conduct a "mega-analysis" (e.g., Eisenhauer, 2021, Scoboria et al., 2017), that is, to pool the raw data of all experiments. Thus, to arrive at a conclusion, we decided to use the full statistical power offered by combining our data sets.

Cross-Experiment Analyses

In summary, we found consistent results regarding the weak hypothesis in that all experiments showed an advantage for emotional distractors. However, regarding the strong hypothesis, for happy faces Experiment 1 contrasted with Experiment 3 to some extent, whereas Experiment 2 and 3 were not fully consistent with regard to angry faces.

Thus, we decided to conduct a cross-experiment analysis to see if meaningful differences remained in this analysis and to use the full power of the combined data set for a final conclusion regarding the weak and strong hypotheses.

Response Times

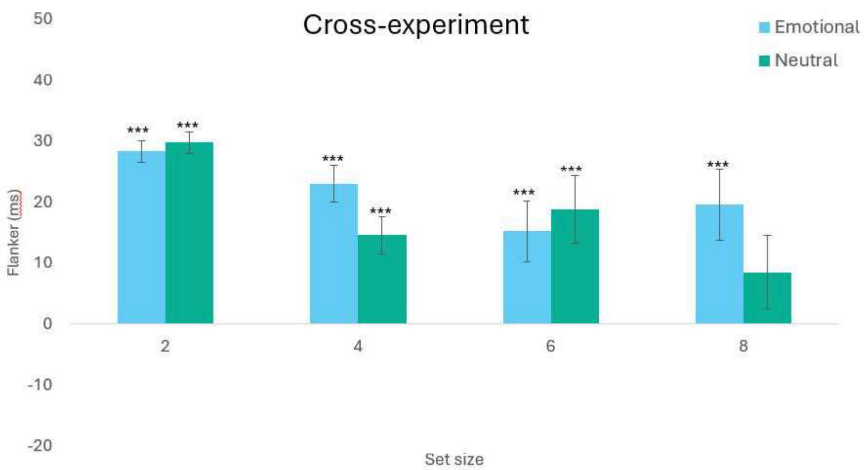
A 4 (Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Face Type: emotional vs. neutral) × 2 (Emotion Type: happy vs. angry) × 2 (Experiment: Experiments 1/2 vs. Experiment 3) repeated-measures MANOVA with the first two factors as within-participant factors, the latter two as between-participants factors, and flanker differences as the dependent variable yielded a significant overall congruency effect, $F(1, 686) = 182.66$, $p < .001$, $\eta p^2 = .210$. There was a significant main effect of load, $F(3, 684) = 10.62$, $p < .001$, $\eta p^2 = .045$. which was further qualified by the cubic trend of the of the Emotion × Load interaction, $F(1, 686) = 4.12$, $p = .043$, $\eta p^2 = .006$; $F(3, 684) = 1.99$, $p = .114$, $\eta p^2 = .009$, for the omnibus test Fs < 1 for linear and quadratic trend. All other effects were not significant, all Fs < 1.43, except $F(1, 686) = 3.31$, $p = .069$, $\eta p^2 = .005$ for the interaction of the two between-participants factors.

Thus, we see that there are no moderations by emotion type (i.e., angry vs. happy) and experiment (1/2 vs. 3). Therefore, Figure 4 shows the flanker effects for the Emotion × Load conditions, with the emotional and the respective neutral flankers combined across experiments.

The cross-experiment graph in Figure 4 reveals why differences between emotional and neutral faces caused the cubic trend of the Emotion × Load interaction to be significant. Not only Load Level 8 but also Load Level 4 showed a numerical difference between the flanker effects for emotional and neutral distractors. As follow-up analyses, we conducted separate MANOVAs for emotional and neutral flankers.

Figure 4

Flanker Effects (i.e., the Difference Between Gender-Incongruent and Congruent Trials) as a Function of Load Level and Emotion (all Experiments combined)



Note. Error bars depict ±1 standard error of the mean; asterisks denote $p < .001$ (***)

A MANOVA for the emotional flanker effects yielded a significant overall congruency effect, $F(1, 686) = 112.24$, $p < .001$, $\eta p^2 = .141$, and an effect of load, $F(3, 684) = 3.60$, $p = .013$, $\eta p^2 = .016$, showing $F(1, 686) = 3.58$, $p = .059$, $\eta p^2 = .005$ for the linear trend and *F*s < 1.26 for the remaining trends, indicating that the flanker effects were somewhat different between load levels. However, as can be seen from Table 1 all simple flanker effects were significant with *p*s <

.001 (one-tailed). All other effects in the MANOVA were not significant, all *F*s < 1.47, except $F(1, 686) = 2.78$, $p = .096$, $\eta p^2 = .004$ for the interaction of the two between-participants factors.

Table 5

T-tests, Effect Sizes, and Bayes Factors for the Flanker Effects (Exp. 1, 2, and 3 combined)

Type	Load	t(689)	p	d_Z	BF_{+0}	Evidence category (in favor of H.)
Emotional						
	2	16.11	< .001	0.61	3.9×10^{46}	Extreme (H_1)
	4	7.61	< .001	0.29	9.3×10^{10}	Extreme (H_1)
	6	3.81	< .001	0.14	108.8	Extreme (H_1)
	8	3.35	< .001	0.13	22.2	Strong (H_1)
Neutral						
	2	16.47	< .001	0.63	2.7×10^{48}	Extreme (H_1)
	4	4.95	< .001	0.19	1.4×10^{4}	Extreme (H_1)
	6	4.18	< .001	0.16	462.8	Extreme (H_1)
	8	1.39	.082	0.05	0.205	*Substantial (H_0)*

Note. BF_{+0} = Bayes factor for the hypothesis of a positive flanker effect. (We used a scale parameter of r = .707 (= √2/2) for the JZS H1 effect size prior.) Evidence categories according to Jeffreys (1961; slightly modified by Wagenmakers et al., 2011)

A corresponding MANOVA for the neutral flanker effects yielded a significant overall congruency effect, $F(1, 686) = 80.19$, $p < .001$, $\eta p^2 = .105$, and a significant load effect, $F(3, 684) = 8.91$, $p = .003$, $\eta p^2 = .038$; linear trend: $F(1, 686) = 8.75$, $p = .003$, $\eta p^2 = .013$; cubic trend: $F(1, 686) = 3.59$, $p = .059$, $\eta p^2 = .005$; quadratic trend: $F < 1$. Importantly, as can be seen from Table 1 the simple flanker effects were significant (with $p < .001$, one-tailed) only up to Level 6; for Load Level 8, however, the flanker effect for neutral distractors was not significant. All other effects in the MANOVA were not significant, all *F*s < 1, except $F(3, 684) = 1.85$, $p = .137$, $\eta p^2 = .008$, for the Load × Face Type interaction.

Again, the flanker differences for emotional and neutral flankers were plagued by some outliers. Therefore, we repeated the tests for the flanker effects at Load Level 8 with a robust one-sample *t* test ($\gamma = .25$), which yielded $t(345) = 4.09$, $p < .001$ (one-tailed), $d_{Z'} = .17$ for emotional distractors, but $t(345) = 0.88$, $p = .190$ (one-tailed), $d_{Z'} = .04$ for neutral distractors. A robust *t* test for dependent measures yielded $t(345) = 2.09$, $p = .019$ (one-tailed), $d_{Z'} = 0.09$, for the difference. Thus, according to these analyses the difference between emotional and neutral faces is even more pronounced than indicated by the conventional analyses.

Table 5 included Bayes factors as well. As can be seen, for 6 out of 8 conditions the Bayes factor indicated "extreme" evidence in favor of the existence of flanker effect. However, whereas there is still "strong" evidence in favor of a flanker effect for emotional faces at Load Level 8, there is "substantial" evidence *for the null hypothesis* for neutral faces at this load level.

To finally conclude that the strong hypothesis holds for both emotions, we separately tested the flanker effects at the highest load level for angry and happy flankers. The conventional tests yielded $t(357) = 2.80$, $p = .003$ (one-tailed), $d_Z = 0.15$ for angry faces and $t(331) = 1.89$, $p = .030$ (one-tailed), $d_Z = 0.10$, for happy faces. Given Experiment 1, it was expectable that the result for happy flankers would be a bit weaker than for angry faces. However, we are be allowed to conclude that the strong hypothesis holds for angry and happy distractors. Moreover, again outliers distorted the picture to some extent. Robust tests ($\gamma = .25$) yielded a clearer picture for angry distractors: $t(179) = 2.90$, $p = .002$ (one-tailed), $d_{Z'} = 0.17$; and happy distractors: $t(165) = 2.89$, $p = .002$ (one-tailed), $d_{Z'} = 0.17$.

Accuracy

A 4 (Load: 2 vs. 4 vs. 6 vs. 8) × 2 (Face Type: emotional vs. neutral) × 2 (Emotion Type: happy vs. angry) × 2 (Experiment: Experiments 1&2 vs. Experiment 3) repeated-measures MANOVA with the first two factors as within-participant factors, the latter two as between-participants flankers, and flanker effects of errors as the dependent variable yielded a significant overall congruency effect, $F(1, 686) = 61.11$, $p < .001$, $\eta p^2 = .082$. There was a significant main effect of load, $F(3, 684) = 13.81$, $p < .001$, $\eta p^2 = .057$. Finally, the main effect of experiment was significant, $F(1, 686) = 5.84$, $p = .016$, $\eta p^2 = .008$; the average flanker effect was a bit larger in Experiments 1 and 2 compared with Experiment 3. All other *F*s < 1.09, except the main effect of emotion, $F(1, 686) = 2.71$, $p = .100$, $\eta p^2 = .004$. Behind this tendency is the fact that the average flanker effect for emotional flankers was numerically larger than the one for neutral distractors (see also below).

For Load Level 2, both flanker effects were significant; emotional distractors: $M = 1.55\%$ ($SD = 5.41\%$), $t(689) = 7.52$, $p < .001$ (one-tailed), $d_Z = 0.29$; neutral distractors: $M = 1.53\%$

(*SD*= 5.74%), $t(689) = 7.01$, $p < .001$ (one-tailed), $d_Z = 0.27$. For Load Level 4, there was a significant effect for emotional distractors, $M = 0.47\%$ (*SD*= 5.56%), $t(689) = 2.22$, $p = .013$ (one-tailed), $d_Z = 0.08$, but not for neutral distractors, $M = 0.20\%$ (*SD*= 5.28%), $|t| < 1.02$. The same holds for Load Level 6: There was a significant effect for emotional distractors, $M = 0.60\%$ (*SD*= 5.96%), $t(689) = 2.65$, $p = .004$ (one-tailed), $d_Z = 0.10$, but not for neutral distractors, $M = 0.21\%$ (*SD*= 5.53%), $|t| < 1$. For Load Level 8, both effects were significant, $M = 0.73\%$ (*SD*= 6.27%), $t(689) = 3.07$, $p = .001$ (one-tailed), $d_Z = 0.12$ for emotional distractors, $M = 0.41\%$ (*SD*= 6.55%), $t(689) = 1.65$, $p = .049$ (one-tailed), $d_Z = 0.06$, for neutral faces.

General Discussion

Through three experiments, we were able to show that emotional faces led to flanker effects even at higher levels of perceptual load than neutral faces. We initially presented only neutral faces to be able to gauge baseline results (pilot). Then we went on to manipulate emotions in the first two experiments, first with neutral and happy faces (Experiment 1), then with neutral and angry faces (Experiment 2). Experiment 3 was a between-participants design to replicate both Experiments 1 and 2.

With some caution we can conclude that the data indicated support for the strong hypothesis: Whereas emotional faces led to a flanker effect at all levels of load, neutral faces led to an effect at all load levels except one, that is the highest level of load. Moreover, as per the robust analysis the difference between neutral and emotional flankers at the highest level was significant.

The conclusion is based finally on a cross-experiment analysis, which allowed us to capture the entire scope of the data. In the interest of a balanced presentation, it should be mentioned again that the results of the individual experiments were not completely conclusive. We acknowledge that in Experiment 1, with happy and neutral faces, there was only some support for the weak hypothesis but no support for the strong one. We acknowledge as well that in Experiment 3 the results for angry faces – if analyzed on their own despite the clear non-significance of any interaction effects related to the angry vs. happy factor – did not show a flanker effect for angry faces at the highest level of load. Thus, it seems as if we are faced by random fluctuations based on the fact that flanker effects at the higher load levels are smaller than initially expected. Hence, our mega-analysis (i.e., the cross-experimental analysis) is the appropriate tool to draw an overall picture.

These findings extend existing research on how emotional faces are influenced by load manipulation. Where researchers such as Wang et al. (2015) and Pessoa et al. (2005) have found neuroscientific evidence that emotional stimuli are influenced by load manipulation, the current behavioral results offer evidence from within a response competition paradigm that emotional flankers conclusively outperform neutral flankers at high levels of load.

Moreover, the cross-experiment analysis provided clarity on the overall results, through which we concluded that there was support for our strong hypothesis.

Perceptual Complexity or Attentional Slippage?

Now we provide a more thorough theoretical discussion of perceptual load effects. In fact, there are three theories that can explain the typical perceptual load effect, that is, the disappearance of flanker effects at higher levels of load. We will briefly present these theories and then discuss how the exception—that is, the finding that emotional faces still showed flanker effects at high levels of load—fits with these theories.

We introduced Lavie's own perceptual load theory (Lavie, 1995; Lavie & Tsal, 1994) as a starting point in our *Introduction.* It focuses on the ease versus difficulty of target processing. If it is difficult to process the target (as in a high perceptual load condition), there is no perceptual capacity left for processing irrelevant stimuli; thus, distractors are not processed. If, however, target processing is easy (as in a low perceptual load condition), perceptual processing unavoidably spills over to other stimuli; thus, distractors are processed and can cause flanker effects when applicable. We expected the perceptual load hypothesis to account for the results of neutral faces and for emotional faces to be an exception to the rule. At higher levels of load, processing the complex relevant information depletes a person's perceptual capacity to an extent that there is no capacity left for the perception of neutral stimuli, but as shown in the aforementioned experiments, emotional stimuli are still processed. Ultimately, the "relevance" of emotional faces led to them being processed, even at higher levels of load. It remains a bit vague how specific types of stimuli (famous faces, Lavie et al., 2003; familiar faces, He & Chen, 2010; emotional faces, the present study) will boost the capacity for perceptual processing (but see He

& Chen, 2010, for speculations employing the dual-pathway theory of perception of Goodale & Milner, 1992).

In their dilution theory, Tsal and Benoni (2010) argued that the irrelevant distractor flankers are processed to the same degree in both high and low load conditions. However, in a typical high load situation, not only the (features of the) distractor but also the (features of the) filler stimuli are processed. Thus, the processing of the distractor is diluted in the high load condition compared with the low load condition (which has no filler stimuli or has very homogeneous fillers). The extant perceptual load literature then does not have a condition that can distinguish between high load and high dilution. So, to distinguish their account from perceptual load theory, the authors manipulated target appearance (a featural difference, e.g., size or color, that perceptually separated the target from the fillers) while retaining the same number of stimuli as in the high load condition (which they termed the "dilution" condition). From the perspective of perceptual load theory, this is a low load condition because the target can easily be identified (hence, distractor interference is expected). From the perspective of dilution theory, almost nothing has changed compared with the high perceptual load condition: The processing features of the fillers still "dilute" the processing of the features of the distractor (hence, no distractor interference is expected). Indeed, a flanker effect did not appear in the "dilution" condition.

Thus, in our gender name categorization paradigm, when there were fewer fillers (pseudo-names), then the trial corresponded to a low dilution condition, and when the relevant distractors increased, then the trial corresponded to a high dilution condition. Accordingly, neutral irrelevant distractors led to flanker effects at low dilution, but the same did not occur in the high dilution condition. Flanker effects that emerged for emotional faces when there was a limited number of fillers persisted even when the fillers increased, indicating that emotional faces in this context were an exception to dilution effects. Because dilution theory focuses on the early stages of visual processing, that is, when processing simple features, it remains unclear how the exception of emotional faces can be integrated into the theory. To be more specific: If we had not carefully prevented emotional faces from showing teeth, then one might have argued that processing the (whiteness of) teeth would have been a basic feature whose processing was not diluted by the features of the fillers. But, as we explained, we specifically used faces that did not show teeth.

For perceptual load theory and dilution theory, the reasoning is not based on shifts in spatial attention, but the situation is different for the third explanation of the typical perceptual load effects, that is, slippage theory (Gaspelin et al., 2014). First, slippage theory is an early selection account: Only stimuli that are attended to are processed on a semantic level. However, "slippage" means a shift in spatial attention to a location that is task-irrelevant. The explanation of typical perceptual load effects by slippage theory is based on contingent capture theory (Folk et al., 1992). This theory states that task-irrelevant stimuli capture attention if they share target-defining features. In low load conditions, this description holds only for the distractor. Therefore, the probability that spatial attention will "slip" to the distractor is large. Accordingly, a large flanker effect can be expected. In high load conditions, the distractor and the fillers share target-defining features; hence, the probability that the distractor will capture attention is low. Accordingly, no substantial flanker effect will be expected. (Interestingly, the result found via the manipulation described by Tsal & Benoni, 2010 [see above], fits slippage theory as well: If the target is always red, no other stimulus in the display shares this feature; hence, there is no "slippage" on the basis of contingent capture.)

Slippage can then explain why flanker effects occur with neutral and emotional faces when there are limited fillers because the distractor shares target-defining features (i.e., here: a stimulus that can easily be categorized as male or female). When a larger number of fillers that share target-defining features is introduced (i.e., here: the name-like string of letters), they grasp attention, which leads to greater competition. Thus, ultimately attentional capture by neutral faces only occurs with low probability when there are a larger number of fillers. We can refer to the concept of a spatial "priority map" as used, for example, by Wolfe (1994, 2021) in his *guided search theory*. The priority map guides attention according to activity peaks. In a low load condition, only a few locations are characterized by notable activity values, that is, for Load Level 2: the target location, the distractor location, and the filler location. In the high load condition with Load Level 8, however, six more locations have a considerable activity value. If we assume—just for the sake of simplicity—that the expected value of activity caused by the face distractor and the pseudo-names are roughly comparable, but the concrete activity values in a given trial differ by added noise, we can easily model a situation in which the probability of an attentional shift toward the distractor (and hence the flanker effect) depends on the number of competitors. If, for example, the probability that a single filler will win the competition is $p = .4$, for seven fillers, then the joint probability that at least one will win the competition is $p = 1 - .6^7 = .97$.

Of course, the competition with pseudo-names also exists if emotional faces are presented instead of neutral faces. However, if we assume that the expected value of the activity of the distractor is considerably exceeded relative to a neutral face, it might always exceed the factual activity values of the fillers; hence, attentional capture still occurs regardless of the increase in the number of fillers. Note that this logic even allows for a graded dependence of attentional capture by the emotional face from the number of fillers. If the expected value of the activity of the emotional distractor is increased only to a degree that the added noise has a low probability of leading to an activity value of a filler that exceeds the one for the distractor in a given trial, we can model a slight dependence of flanker effects by the number of distractors. The joint probability that at least one filler activity exceeds the distractor's activity becomes larger with a larger number of fillers. If, for example, the probability that a single filler will win the competition is $p = .03$, then for seven fillers, the joint probability that at least one will win the competition is $p = 1 - .97^7 = .19$.

Note, that the value of stimuli is indeed integrated into the latest version of "guided search theory" (Wolfe, 2021) as one of five types of attentional guidance (based, e.g., on Anderson et al., 2011; see also Wentura et al., 2014).

Thus, across these theories, the most concrete explanation for the mechanisms behind the current effects is provided by slippage theory. The propensity of emotional faces to attract attention is greater than the one neutral face. Emotional faces capture attention at a point when neutral faces cease to do so.

Attentional capture by Emotional Faces

Our findings emphasize the special role that emotional faces play, in that their perception does not diminish even at higher levels of load (Müller-Bardorff et al., 2016; Neumann et al., 2011; Schindler et al., 2021). Thus, our conclusion that it is in fact the emotional faces that lead to distractor interference, regardless of the degree of complexity of the center task, offers further support for the crucial role that emotional faces play in our daily lives (Frischen et al., 2008).

We should briefly address again the study by Gupta and colleagues (2016; see also Gupta & Srinivasan, 2015) because superficially the study was somewhat similar to ours (i.e., it tested interference by emotional faces under low and high load conditions as well). We found an overall advantage of emotional flankers over neutral flankers (i.e., no difference between happy and angry faces), whereas Gupta et al. found an advantage for happy faces over angry faces at higher levels of load. Our results differed, but what differed even more were the paradigm and the task. They used a circular target presentation and presented faces in the center of the screen. Moreover, the frequencies of the emotional distractors were decreased, as the authors were concerned about habituation. Finally, they assessed whether the presence of emotional distractors non-specifically slowed down responses (whereas we assessed flanker effects). Again, an explanation based on spatial attention is helpful here: Whereas our results are best explained by attentional capture towards the peripheral emotional face, spatial attention is initially on the centered face in the experiments by Gupta et al. and differences in RTs between different face types can be related to attentional dwelling (i.e., how easy it is to draw attention away from a task-irrelevant stimulus). Thus, the basic process is a different one in our experiments (i.e., attentional capture) than in the study by Gupta et al. (2016; i.e., attentional dwelling). Hence, the results do not contradict each other. It is open to further research to relate these two versions of the paradigm to each other.

Furthermore, as we know from Wirth and Wentura (2020, 2019, 2023), bias toward emotional faces can be contextual. They used a variation of the dot-probe paradigm where they showed that, within a social attentional set, there was a bias toward angry faces. But the bias toward happy faces occurred irrespective of target type or attentional set. As in the conventional dot-probe paradigm, the task employed cueing. What we can conclusively infer is that, within the attentional set of the social condition, attention was spatially directed toward angry faces, thus indicative of a context-based attentional bias toward anger, but also a context-independent attentional bias toward happiness. The current results further add to these findings by showing that an attentional "bias" toward emotional faces persists even when participants contend with increasingly complex information. Thus, the takeaway message is that emotional faces do not simply lead to an attentional bias, but they explicitly outperform neutral faces at high levels of load.

Constraints of Generality

One question we cannot answer with the present study is whether its results will be found with other emotion types that have a strong component of arousal such as disgust, and fear. We have used angry and happy emotional faces and have found results that are supportive of our hypothesis within the context of the perceptual load paradigm. Thus, we might not be able to extend the findings to all emotions.

The participants in our study were collected via Prolific, an online data collection platform. We specifically collected data from an English-speaking population in the USA. As our theoretical considerations and hypotheses relate to general human cognitive processes, this gave us the opportunity to test and possibly falsify our hypothesis. For the moment, they have passed the OSFtest.

Conclusion

Across three experiments where we manipulated emotion in the conventional perceptual load paradigm by Lavie et al. (2003), we found that emotional flankers were an exception to the perceptual load hypothesis. Neutral stimuli led to flanker effects at lower levels of load; however, at the highest load level, the neutral flanker effect disappeared. Emotional stimuli displayed a flanker effect at all levels of load. Thus, emotional face stimuli tended to act as familiar faces did in He and Chen's (2010) gender-name categorization experiment. The result can best be explained by an increased probability of attentional slippage (Gaspelin et al., 2014) towards emotional faces.

References

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. Proceedings of the National Academy of Sciences of the United States of America, *108*(25), 10367-10371. https://doi.org/10.1073/pnas.1104047108

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

Broadbent, D. E. (1958). Perception and communication. Pergamon Press.

Carmel, D., Saker, P., Rees, G., & Lavie, N. (2007). Perceptual load modulates conscious flicker perception. *Journal of Vision*, *7*(14), 14. https://doi.org/10.1167/7.14.14

Champely , S. (2018). *PairedData: Paired Data Analysis. R package version 1.1.1.* https://CRAN.R-project.org/package=PairedData

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90. https://doi.org/10.1037/h0039515

Eisenhauer, J. G. (2021). Meta-analysis and mega-analysis: A simple introduction. *Teaching Statistics*, *43*(1), 21–27. https://doi.org/10.1111/test.12242

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 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

Forster, S., & Lavie, N. (2021). Faces are not always special for attention: Effects of response–relevance and identity. *Vision Research*, *189*, 1–10. https://doi.org/10.1016/j.visres.2021.08.002

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin, 134*(5), 662–676. https://doi.org/10.1037/0033-2909.134.5.662

Gaspelin, N., & Luck, S. J. (2019). Inhibition as a potential resolution to the attentional capture debate. *Current Opinion in Psychology*, *29*, 12–18. https://doi.org/10.1016/j.copsyc.2018.10.013

Gaspelin, N., Ruthruff, E., & Jung, K. (2014). Slippage theory and the flanker paradigm: An early-selection account of selective attention failures. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(3), 1257–1273. https://doi.org/10.1037/a0036179

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. *Social & Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Gupta, R., Hur, Y. J., & Lavie, N. (2016). Distracted by pleasure: Effects of positive versus negative valence on emotional capture under load. *Emotion*, *16*(3), 328–337. https://doi.org/10.1037/emo0000112

Gupta, R., & Srinivasan, N. (2015). Only irrelevant sad but not happy faces are inhibited under high perceptual load. *Cognition and Emotion*, *29*(4), 747–754. https://doi.org/10.1080/02699931.2014.933735

Globisch, J., Hamm, A. O., Esteves, F., & Öhman, A. (1999). Fear appears fast: Temporal course of startle reflex potentiation in animal fearful subjects. *Psychophysiology*, *36*(1), 66–75. https://doi.org/10.1017/S0048577299970634

Hay, J. L., Milders, M. M., Sahraie, A., & Niedeggen, M. (2006). The effect of perceptual load on attention-induced motion blindness: The efficiency of selective inhibition. *Journal of*

Experimental Psychology: Human Perception and Performance, *32*(4), 885–907. https://doi.org/10.1037/0096-1523.32.4.885

He, C., & Chen, A. (2010). Interference from familiar natural distractors is not eliminated by high perceptual load. *Psychological Research PRPF*, *74*(3), 268–276. https://doi.org/10.1007/s00426-009-0252-0

Huffman, G., Al-Aidroos, N., & Pratt, J. (2017). Salience drives non-spatial feature repetition effects in cueing tasks. *Attention, Perception, & Psychophysics*, *79*(1), 212–222. https://doi.org/10.3758/s13414-016-1215-3

Jenkins, R., Lavie, N., & Driver, J. (2003). Ignoring famous faces: Category-specific dilution of distractor interference. *Perception & Psychophysics*, *65*(2), 298–309. https://doi.org/10.3758/BF03194801

Jenkins, R., Lavie, N., & Driver, J. (2005). Recognition memory for distractor faces depends on attentional load at exposure. *Psychonomic Bulletin & Review*, *12*, 314–320. https://doi.org/10.3758/BF03196378

Jonides, J., & Irwin, D. E. (1981). Capturing attention. *Cognition*, *10*(1–3), 145–150. https://doi.org/10.1016/0010-0277(81)90038-x

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

Kennedy, J. E. (2024). Planning Falsifiable Confirmatory Research. *Psychological Methods*. https://doi.org/10.1037/met0000639

Kim, J., Kang, M.-S., Cho, Y. S., & Lee, S.-H. (2017). Prolonged Interruption of Cognitive Control of Conflict Processing Over Human Faces by Task-Irrelevant Emotion Expression. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01024

Klein, R. M., Wang, Y., Dukewich, K. R., He, S., & Hu, K. (2015). On the costs and benefits of repeating a nonspatial feature in an exogenous spatial cuing paradigm. *Attention,* *Perception, & Psychophysics*, *77*(7), 2293–2304. https://doi.org/10.3758/s13414-015-0941-2

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, *111*(4), 880–913. https://doi.org/10.1037/0033-295X.111.4.880

Lachter, J., Remington, R., & Ruthruff, E. (2009). Space, object, and task selection. *Attention, Perception & Psychophysics*, *71*, 995–1014. https://doi.org/10.3758/APP.71.5.995

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451. https://doi.org/10.1037//0096-1523.21.3.451.

Lavie, N., & Cox, S. (1997). On the efficiency of visual selective attention: efficient visual search leads to inefficient distractor rejection. *Psychological Science*, *8*(5), 395-396. https://doi.org/10.1111/j.1467-9280.1997.tb00432.x

Lavie, N., & Fox, E. (2000). The role of perceptual load in negative priming. *Journal of Experimental Psychology: Human Perception and Performance, 26*(3), 1038–1052. https://doi.org/10.1037/0096-1523.26.3.1038

Lavie, N., & De Fockert, J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception & Psychophysics*, *65*(2), 202–212. https://doi.org/10.3758/BF03194795

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1346–1358. https://doi.org/10.1037/a0016454

Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological Science*, *14*(5), 510–515. https://doi.org/10.1111/1467-9280.03453

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197. https://doi.org/10.3758/BF03213897

Lee, E., Kang, J. I., Park, I. H., Kim, J.-J., & An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Research*, *157*(1), 77–85. https://doi.org/10.1016/j.psychres.2007.02.005

Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W., & Theeuwes, J. (2021). Progress toward resolving the attentional capture debate. *Visual Cognition*, *29*(1), 1–21. https://doi.org/10.1080/13506285.2020.1848949

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Macdonald, J. S. P., & Lavie, N. (2008). Load induced blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1078–1091. https://doi.org/10.1037/0096-1523.34.5.1078

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*(1), 15–20. https://doi.org/10.1037/0021-843X.95.1.15

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate Cortex. *Science*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Moriya, J., Koster, E. H., & De Raedt, R. (2014). The influence of working memory on the anger superiority effect. *Cognition and Emotion*, *28*(8), 1449-1464. https://doi.org/10.1080/02699931.2014.890094

Müller-Bardorff, M., Schulz, C., Peterburs, J., Bruchmann, M., Mothes-Lasch, M., Miltner, W., & Straube, T. (2016). Effects of emotional intensity under perceptual load: An event-related potentials (ERPs) study. *Biological Psychology*, *117*, 141–149. https://doi.org/10.1016/j.biopsycho.2016.03.006

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin & Review*, *23*(5), 1316–1340. https://doi.org/10.3758/s13423-015-0982-5

Neumann, M. F., Mohamed, T. N., & Schweinberger, S. R. (2011). Face and object encoding under perceptual load: ERP evidence. *NeuroImage*, *54*(4), 3021–3027. https://doi.org/10.1016/j.neuroimage.2010.10.075

Neumann, M. F., & Schweinberger, S. R. (2008). N250r and N400 ERP correlates of immediate famous face repetition are independent of perceptual load. *Brain Research*, *1239*, 181–190. https://doi.org/10.1016/j.brainres.2008.08.039

Norberg, J., Peira, N., & Wiens, S. (2010). Never mind the spider: late positive potentials to phobic threat at fixation are unaffected by perceptual load. *Psychophysiology*, *47*(6), 1151-1158. https://doi.org/10.1111/j.1469-8986.2010.01019.x

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381–396. https://doi.org/10.1037/0022-3514.80.3.381

Okon-Singer, H., Tzelgov, J., & Henik, A. (2007). Distinguishing between automaticity and attention in the processing of emotionally significant stimuli. *Emotion*, *7*(1), 147–157. https://doi.org/10.1037/1528-3542.7.1.147

Pandey, S., & Gupta, R. (2022). Irrelevant positive emotional information facilitates response inhibition only under a high perceptual load. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-17736-5

Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. *Attention*, 155-189.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, *99*(17), 11458–11463. https://doi.org/10.1073/pnas.172403899

Pessoa, L., Padmala, S., & Morland, T. (2005). Fate of unattended fearful faces in the amygdala is determined by both attentional resources and cognitive modulation. *NeuroImage*, *28*(1), 249–255. https://doi.org/10.1016/j.neuroimage.2005.05.048

Pinsk, M. A., Doniger, G. M., & Kastner, S. (2004). Push-Pull Mechanism of Selective Attention in Human Extrastriate Cortex. *Journal of Neurophysiology*, *92*(1), 622–629. https://doi.org/10.1152/jn.00974.2003

Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Remington, R. W., & Folk, C. L. (2001). A dissociation between attention and selection. *Psychological Science*, *12*(6), 511–515. https://doi.org/10.1111/1467-9280.00394

Rubin, M. (2021). When to adjust alpha during multiple testing: a consideration of disjunction, conjunction, and individual testing. *Synthese*, *199*(3-4), 10969-11000. https://doi.org/10.1007/s11229-021-03276-4

Sadeh, N., & Bredemeier, K. (2011). Individual differences at high perceptual load: The relation between trait anxiety and selective attention. *Cognition and Emotion*, *25*(4), 747–755. https://doi.org/10.1080/02699931.2010.500566

Sato, S., & Kawahara, J. I. (2015). Attentional capture by completely task-irrelevant faces. *Psychological Research*, *79*(4), 523–533. https://doi.org/10.1007/s00426-014-0599-8

Schindler, S., Tirloni, C., Bruchmann, M., & Straube, T. (2021). Face and emotional expression processing under continuous perceptual load tasks: An ERP study. *Biological Psychology*, *161*, 108056. https://doi.org/10.1016/j.biopsycho.2021.108056

Scoboria, A., Wade, K. A., Lindsay, D. S., Azad, T., Strange, D., Ost, J., & Hyman, I. E. (2017). A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies. *Memory*, *25*(2), 146–163. https://doi.org/10.1080/09658211.2016.1260747

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Simons, D. J., & Chabris, C. F. (1999). Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception*, *28*(9), 1059-1074. https://doi.org/10.1068/p281059

Spielberger, C. D., Gorsuch, R., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press.

Tannert, S., & Rothermund, K. (2020). Attending to emotional faces in the flanker task: Probably much less automatic than previously assumed. *Emotion*, *20*(2), 217–235. https://doi.org/10.1037/emo0000538

Tipples, J., & Sharma, D. (2000). Orienting to exogenous cues and attentional bias to affective pictures reflect separate processes. *British Journal of Psychology*, *91*(1), 87-97. https://doi.org/10.1348/000712600161691

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. https://doi.org/10.1037/h0027242

Tsal, Y., & Benoni, H. (2010). Diluting the Burden of Load: Perceptual Load Effects Are Simply Dilution Effects. *Journal of Experimental Psychology. Human Perception and Performance*, *36*, 1645–1656. https://doi.org/10.1037/a0018172

Tukey, J.W. (1977). Exploratory Data Analysis. Reading, MA: AddisonWesley Pub. Co.

Vuilleumier, P., & Righart, R. (2011). Attention and automaticity in processing facial expressions. *Oxford handbook of face perception*, 449-478. https://doi.org/10.1093/oxfordhb/9780199559053.013.0023

Wang, L., Feng, C., Mai, X., Jia, L., Zhu, X., Luo, W., & Luo, Y. (2016). The Impact of Perceptual Load on the Non-Conscious Processing of Fearful Faces. *PLOS ONE*, *11*(5), e0154914. https://doi.org/10.1371/journal.pone.0154914

Wentura, D., & Rothermund, K. (2003). The "Meddling-In" of Affective Information: A General Model of Automatic Evaluation Effects. *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, 51–86.

Wilcox, R. R. (2013). *Introduction to robust estimation and hypothesis testing (3rd ed.)*. Academic Press. https://doi.org/10.1016/C2010-0-67044-1

Wirth, B. E., & Wentura, D. (2018). Furious snarling: Teeth-exposure and anxiety-related attentional bias towards angry faces. *PLOS ONE*, *13*(11), e0207695. https://doi.org/10.1371/journal.pone.0207695

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. *Cognition and 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), nsad070. https://doi.org/10.1093/scan/nsad070

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Yi, D.-J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: Dissociable effects of perceptual and working memory load. *Nature Neuroscience*, *7*(9), 992–996. https://doi.org/10.1038/nn1294

Zhang, J., Bürkner, P.-C., Kiesel, A., & Dignath, D. (2023). How emotional stimuli modulate cognitive control: A meta-analytic review of studies with conflict tasks. *Psychological Bulletin*, *149*(1–2), 25–66. https://doi.org/10.1037/bul0000389

Zivony, A., & Lamy, D. (2016). The role of dimension relevance in features' access to response-selection mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1873–1885. https://doi.org/10.1037/xhp0000289

Appendix A

Table A1

The Target Names

Female Names	Male Names
Amelia	Alexander
Charlotte	Andrew
Elizabeth	Anthony
Evelyn	Lucas
Hannah	Oliver
Madison	Robert
Olivia	Rupert
Sophia	Theodore

Appendix B

Analyses with RT as the Dependent Variable

In the *Results* sections of the main text we analyzed flanker differences as the dependent variable to provide a condensed report of the main results. For the sake of transparency, we report here the analyses with RTs as the dependent variable.

Pilot Study

A 4 (load) × 2 (congruency) repeated measures MANOVA with mean RTs as dependent variable yielded the effects reported in Table B1. The corresponding descriptive values were given in Table 1. As can be seen there is (additionally to the effects reported in the main text) a significant load effect, which only indicates the rather trivial increase of RTs with increasing load

Table B1

Results of a 4 (load) × 2 (congruency) repeated measures MANOVA with RT as the dependent variable (Pilot Study)

Effect	$F(df_Z, 66-df_Z)$	df_Z	p	η_p^2
Load (L)	873.07	3	< .001	.977
Congruency (C)	13.41	1	< .001	.171
L × C	.46	3	.714	.021

Note. The last two rows correspond to the effects reported in the main text (overall congruency effect and Load effect of the repeated-measures MANOVA with flanker differences as the dependent variable)

Experiment 1

A 2 (face type) × 4 (load) × 2 (congruency) repeated measures MANOVA with mean RTs as dependent variable yielded the effects reported in Table B2. The corresponding descriptive values were given in Table 2. As can be seen there is (additionally to the effects reported in the main text) the (trivial) significant effect of load. The table additionally shows the main effect of faced type and the interaction of load and face type (which are both not significant).

Table B2

Results of a 2 (face type) × 4 (load) × 2 (congruency) repeated measures MANOVA with RT as the dependent variable (Experiment 1)

Effect	$F(df_Z,144\text{-}df_Z)$	df_Z	*p*	η_p^2
Load (L)	1414.30	3	< .001	.968
Congruency (C)	22.85	1	< .001	.138
Face type (F)	.33	1	.569	.002
L × F	.62	3	.600	.013
L × C	4.03	3	.009	.079
F × C	.03	1	.863	.000
L × C × F	1.60	3	.193	.033

Note. All effects including Congruency (C) correspond to the effects reported in the main text.

Experiment 2

A 2 (face type) × 4 (load) × 2 (congruency) repeated measures MANOVA with mean RTs as dependent variable yielded the effects reported in Table B3. The corresponding descriptive values were given in Table 3. As can be seen there are (additionally to the effects reported in the main text) a (trivial) significant main effect of load.

Table B3
Results of a 2 (face type) × 4 (load) × 2 (congruency) repeated measures MANOVA with RT as the dependent variable (Experiment 1)

Effect	$F(df_Z,180\text{-}df_Z)$	df_Z	P	η_p^2
Load (L)	2620.62	3	< .001	.978
Congruency (C)	82.22	1	< .001	.315
Face type (F)	.89	1	.346	.005
L × F	.53	3	.663	.009
L × C	1.68	3	.173	.028
F × C	1.56	1	.214	.009
L × C × F	2.87	3	.038	.046

Note. All effects, including Congruency (C), correspond to the effects reported in the main text.

Experiment 3

A 2 (face type) × 4 (load) × 2 (congruency) × 2 (Emotion Type: happy vs. angry) repeated measures MANOVA with the first three factors as within-participant factors, emotion type as a between-participants factor, and mean RTs as dependent variable yielded the effects reported in Table B4. The corresponding descriptive values were given in Table 4. As in Experiment 2 there was additionally only the (trivial) significant main effect of load.

Table B4
Results of a 2 (face type) × 4 (load) × 2 (congruency) as within × 2 (emotion type: happy vs. angry) as between, repeated measures MANOVA with mean RTs as dependent variable with RT as the dependent variable (Experiment 3)

Effect	$F(df_Z,365\text{-}df_Z)$	df_Z	P	η_p^2
Load (L)	4545.03	3	< .001	.974
Congruency (C)	93.30	1	< .001	.204
Face type (F)	.48	1	.489	.001
L × F	1.09	3	.354	.009
Emotion Type	.92	1	.337	.003
L × C	5.99	3	< .001	.05
F × C	1.00	1	.318	.003
L × C × F	.57	3	.638	.005

Note. All effects, including Congruency (C), correspond to the effects reported in the main text.

Appendix C

The F-test for a linear trend is equivalent to a one-sample t-test for a specific weighted sum of the flanker differences for the four load levels (see, e.g., Maxwell & Delaney, 2004):

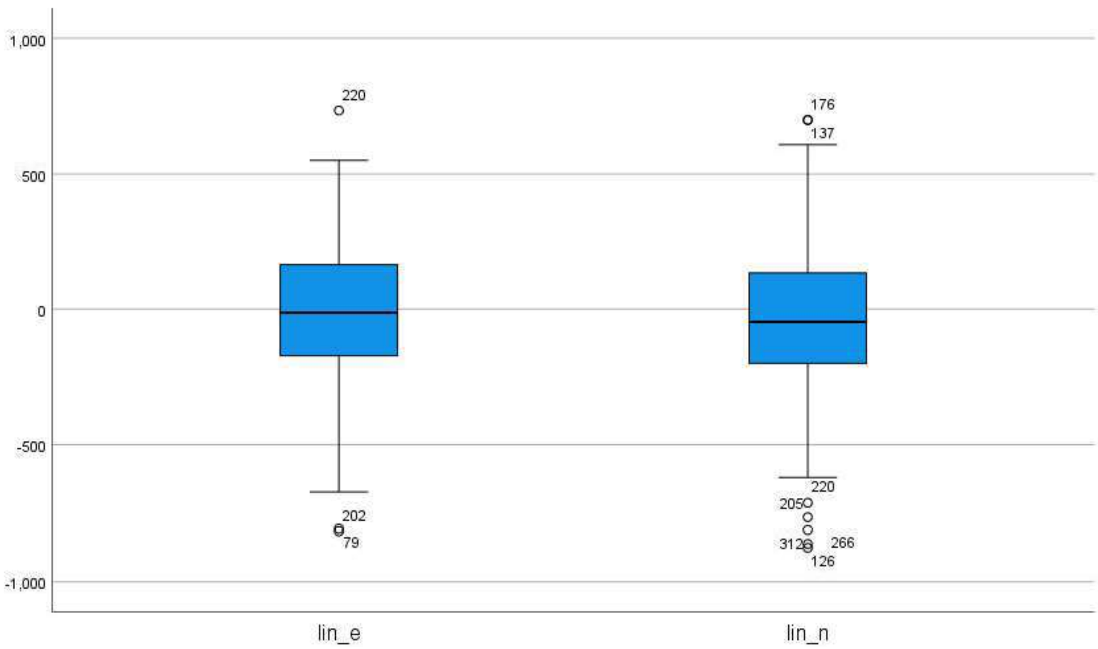
LinT = (-1.5)•load2 + (-0.5)•load4 + (0.5)•load6 + (1.5)•load8

with "loadx" for "flanker difference at load level x"

Therefore, the F-test for the Face Type × Linear Trend interaction is equivalent to a t-test for dependent measures comparing *LinT* for angry and neutral faces. Figure A1 shows the linear trends for emotional (*lin_e*) and neutral (*lin_n*) distractors. As can be seen, these variables are plagued by outliers at both tails of the distribution.

Figure C1

Boxplots for the Linear Trend Variables for Emotional (lin_e) and Neutral (lin_n) Distractors



Robust one-sample t-tests (yuen.t.test function from the R package PairedData; Champely, 2018; see Wilcox, 2013, with regard to robust testing) with a trimming of $\gamma = .25$, which yielded $|t| < 1$ for the linear trend for emotional distractors, and $t(183) = 2.70$, $p = .007$, $d_Z = 0.15$ for the linear trend for neutral distractors. A robust t-test for dependent measures (yuend function from the R package WRS2; $\gamma = .25$; Mair & Wilcox, 2020) yielded $t(183) = 1.65$, $p = .05$ (one-tailed), $d_{Z'} = 0.10$ for the difference.

Appendix D

In Experiment 3, for exploratory reasons we added the trait scale from the State-Trait Anxiety Inventory (STAI; Laux, Glanzmann, Schaffner, & Spielberger, 1981). To keep the analyses concise, we conducted two 2 (Face Type: emotional vs. neutral) × 4 (Load: 2 vs. 4 vs. 6 vs. 8) repeated-measures MANOVAs with flanker differences as the dependent variable, now by adding (*z*-standardized) STAI scores as a covariate, one for the sample with angry faces and one for the sample with happy faces. For the sample that saw happy faces, no effects involving the STAI scores were significant, all *F*s < 1. For the sample that saw angry faces, no effects involving STAI scores were significant, all *F*s < 1, except $F(3, 174) = 1.66$, $p = .177$, $\eta p^2 = .028$ for the load × STAI interaction, and $F(1, 176) = 2.24$, $p = .136$, $\eta p^2 = .013$ for the main effect of STAI,. Thus, the average flanker effect—across load levels and irrespective of whether the distractor was angry or neutral—showed a tendency toward a positive correlation with STAI. In fact, a rank correlation (Spearman's ρ) – which might be more appropriate given the distribution of the flanker variable – was significant ($r = .14$, $p = .027$, one-tailed; see Wirth & Wentura, 2018).