Journal of Articles in Support of the Null Hypothesis
Vol. 19, No. 1

www.jasnh.com

No Emotional Stroop Effect from Masked and Unmasked Stimuli in Non-Clinical Young Adults Tested Online

Gary D. Fisk
Department of Psychology and Sociology, Georgia Southwestern State University

Steven J. Haase
Department of Psychology, Shippensburg University

Emotional Stroop effects ("murder" slows color identification) produced by briefly presented, pattern-masked word stimuli have not replicated well. A masked emotional Stroop effect study was conducted with an improved, more informative experimental design. In Experiment 1, neutral or emotional words were presented in color (red, yellow, green, or blue) for a short (21 ms), medium (38 ms), or long (103 ms) duration. The prime words in Experiment 1 were backward-masked to decrease visibility, whereas Experiment 2 had unmasked words that were easily visible. Neither experiment produced the anticipated emotional Stroop effect. The emotional Stroop effect may be an example of fragile data: a psychological phenomenon that depends upon very narrow experimental parameters.

Keywords: emotional Stroop effect, subliminal, unconscious perception, backward masking, attentional bias

Author Note:
Gary D. Fisk: https://orcid.org/0000-0001-6203-0713
Steven J. Haase: https://orcid.org/0000-0002-8935-2111
Web site with data, analyses, and supporting materials: https://osf.io/jp43t/

Correspondence should be addressed to Gary Fisk, Professor of Psychology, Department of Psychology and Sociology, Georgia Southwestern State University, 800 GSW Drive, Americus, GA 31709; Email: gary.fisk@gsw.edu; phone 229-931-2296; fax 229-931-2315

The emotional Stroop effect occurs when emotional stimuli—typically words with a negative emotional valence—slow responding on a color identification task. For example, a trial with the word failure displayed in a red font might delay the reporting of a "red" color choice compared to a neutral word. This emotional interference is heightened in people with emotional disorders, thereby making the emotional Stroop task an important research tool for studying a wide range of psychological disorders (for reviews, see Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007; Cisler et al., 2011; Phaf & Kan, 2007; Williams, Mathews, & MacLeod, 1996). Increased sensitivity to emotional stimuli, particularly in people with emotional disorders, might be attributed to a hypervigiliance or attentional bias towards emotional content (Williams et al., 1996). An alternative interpretation is that emotional Stroop effects are caused by a generalized slowing of responses to emotional stimuli (Algom, Chajut, & Lev, 2004).[1]

A subset of the emotional Stroop effect literature is directed towards uncovering a potentially unconscious or automatic mechanism responsible for producing the interference effect. The display technology in these studies employs brief (about 16 to 20 ms) prime word stimuli, followed by a masking stimulus to prevent awareness of the prime word. The participant responds with a color choice. To illustrate, a typical trial from these studies might display an emotional word (example: failure) for 20 ms. This brief prime stimulus is immediately followed by a pattern masking stimulus in the same color and location as the emotional word (example: XDMGNGQ; see MacLeod & Rutherford, 1992) with the masking stimulus severely decreasing the visibility of the prime word.[2] Participants respond by reporting the stimulus color (red in this example), not the prime word meaning. The original studies with masked prime stimuli showed an emotional Stroop effect even though the participants were seemingly unaware of the masked primes, suggesting that emotional Stroop effects originated from an unconscious or automatic process (MacLeod & Hagan, 1992; MacLeod & Rutherford, 1992; Van Den Hout, Tenney, Huygens, Merckelbach, & Kindt, 1995). However, numerous later studies with similar designs and methodologies showed little or no effect. A non-systematic review suggests that these studies are about equally divided between positive (Bradley, Mogg, Millar, & White, 1995; Dejonckheere, Braet, & Soetens, 2003; Jansson & Lundh, 2006; Kyrios & Iob, 1998; Leventhal et al., 2008; Lundh, Wikström, Westerlund, & Öst, 1999; MacLeod & Hagan, 1992; MacLeod & Rutherford, 1992; Mogg, Bradley, Williams, & Mathews, 1993; Mogg, Kentish, & Bradley, 1993; Pury, 2002; Putman, Hermans, & van Honk, 2004; Van Den Hout et al., 1995; Van Honk, Tuiten, de Haan, van den Hout, & Stam, 2001; Wikström, Lundh, & Westerlund, 2003) and negative findings (Arntz, Appels, & Sieswerda, 2000; Egloff & Hock, 2003; Franken, Kroon, Wiers, & Jansen, 2000; Jansen, Huygens, & Tenney, 1998; Kampman, Keijsers, Verbraak, Näring, & Hoogduin, 2002; Mahmoudian, Alborzi, Kazemian, Mahmoudian, & Barzegar, 2014; Mogg & Bradley, 2002; Munafò, Mogg, Roberts, Bradley, & Murphy, 2003; Sackville, Schotte, Touyz, Griffiths, & Beumont, 1998; Wikström, Lundh, Westerlund, & Högman, 2004). The frequency of published studies on masked, emotional Stroop effect effects declined after two influential meta-analytic reviews raised doubts about effects from masked stimuli (Bar-Haim et al., 2007; Phaf & Kan, 2007).

Although the previous findings of masked emotional Stroop studies are inconsistent, these ambiguities should be interpreted within a wider perspective of experimental psychology research aimed at isolating unconscious perceptual processes. The broad range of research on unconscious perception and subliminal influence has a long history of controversies and failed replications (for classic critical reviews, see Eriksen, 1960; Holender, 1986). A more recent survey of investigators in this field suggests that most researchers believe that unconscious processing exists (94%), yet only about one-third feel that the empirical evidence shows strong support (Peters & Lau, 2015). This plausibility of unconscious processing accompanied by controversial and sometimes negative empirical evidence has led to an emphasis on methodological issues, improved statistical methods (e.g., Bayesian), and new masking technologies (e.g., continuous flash suppression). Given this context, we feel that the 1990s methodology of masked emotional Stroop effects deserves a reconsideration, with the hope that more robust research findings could be possible with better methodology.

Our interest in masked emotional Stroop effects was sparked by the discrepancy between an emotional Stroop literature with positive findings (see above) and the findings from studies of masked traditional Stroop stimuli (e.g., BLUE; color words, not emotional words). In general, Stroop experiments with masked color word stimuli do not produce a Stroop effect at stringently set display settings that eliminate all behavioral evidence of prime stimulus sensitivity (Cheesman & Merikle, 1984; Kouider & Dupoux, 2004; Lorentz et al., 2015; Sand, 2016; Sand & Nilsson, 2016; Severance & Dyer, 1973; Tzelgov, Porat, & Henik, 1997; Van den Bussche et al., 2013). The traditional Stroop effect size for near-threshold display presentations increases with increasing prime visibility, suggesting a dose-response relationship (Fisk & Haase, 2020a, 2020b).

The differences between the masked emotional Stroop and classic Stroop studies can be illustrated by comparing influential studies in each area conducted by different investigators. A prominent classic Stroop study by Cheesman & Merikle (1984) utilized a tachistoscope to briefly present classic Stroop stimulus

1 Algom et al. (2004) have argued that emotional Stroop effects are a "generic slowdown" (p. 323) that does not require selective attention, and thus should not be called a Stroop effect. This argument may have merit. We use the term emotional Stroop effect in this paper to stay consistent with the broader literature on this topic.

2 Previous studies often describe very brief, masked prime stimulus presentations as being "subliminal." A problem with this description is the word subliminal can have multiple meanings, such as being a synonym for unconscious or having energy levels below a psychophysical threshold (Lundh, Wikström, Westerlund, & Öst, 1999). To decrease confusion, we have chosen to avoid using the word subliminal. We refer to these stimuli as masked stimuli to emphasize their technical display characteristics rather than psychological characteristics.

words (blue, green, yellow, and orange) and a pattern mask of XXXXXX characters. The duration of the prime word stimuli was determined for each participant prior to experimental trials with threshold testing methods. This threshold setting procedure started with stimulus durations at 100 ms and decreased progressively to the point where the participants could no longer show behavioral sensitivity to the stimuli—an objective threshold with 25% accuracy (i.e., choosing one out of four color words correctly is random performance). The color word prime stimuli were presented at three durations: just below each participant's objective threshold, at an intermediate level with 55% detection accuracy, and a duration with 90% or greater accuracy. The results, in brief, showed no Stroop effect at the shortest objective threshold durations, with increasing Stroop effects at the intermediate and longest duration settings (see also Fisk & Haase, 2020a). The obtained effects at the intermediate duration levels were described as subjective threshold effects, at which the participants showed some behavioral sensitivity to the stimuli yet would typically deny stimulus visibility in a verbal report. In the emotional Stroop field, MacLeod and colleagues conducted two studies that were similar to Cheesman and Merikle: briefly presented words (emotional or neutral) followed by a masking stimulus (MacLeod & Hagan, 1992; MacLeod & Rutherford, 1992). A critical difference is that all of the masked priming stimuli were presented at very brief durations of 20 ms on a computer display. Although threshold testing was not conducted, this brief duration setting would likely correspond closely to the objective threshold display conditions of Cheesman and Merikle. MacLeod and colleagues found that emotional Stroop effects were elicited with the prime stimuli at a very stringent display setting, unlike the findings of Cheesman and Merikle. In summary, the critical issue is that classic Stroop effects, which are very robust effects, cannot be obtained with stimuli presented at an objective threshold (Cheesman & Merikle, 1984), yet emotional Stroop effects may possibly occur at this very stringent display setting (MacLeod & Hagan, 1992; MacLeod & Rutherford, 1992). This intriguing discrepancy between two very similar research approaches has not been closely examined. Typically, the classic Stroop task research literature focuses on classic Stroop task findings and a similar pattern occurs in the emotional Stroop literature. In addition, it has been proposed that perception of stimuli at an objective threshold, like the masked emotional Stroop studies, would be compelling evidence of unconscious perception (Snodgrass, Bernat, & Shevrin, 2004), but this evidence has been difficult to reliably demonstrate.

A plausible explanation for the discrepancy between emotional and traditional Stroop study findings is that emotional stimuli may receive preferential processing over neutral stimuli. Evidence for preferential processing has been obtained from emotional stimuli such as angry faces and snakes, possibly reflecting an evolutionary preparation to deal with threatening situations (Capitão et al., 2014; Dimberg, Thunberg, & Elmehed, 2000; Morris, Öhman, & Dolan, 1998; Öhman, Flykt, & Esteves, 2001; Yang, Zald, & Blake, 2007). This raises the possibility that emotional Stroop and traditional Stroop rely on different mechanisms, with emotional Stroop being more likely to reflect unconscious processing due to having a higher processing priority.

Our investigation into the differences between the traditional Stroop and emotional Stroop study findings began with the possibility that seemingly trivial differences in display parameter settings might explain the discrepancy. Using the display parameters from masked emotional Stroop studies with traditional Stroop words (e.g., red presented from 18 to 108 ms, then XDFWPLM as a masking stimulus) did not yield Stroop effects at brief durations, but did yield robust effects at longer durations with greater prime visibility (Fisk & Haase, 2020b). These findings were similar to previous studies performed with traditional Stroop stimuli, thereby ruling out display differences as a technical explanation for the discrepancy. In a second study, we compared an experimental design commonly used in masked emotional Stroop studies (brief, immediately masked primes vs. an unmasked condition; a binary, all-or-nothing design) to an experiment with varied degrees of prime visibility (multiple prime duration conditions with masking stimuli on every trial). Traditional Stroop color word stimuli were used. This experimental design idea was based, in part, on the realization that the experimental designs from masked emotional Stroop studies are not fully factorial. We concluded that the design with different degrees of prime visibility was more informative and interpretable than the binary design (masked vs. unmasked trials) commonly used in masked emotional Stroop studies (Fisk & Haase, 2020a).

The present study extends this previous work (Fisk & Haase, 2020a, 2020b) towards the question of why masked emotional Stroop studies sometimes fail to produce strong unconscious processing evidence. This is a conceptual replication: Can a robust and reliable masked emotional Stroop effect be obtained with an improved experimental design? Experiment 1 was based on the idea that varying the degree of prime stimulus visibility in an emotional Stroop paradigm would be more sensitive to unconscious processes than previous studies based upon the view that consciousness is a strictly conscious/unconscious dichotomy (binary experimental designs). The resulting experiment was like Experiment 2 of Fisk and Haase (2020a): three levels of prime stimulus duration (short, medium, and long) with masking stimuli on each trial (see also Cheesman & Merikle, 1984). The short duration condition (approximately 21 ms) was chosen to correspond closely to the display parameters used in most emotional Stroop studies with masked stimuli. In both experiential and sensitivity terms, these very brief stimuli are very difficult to perceive due to having identification *d'* values near zero (Fisk & Haase, 2011). In contrast, the medium duration (approximately 38 ms) and long duration conditions (approximately 103 ms) have greater prime stimulus visibility. These correspond, respectively, to values commonly used in masked priming studies (e.g. mean of 42 ms from a meta-analysis; Van den Bussche, Van den Noortgate, & Reynvoet, 2009) and an upper point where conscious perception of the stimulus is likely to occur on most trials. Thus, a range of prime visibility values are covered, which may yield results that are more interpretable than the traditional all-or-nothing approach to stimulus visibility (Fisk & Haase, 2020a). An important difference from the previous studies was the comparison of emotional words (negative valence) vs. neutral words instead of using traditional Stroop color words that are congruent or incongruent to the font color. An additional strength of the experimental design, which has masking stimuli on every trial, is being fully factorial compared to the designs commonly used in earlier studies. Experiment 2 was conducted with unmasked stimuli that remained on the display until a response was made or a response time limit was reached. This similar to the unmasked, within-subject conditions of previous studies and provides a reference point for interpreting the results from Experiment 1.

A secondary hypothesis was that the results might partially depend upon the emotional characteristics of the participants (e.g., MacLeod & Hagan, 1992; Pury, 2002). To address this possibility, the non-clinical sample was also assessed on the Scale of Positive and Negative Experiences (SPANE; Diener et al., 2009), a brief self-report of recently experienced positive or negative emotions. An individual differences hypothesis was that people scoring high in negative emotional states might have a larger emotional Stroop effect than people with lower levels of negative emotions.

Method

Participants

For Experiment 1, 48 undergraduate students were recruited from lower-level Psychology courses at Georgia Southwestern State University. Three participants were removed from the sample for self-reported color insensitivity, resulting in a final N = 45. The sample was mostly female (80%), young adults (mean age of M = 21, SD = 3.4). For Experiment 2, 51 students were tested, with none being removed for self-reported color insensitivity. One person was removed for having timed-out trials on almost every trial. This sample was also mostly female (80%) and young, M = 20 (SD = 4.9) years old. Ethical research procedures were followed, including IRB approval (GSW IRB #FA20-02XPED), informed consent, and debriefing.

These sample sizes were similar to previous masked emotional Stroop effect studies (N = 31 for MacLeod & Hagan, 1992; N = 47 for MacLeod & Rutherford, 1992; N = 46 for Mogg, Bradley, et al., 1993). These sample sizes were also slightly larger than our previous masked priming studies that obtained robust Stroop effects from traditional Stroop color word stimuli (Ns = 32, 38, 42; Fisk & Haase, 2020a, 2020b).

Hardware, Software, and Materials

The experimental data were collected via PsyToolkit version 3.2.0 (Stoet, 2010, 2017), which is a web-based platform. Online data collection was chosen over a traditional laboratory-based approach due to the need for social distancing during the COVID-19 pandemic. The validity of online testing with masked prime stimuli was recently demonstrated with PsychoPy (a similar online testing platform) for 16, 33, and 50 ms stimuli (Angele, Baciero, Gomez, & Perea, in press). The participants used their own computers (desktop or laptop) to perform the experiment, thereby making the exact computer hardware characteristics, such as display refresh rates, uncertain. The PsyToolkit software programs for the two experiments are available at the data and materials web site.

Some emotional Stroop studies have used small stimulus sets (e.g., 8 emotional words; Van Den Hout et al., 1995) that necessitated presenting each stimulus on multiple trials over the course of the experiment. The interpretation of repeated stimuli in masked priming experiments can be problematic due to possible inter-trial repetition priming effects (Abrams & Greenwald, 2000; Damian, 2001; Peremen, Hilo, & Lamy, 2013). Emotional Stroop research has also shown that repeated use of the same stimuli may lead to an effect decrement over time, possibly due to emotional habituation (McKenna & Sharma, 1995; Experiment 2). Given this issue, stimulus repetition was completely eliminated by using a unique stimulus for each trial. There were 60 emotional words selected from previous emotional Stroop studies (Frings, Englert, Wentura, & Bermeitinger, 2010; Lundh et al., 1999; Uzzaman, 2017; Van Den Hout et al., 1995; Vasey, Daleiden, Williams, & Brown, 1995). Close semantic and/or orthographic neighbors were avoided, such as torture/torment. These emotional words were matched with 60 neutral words from the Corpus of Contemporary English (https://www.english-corpora.org/coca/) that had the same length and a similar frequency of occurrence (Ben-Haim et al., 2016; Larsen, Mercer, & Balota, 2006). A paired t-test of word frequency showed no statistically significant difference between the neutral words and the emotional words. All of the words, both emotional and neutral, were 4 to 8 letters in length. The masking stimuli were 120 randomly-generated 8 character strings of upper-case consonant letters. The practice trial word stimuli—"house," "garden," "pizza," and "post"—were not used for the experimental trials. A complete list of word stimuli and their characteristics is available on the supporting web site.

Masked emotional Stroop studies suggest that clinical patients, such as people with anxiety disorders, may have elevated emotional Stroop effects (Williams et al., 1996). Emotional differences have also been shown in non-clinical samples by median-split comparisons of people with high vs. low anxiety (MacLeod & Rutherford, 1992; Pury, 2002). To address this possibility, the Scale of Positive and Negative Experiences (SPANE) survey (Diener et al., 2009) was given to each participant. The SPANE instructions ask participants to consider their emotional experiences and events from the last four weeks. Participants respond to emotional words (examples: positive, negative, good, bad) with a five-point Likert-scale ranging from 1 representing "Very rarely or never" to ratings of 5 representing "Very often or always." For example, a response to the positive word "good" could be a 4 rating of "often" to indicate feeling good often over the last four weeks. A second example is that people who infrequently feel "sad" could respond to the word "sad" with a 2 rating (rarely). The scores from the positive valence words (e.g., positive, good, happy) or negative valence words (e.g., negative, bad, unpleasant) are summed to form a positive feelings score (SPANE-P) and a negative feelings score (SPANE-N), respectively. The difference between SPANE-P and

SPANE-N is a balance score, SPANE-B. To examine individual differences based on emotional characteristics, the emotional Stroop effect in participants with higher self-reported negative emotions were compared to people with lower anxiety. For both experiments, people with SPANE-N scores of 16 and higher were compared to people with SPANE-N scores below 16.

The display sequence for an example trial is shown in Figure 1. The font type was Courier (a monospaced font for precise letter positioning) presented in an 18-point size on a black background. The sequence began with a fixation screen (500 ms) with white dashes indicating the location where the prime word was about to appear in the center of the display. Fixation was followed by a prime stimulus (short, medium, or long duration; see below). The stimulus colors were either red (RGB values of 255,0,0), yellow (255,255,0), green (0,255,0), or blue (0,0,255). For Experiment 1, the prime stimulus was immediately followed by a pattern mask stimulus of eight upper-case, randomly generated consonant letters. The backward masking stimulus was the same color and presented in the same location as the prime stimulus. The same color was used for both the prime and the masking stimulus to decrease the visibility of the prime stimulus. These display characteristics—a briefly presented prime word followed by a masking stimulus of the same color—are similar to the stimuli used in masked emotional Stroop experiments (for a review, see Fisk & Haase, 2020b). The additional medium and long prime duration conditions was based upon our earlier work with traditional Stroop stimuli that showed how bracketing a range of prime stimulus visibility—difficult to relatively easy—led to results that were more informative and more interpretable (Fisk & Haase, 2020a). In Experiment 2, the word stimuli were unmasked. In other words, the neutral or emotional words were presented and remained on the screen until a color choice response was made or 2000 ms had passed.

A technical challenge with briefly presented, masked stimuli is

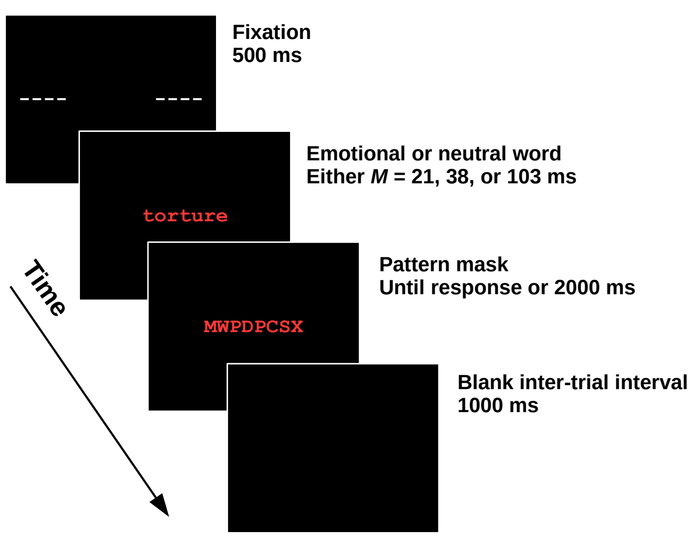

Figure 1. Display sequence characteristics for Experiment 1. For Experiment 2, the word stimulus simply remained on the screen without being followed by a masking stimulus until a response was made.

that the actual display times on the monitor might be longer than the scripted timing values coded into the software. This issue was partly addressed in Experiment 1 with software code to log the time between the onset and offset of the prime stimulus. These logged values represent an internal timing within the PsyToolkit application (not actual display measurements) that should approximately correspond to the display timings. The short condition (coded for 16 ms) had a logged prime duration of $M = 1$ ms ($SD = 4$). Similarly, the medium (coded for 32 ms) and long (coded for 96 ms) conditions had logged duration times of $M = 38$ ($SD = 19$), and $M = 103$ ($SD = 5$) ms, respectively. These results suggest that the actual display times may be about 8 to 16 ms longer (one additional refresh cycle for contemporary computers) than the coded values. For brevity, we will refer to these prime duration conditions as short, medium, and long display times. In Experiment 2, all prime words were displayed until a response was made, which was typically between 300 and 2000 ms.

Design

The standard design of masked emotional Stroop studies compares a masked stimulus condition (e.g., 16 ms prime stimulus, then a same-colored pattern mask until response) to an unmasked condition (e.g., prime word presented until a response). We feel that this design is problematic to interpret and possibly insensitive to marginal awareness (subjective threshold effects), which may possibly lead to failed replications (Fisk & Haase, 2020a, 2020b). To overcome this issue, the prime stimuli for Experiment 1 were presented for three different durations: short, medium, and long (see above). The goal of varying stimulus durations is to bracket a wide range of prime stimulus awareness, such as low or no awareness (short) to likely awareness (long), as well as potentially intermediate, subjective-threshold type states (medium). All of the display conditions utilized pattern–masked stimuli to enable an unambiguous interpretation that might arise from comparing conditions with different numbers of stimuli (Purcell, Stewart, & Stanovich, 1983). This approach stands in contrast to the design of most masked emotional Stroop experiments, in which a masked condition (two stimuli—prime word then masking stimulus) is compared to an unmasked condition (prime word alone; for further details, see Fisk & Haase, 2020a). Another factorial design issue from previous studies is that the unmasked prime stimuli were presented at durations that were not equivalent to the durations of the prime stimuli in the masked condition. This imbalance was avoided in the current design. Altogether, Experiment 1 had six conditions: three levels of prime duration (short, medium, and long) and two levels of word content (emotional and neutral). Each condition had 20 trials. Experiment 2, in contrast, had only unmasked word stimuli, resulting in a single display condition and two levels of word stimuli (emotional and neutral). Each stimulus condition in Experiment 2 had 60 trials.

Testing Protocol

The testing procedure began with an informed consent statement. After consent, the testing question began with survey questions about basic demographics (age, etc.) and then the SPANE survey. The emotional Stroop testing started with instructions that described the appearance of the stimuli (prime stimulus, then mask for Experiment 1; word only for Experiment 2) and the task of choosing the keyboard key that matched the color of the presented stimulus (s = red, d = yellow, k = green, l = blue). A graphic representation showed how the hands should be place on the keyboard, with the index and the middle finger of the right hand on the k (green) and l (blue). The left hand was a mirror opposite: middle finger placed upon the s key (red) and index finger placed upon d (yellow). Instructions were followed by a block of practice trials to familiarize the participants with the color − key correspondence. The prime stimuli for the practice trials of Experiment 1 were presented for 80 ms (coded value). Key pressing practice continued until participants had six correct responses in a row or a total of 24 practice trials were completed. Error messages were given for feedback on incorrect practice trials to encourage accurate responding. After practice, five blocks of 24 experimental trials were given, resulting in a total of 120 experimental trials. The task during the experimental trials was to identify the color of the prime and the masking stimuli (see Figure 1) of red, yellow, green, or blue by pressing the s, d, k, or l keys, respectively. The meaning of the prime word stimulus was irrelevant to the color choice task. Like the practice trials, an error feedback message was given when incorrect color choices occurred (1000 ms in duration). Break messages in between the blocks informed the participants of their progress and encouraged them to take a brief rest. A new block of trials was initiated after a break message by pressing the space bar. The entire testing period lasted about 20 minutes. At the end, a debriefing message explained the purpose of the experiment and thanked participants for their participation. Experiment 2 followed a similar procedure.

Data analysis

Before analysis, the response time from stimulus onset for each trial of Experiment 1 was calculated by adding the logged prime duration to the response time recorded from the onset of the masking stimulus. This summation equates all of the prime duration conditions from Experiment 1 in regard to color onset, which starts with the prime stimulus display. Next, the response time data were filtered to remove trials that were impossibly fast (< 250 ms; almost 0%) and inaccurate responses (4.0% mistakes, .9% time out). Non-directional, dependent-samples *t*-tests compared mean response times for the emotional vs. neutral word conditions at each prime stimulus duration (short, medium, and long). The non-directional test choice was based on evidence that emotional stimuli can either facilitate responding (e.g., Capitão et al., 2014) or produce Stroop-like interference, although most studies report interference effects (Williams et al., 1996). A similar comparison of emotional vs. neutral stimulus trials was conducted for Experiment 2.

The negative *t*-test results were followed up with Bayesian analyses to determine the degree of evidence favoring the null hypothesis. The prior value was based upon results from our previous work with traditional Stroop stimuli that had

Table 1. Response time means (standard deviations) in milliseconds are shown for the emotional and neutral stimulus conditions. The differences were calculated as emotional word trials minus neutral word trials, in which positive difference scores represent slower responses in emotional stimulus conditions. All values were rounded to the nearest millisecond.

Experiment	Emotional	Neutral	Mean Difference	SE Difference	95% CI - Lower	95% CI - Upper
Exp. 1 – short	755 (126)	764 (129)	-9	11	-32	14
Exp. 1 – medium	778 (141)	772 (147)	7	11	-16	29
Exp. 1 – long	795 (138)	806 (166)	-11	13	-37	15
Exp. 2	770 (129)	766 (130)	4	7	-11	19

experimental design and prime masking characteristics similar to the present study (Fisk & Haase, 2020a; Experiment 2). The medium prime duration condition from this experiment yielded a small traditional Stroop response time congruency difference of 16 ms, with Cohen's d = .38. Comparable small effects have been reported in emotional Stroop experiments (e.g., Frings et al., 2010; 14 and 20 ms). Accordingly, the prior is described by a Cauchy distribution centered at zero and having a width parameter of .12. This yields an 80% probability of having a .38 effect size.

Most of the statistical analyses were performed with JASP version 0.9.2. Data summaries and a few specialized analyses (e.g., fast vs. slow trial effects) were performed in the LibreOffice Calc spreadsheet application. The power analysis was done with an online power contour estimation tool (https://shiny.york.ac.uk/powercontours/) from the University of York (Baker et al., 2021).

Results

In Experiment 1 (masked word stimuli), there was no evidence for an emotional Stroop effect. The short duration condition that is similar to previous studies had no statistically significant response time difference between the emotional trials and the neutral trials ($t(44) = -.77$, $p = .45$, $d = -.12$). The same non-significant outcome also occurred for the medium ($t(44) = .58$, $p = .56$, $d = .09$) and long duration conditions ($t(44) = -.86$, $p = .40$, $d = -.13$). The medium duration condition showed a trend towards slower responding on emotional trials (M = 7 ms), but the short and the long conditions showed opposite trends of faster responding on emotional trials (M = -9 and -11 ms, respectively; Table 1). The 95% confidence intervals of the effect size included 0 (Table 1). Mean Cohen's d sensitivity based upon individual calculations also produced effect size values near zero: Short $M_{Cohen's\ d} = -.02$, Medium $M_{Cohen's\ d} = .01$, and Long $M_{Cohen's\ d} = -.05$.

Bayesian analyses of Experiment 1 yielded Bayes factors near 1.0 for both the null hypothesis (short $BF_{01} = 1.50$, medium $BF_{01} = 1.59$, and long $BF_{01} = 1.44$) and the alternative hypothesis (short $BF_{10} = .67$, medium $BF_{10} = .63$, and long $BF_{10} = .69$). Thus, the evidence for either hypothesis is inconclusive from a Bayesian perspective. Consistent with this view, the upper 95% confidence interval boundaries included positive values that are the size of the anticipated small emotional Stroop effect (Table 1).

A fine-grained examination of prime display timings showed a small number of trials (3.6%) programmed for the short condition exceeded 30 ms, making these trials more like the medium display condition. Similarly, 0.2% of the trials coded for the medium prime duration exceeded 70 ms. A recoding of the dataset based on these empirically logged timings (not scripted timings) did not change the outcomes of Experiment 1 for the short ($t(44) = -.67$, $p = .50$) or the medium conditions ($t(44) = .60$, $p = .55$).

A secondary hypothesis that predicted larger effects for people who had higher levels of emotionality was also not supported. There were no significant correlations between response time difference scores (emotional trials minus neutral trials) and SPANE-N (short $r(43) = .06$, $p = .68$; medium $r(43) = -.14$, $p = .34$, and long $r(43) = .04$, $p = .81$). Bayesian analyses with a stretched beta prior width value of 1 showed substantial support for the null hypothesis in all three duration conditions: short $BF_{01} = 4.9$, medium $BF_{01} = 3.5$, and long $BF_{01} = 5.2$. In addition, no significant correlations were found between response time difference scores and the SPANE-P or SPANE-B measures (available online). Similarly, a median-split comparison on the SPANE-N variable for Experiment 1 showed significant response time differences for the different prime duration conditions ($F(2, 86) = 10.07$, $p < .001$, $\eta^2_p = .190$), but no other main effects or interactions were significant.

Like Experiment 1, Experiment 2 (unmasked stimuli) showed no significant response time difference between the neutral word and the emotional word conditions, $t(49) = .56$, $p = .58$, $d = .08$. The mean response time difference was 4 ms longer on the emotional trials (Table 1). Mean Cohen's *d* sensitivity calculated from individual scores was also near zero, $M_{Cohen's\ d} = .02$. The Bayesian evidence for both the null hypothesis and the alternative hypothesis were inconclusive, with both Bayes factors near 1 ($BF_{01} = 1.63$; $BF_{10} = .61$). There was also no correlation between response time difference scores (emotional trials minus neutral trials) and SPANE-N, $r(48) = -.01$, $p = .92$. A factorial analysis of variance of stimulus type (neutral vs. emotional) and SPANE-N revealed no significant main effects or interactions.

Additional exploratory analyses included choice accuracy as a dependent variable, testing order effects, ANCOVA with SPANE-N as a covariate, and fast vs. slow trial effects (Frings et al., 2010). Noteworthy findings did not emerge from these analyses. The details are available at the supporting web site.

A post-hoc power analysis was conducted to better understand the failure to obtain an emotional Stroop effect. The "response time" example of Baker and colleagues (2021) was followed. Within and between-subjects standard deviations were calculated for the three conditions of Experiment 1 and for Experiment 2. Power contour estimate models were run with the following values: $N = 50$, trial size k of 40 (Experiment 1 duration conditions) or 120 (Experiment 2), and the anticipated emotional Stroop effect size of 20 ms (Fisk & Haase, 2020a; Frings et al., 2010). The short, medium, and long duration conditions of Experiment 1 had 40%, 40%, and 35% power, respectively. In contrast, Experiment 2 had 80% power due to having more trials than the conditions of Experiment 1. Additional details are available online.

Discussion

Evidence of masked emotional Stroop effects was not obtained, which is consistent with some studies showing little to no masked emotional Stroop effect. However, a strong conclusion about the presence or absence of effects from masked prime stimuli is not possible due to the lack of an emotional Stroop effect in Experiment 2, in which unmasked stimuli also produced no effect. In both experiments, traditional statistical tests produced non-significant findings and small effect sizes near zero. However, the present data are insufficient from a Bayesian perspective to provide strong support for a null hypothesis outcome, thereby making the current results inconclusive. The lack of clear-cut findings is most likely attributable to methodological shortcomings.

A detailed comparison to the effect sizes from previous studies is also informative about the present lack of significant findings. Phaf and Kahn (2007) reviewed over 70 emotional Stroop studies in a meta-analysis. The present Experiment 1 is similar to their category of "suboptimal" (i.e., masked stimuli) with mixed stimulus trials (i.e., emotional and neutral words in the same block) for non-clinical participants. The effects were $r = -.017$ (16 studies of low anxiety, non-clinical participants) and $r = .037$ (11 studies of high anxiety, non-clinical participants; their Table 1). The comparable present condition, the short duration masked stimuli of Experiment 1, had a Cohen's *d* effect size corresponding to an *r* value of -.010 (from individual calculations). Furthermore, we hypothesized that our novel experimental design that included prime stimuli with more energy (i.e., the medium and long durations) might be more sensitive to subtle effects than previous studies with masked stimuli that used exclusively short durations. However, this anticipated finding was not obtained. These additional conditions yielded effect sizes corresponding to *r* values of .005 (medium, from individuals) and -.025 (long, from individuals), which are between the effect sizes reported by Phaf and Kahn for briefly presented, masked stimuli. Altogether, these results from Experiment 1 are similar to Phaf and Kahn's conclusion: "Rather surprisingly, the effect sizes for suboptimal [i.e., masked] presentation—we only had sufficient studies with mixed presentation—were all close to zero and no effect even approached significance." (p. 190).

For "optimal" unmasked stimuli, Phaf and Kahn (2007) report effect sizes of $r = -.11$ (32 studies; low anxiety, non-clinical) to $r = .056$ (20 studies; high anxiety, non-clinical) for mixed trials (neutral and emotional words in the same block). Our Experiment 2 results corresponded to an *r* value of $r = .010$ (mean from individuals). In contrast, Phaf and Kahn noted that studies with unmasked stimuli and blocked trials produced much larger effects, $r = .048$ (low anxiety, non-clinical) and $r = .135$ (high anxiety; non-clinical). This blocking approach may contribute to larger effects because of sustained emotional responses that carry-over between trials in a block. This overall pattern strongly suggests that the decision to use mixed trials in a block may have decreased the ability to find a significant emotional Stroop effect. If sustained effects occur, mixing trial types in a block may be counterproductive because the neutral and emotional effects are basically canceled out inside of a block (for a methodology review, see Ben-Haim et al., 2016). Our decision to use a mixed design was justified given that this is the design used by most masked emotional Stroop effect studies (Phaf & Kan, 2007). The statistical separation of fast vs. slow trial effects (Frings et al., 2010) explored the possibility that sustained emotional effects might produce an influence on subsequent trials in a mixed design, yet it did not yield evidence of a difference. The potential importance of sustained effects suggests the possibility that a blocked approach might yield stronger differences compared to the present mixed trials approach.

A lack of statistical power does not fully explain the absence of an emotional Stroop effect. The sample sizes of N = 44 (Experiment 1) and N = 50 (Experiment 2) were comparable to or slightly larger than previous studies (see Methods). A power analysis based on sample size and trial numbers shows that the conditions in Experiment 1 may have been insufficient to detect an emotional/neutral response difference (about 40% power). In contrast, Experiment 2 had a desirable level of power (about 80%) due to having more trials (120) than the conditions of Experiment 1 (40). In addition, Experiment 2 used more salient, unmasked stimuli that should, theoretically, have resulted in a stronger impact. We also note that our previous studies with traditional color-based Stroop words have yielded robust Stroop effects with comparable sample sizes and trials per condition (Fisk & Haase, 2020a, 2020b). This discrepancy between the present results and the previous results suggests that the emotional Stroop effect is much weaker than the traditional Stroop effect. Altogether, statistical power could have been better for the masked conditions of Experiment 1, but power deficiencies do not explain the lack of significant findings in Experiment 2.

A significant methodological departure from previous studies is the use of an online experimentation platform—PsyToolkit. The methodology is similar to recent online study conducted with PsychoPy that examined potentially automatic emotional Stroop effects in response to unmasked COVID-19 related stimuli (Ypsilanti, Mullings, Hawkins, & Lazuras, 2021). Two experiments did not show an emotional Stroop effect even though their sample sizes were approximately twice (Experiment 1) or three times (Experiment 2) larger than the present samples. This result, in conjunction with the present results, suggests the possibility that online testing may be problematic for finding small response time differences. This could arise due to the higher variability in hardware that occurs when participants use their own devices rather than using a carefully controlled, traditional laboratory-based approach. We cannot rule this possibility out, but our opinion is that the online testing was not the major shortcoming of the present experiments. The accuracy of logged timings for the brief stimulus displays was comparable to what we have seen in past experiments on a single laboratory computer. Similarly, our previous results from a laboratory-based experiment with masked, classic Stroop stimuli produced response time standard deviations from 117 to 137 ms (Experiment 1; Fisk & Haase, 2020a), that were only slightly less variable than the present results (*SD*s of 126 to 166; see Table 1). This previous study had very large classic Stroop effects (83 ms from unmasked stimuli) that easily achieved statistical significance. Perhaps the combination of a smaller emotional Stroop effect (e.g., 20 ms) plus slightly higher variability in response time recordings decreases the ability to find a significant small effect. However, the power analysis results from Experiment 2 (see above) argue against this interpretation.

There are many other possibilities for the inability to find a significant emotional Stroop effect. The first possibility is that the emotional stimuli were inadequate to elicit a strong emotional response in a general sample of young adult, college students. In support of this possibility, Uzzaman (2017) found no emotional Stroop effects in an undergraduate student population with unmasked, emotional words (examples: war, hate; Experiment 2), but did find a stronger effect with taboo word stimuli (examples: shit, bitch; Experiments 3 and 4; see their appendix for a complete stimulus list). Altogether, this suggests an effect might have been obtained in the present experiments with stronger emotional word stimuli. However, the emotional stimuli used in the present study were drawn from previous studies that reported positive findings (the stimuli are available in the supporting web site). It is unclear why the emotional word stimuli would elicit a significant emotional Stroop effect in previous studies yet not elicit a similar effect in the present experiments. Thus, it seems unlikely that insufficient emotional impact adequately explains the negative findings.

A second additional explanation related to stimulus arousal is participant characteristics, such as emotional traits or states. In general, participants with a higher degree of emotional traits tend to show greater emotional Stroop effects from masked stimuli (e.g., MacLeod & Hagan, 1992; MacLeod & Rutherford, 1992) and unmasked stimuli (Phaf & Kan, 2007). Although some studies have compared clinical to non-clinical samples (Bradley et al., 1995; Kyrios & Iob, 1998; Lundh et al., 1999; Mogg, Bradley, et al., 1993), most studies reporting a masked emotional Stroop effect have been based upon non-clinical participant samples (Dejonckheere et al., 2003, 2003; Lundh et al., 1999; MacLeod & Hagan, 1992; MacLeod & Rutherford, 1992; Mogg, Kentish, et al., 1993; Pury, 2002; Putman et al., 2004; Van Den Hout et al., 1995; Van Honk et al., 2001; van Honk et al., 1998; Wikström et al., 2003). In addition, median-split comparisons of participants on self-reported emotional states were conducted like previous studies (MacLeod & Rutherford, 1992; Mogg, Kentish, et al., 1993; Pury, 2002), yet did not yield noteworthy differences between people with high or low degrees of emotionality. The Experiment 1 median SPANE-B (balanced, a combination of negative and positive scales) score was 7, suggesting that the participants were inclined to report more positive emotions. It's possible that this bias towards positive emotions prevented finding increased responses to negative stimuli, possibly through relatively few people with negative emotions (i.e., a restriction of range). Altogether, previous research suggests that it should be quite feasible to obtain a masked emotional Stroop effect with a non-clinical sample of college student participants, especially among those who self-report higher levels of negative emotions.

A third possible issue is word repetition. Most previous studies have used small word sets, with each word presented repeatedly over the course of the experiment. We avoided this repetition due to potential interpretation problems and the possibility of emotional fatigue or habituation over the course of an experiment. Thus, the present strategy of presenting each word once for the entire experiment should improve—not decrease—the size of the emotional Stroop effect. In support of our approach, some studies with positive findings have used a large number of words, such as 200 words (from five categories; Mogg, Bradley, et al., 1993) or 160 words (from four categories; Mogg, Kentish, et al., 1993). Altogether, stimulus repetition seems like an unnecessary factor for obtaining an emotional Stroop effect from masked stimuli.

A fourth issue is response modality. The present study utilized key press responses, like many experimental psychology studies. However, some masked emotional Stroop studies have used spoken responses, like saying the word "green" out loud into a microphone (MacLeod & Hagan, 1992; MacLeod & Rutherford, 1992). Spoken responses might make a difference for finding an emotional Stroop effect.

In closing, the present experiments do not provide strong evidence for or against potentially unconscious processes in masked emotional Stroop effect studies. However, the results may provide some insight into the conditions necessary for eliciting an emotional Stroop effect in non-clinical participants. The present finding may represent "fragile data," a term coined by Uttal (2000, pp. 76-78). Uttal's concern was that the complexity of modern cognitive science experiments—highly specific dependencies on particular stimuli, displays, experimental designs, and populations—posed a significant barrier to drawing strong and replicable conclusions.

> "As I studied and reviewed the "high level" [cognitive] literature, I came to a rather surprising general conclusion. The reliability, durability, and presumably the validity of the data from the sample of experiments with which I was concerned seemed to evaporate. Data, as well as conclusions, seemed to last only for a few issues of the journal in which they had been published before some criticism of it emerged." (p. 77)

Furthermore, the high degree of experimental complexity may lead to "irreproducible or vanishing findings" (p. 77), a view that presaged the recent replication crisis in experimental psychology. The emotional Stroop paradigm may be an example of Uttal's fragile data. Emotional Stroop effects may only be attainable under a rather narrow combination of specific conditions: word content, display parameters, experimental design, response task, and participant emotional state or traits. In particular, we feel that the mixture of emotional and neutral words in a block may be detrimental and recommend that future research should use blocked stimulus presentations (see Ben-Haim et al., 2016). Investigators who work on this paradigm must pay close attention to configuring these variables in order to obtain a measurable effect.

References

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science, 11,* 118–124. https://doi.org/10.1111/1467-9280.00226

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

Angele, B., Baciero, A., Gomez, P., & Perea, M. (in press). Does online masked priming pass the test? The effects of prime exposure duration on masked identity priming. *Behavior Research Methods*. Retrieved from https://www.uv.es/mperea/MaskedPriming_OnlineBRM.pdf

Arntz, A., Appels, C., & Sieswerda, S. (2000). Hypervigilance in borderline disorder: A test with the emotional Stroop paradigm. *Journal of Personality Disorders, 14*(4), 366–373. https://doi.org/10.1521/pedi.2000.14.4.366

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods, 26*(3), 295–314. https://doi.org/10.1037/met0000337

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M., & van IJzendoorn, M. (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

Ben-Haim, M. S., Williams, P., Howard, Z., Mama, Y., Eidels, A., & Algom, D. (2016). The emotional Stroop task: Assessing cognitive performance under exposure to emotional content. *Journal of Visualized Experiments,* 112. https://doi.org/10.3791/53720

Bradley, B. P., Mogg, K., Millar, N., & White, J. (1995). Selective processing of negative information: Effects of clinical anxiety, concurrent depression, and awareness. *Journal of Abnormal Psychology, 104*(3), 532–536. https://doi.org/10.1037/0021-843X.104.3.532

Capitão, L. P., Underdown, S. J. V., Vile, S., Yang, E., Harmer, C. J., & Murphy, S. E. (2014). Anxiety increases breakthrough of threat stimuli in continuous flash suppression. *Emotion, 14*(6), 1027–1036. https://doi.org/10.1037/a0037801

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception and Psychophysics, 36,* 387–395. https://doi.org/10.3758/BF03202793

Cisler, J. M., Wolitzky-Taylor, K. B., Adams, T. G., Babson, K. A., Badour, C. L., & Willems, J. L. (2011). The emotional Stroop task and posttraumatic stress disorder: A meta-analysis. *Clinical Psychology Review, 31*(5), 817–828. https://doi.org/10.1016/j.cpr.2011.03.007

Damian, M. F. (2001). Congruity effects evoked by subliminally presented primes: Automaticity rather than semantic processing. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 154–165. https://doi.org/10.1037/0096-1523.27.1.154

Dejonckheere, P. J. N., Braet, C., & Soetens, B. (2003). Effects of thought suppression on subliminally and supraliminally presented food-related stimuli. *Behaviour Change, 20*(04), 223–230. https://doi.org/10.1375/bech.20.4.223.29379

Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Chu, K. P., Choi, D., & Oishi, S. (2009). New measures of well-being. In E. Diener (Ed.), *Assessing well-being: The collected works of Ed Diener* (pp. 247–266). Spring Science+Business Media.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*(1), 86–89. https://doi.org/10.1111/1467-9280.00221

Egloff, B., & Hock, M. (2003). Assessing attention allocation toward threat-related stimuli: A comparison of the emotional Stroop task and the attentional probe task. *Personality and Individual Differences, 35*(2), 475–483. https://doi.org/10.1016/S0191-8869(02)00265-9

Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *The Psychological Review, 67,* 279–300. https://doi.org/10.1037/h0041622

Fisk, G. D., & Haase, S. J. (2011). The relationship between the objective identification threshold and priming effects does not provide a definitive boundary between conscious and unconscious perceptual processes. *Consciousness and Cognition, 20*(4), 1221–1231. https://doi.org/10.1016/j.concog.2011.03.016

Fisk, G. D., & Haase, S. J. (2020a). Binary vs. Continuous experimental designs for the study of unconscious perceptual processing. *Consciousness and Cognition, 81,* 102933. https://doi.org/10.1016/j.concog.2020.102933

Fisk, G. D., & Haase, S. J. (2020b). Classic Stroop color words produce no Stroop effect when the display characteristics are based upon emotional Stroop studies with subliminal presentations. *Psychological Reports, 123*(4), 1207–1225. https://doi.org/10.1177/0033294119843220

Franken, I. H. A., Kroon, L. Y., Wiers, R. W., & Jansen, A. (2000). Selective cognitive processing of drug cues in heroin dependence. *Journal of Psychopharmacology, 14*(4), 395–400. https://doi.org/10.1177/026988110001400408

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

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9*(1), 1–66. https://doi.org/10.1017/S0140525X00021269

Jansen, A., Huygens, K., & Tenney, N. (1998). No evidence for a selective processing of subliminally presented body words in restrained eaters. *International Journal of Eating Disorders, 24*(4), 435–438. https://doi.org/10.1002/(SICI)1098-108X(199812)24:4<435::AID-EAT11>3.0.CO;2-B

Jansson, B., & Lundh, L. G. (2006). The interactive role of worried mood and trait anxiety in the selective processing of subliminally presented threat words. *Personality and Individual Differences, 41*(7), 1195–1204. https://doi.org/10.1016/j.paid.2006.03.026

Kampman, M., Keijsers, G., Verbraak, M., Näring, G., & Hoogduin, C. (2002). The emotional Stroop: A comparison of panic disorder patients, obsessive–compulsive patients, and normal controls, in two experiments. *Journal of Anxiety Disorders, 16*(4), 425–441. https://doi.org/10.1016/S0887-6185(02)00127-5

Kouider, S., & Dupoux, E. (2004). Partial awareness creates the "illusion" of subliminal semantic priming. *Psychological Science, 15*(2), 75–81. https://doi.org/10.1111/j.0963-7214.2004.01502001.x

Kyrios, M., & Iob, M. A. (1998). Automatic and strategic processing in obsessive-compulsive disorder. *Journal of Anxiety Disorders, 12*(4), 271–292. https://doi.org/10.1016/S0887-6185(98)00015-2

Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion, 6*(1), 62–72. https://doi.org/10.1037/1528-3542.6.1.62

Leventhal, A. M., Waters, A. J., Breitmeyer, B. G., Miller, E. K., Tapia, E., & Li, Y. (2008). Subliminal processing of smoking-related and affective stimuli in tobacco addiction. *Experimental and Clinical Psychopharmacology, 16*(4), 301–312. https://doi.org/10.1037/a0012640

Lorentz, E., Gould, L., Mickleborough, M., Ekstrand, C., Boyer, M., Cheesman, J., & Borowsky, R. (2015). All in one fell Stroop: Examining consciousness thresholds with a multiple response paradigm. *Psychology of Consciousness: Theory, Research, and Practice, 2*(2), 111–133. https://doi.org/10.1037/cns0000049

Lundh, L. G., Wikström, J., Westerlund, J., & Öst, L. G. (1999). Preattentive bias for emotional information in panic disorder with agoraphobia. *Journal of Abnormal Psychology, 108*(2), 222–232. https://doi.org/10.1037/0021-843X.108.2.222

MacLeod, C., & Hagan, R. (1992). Individual differences in the selective processing of threatening information, and emotional responses to a stressful life event. *Behaviour Research and Therapy, 30*(2), 151–161. https://doi.org/10.1016/0005-7967(92)90138-7

MacLeod, C., & Rutherford, E. M. (1992). Anxiety and the selective processing of emotional information: Mediating roles of awareness, trait and state variables, and personal relevance of stimulus materials. *Behaviour Research and Therapy, 30*(5), 479–491. https://doi.org/10.1016/0005-7967(92)90032-C

Mahmoudian, E., Alborzi, S., Kazemian, M., Mahmoudian, Z., & Barzegar, M. (2014). The power of subliminal vs supraliminal versions of the combi emotional Stroop test in measuring attentional bias for test and social anxiety. *European Online Journal of Natural and Social Sciences, 3,* 398–409.

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

Mogg, K., & Bradley, B. P. (2002). Selective processing of smoking-related cues in smokers: Manipulation of deprivation level and comparison of three measures of processing bias. *Journal of Psychopharmacology, 16*(4), 385–392. https://doi.org/10.1177/026988110201600416

Mogg, K., Bradley, B. P., Williams, R., & Mathews, A. (1993). Subliminal processing of emotional information in anxiety and depression. *Journal of Abnormal Psychology, 102*(2), 304–311. https://doi.org/10.1037/0021-843X.102.2.304

Mogg, K., Kentish, J., & Bradley, B. P. (1993). Effects of anxiety and awareness on colour-identification latencies for emotional words. *Behaviour Research and Therapy, 31*(6), 559–567. https://doi.org/10.1016/0005-7967(93)90107-6

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature, 393*(6684), 467–470. https://doi.org/10.1038/30976

Munafò, M., Mogg, K., Roberts, S., Bradley, B. P., & Murphy, M. (2003). Selective processing of smoking-related cues in current smokers, ex-smokers and never-smokers on the modified Stroop task. *Journal of Psychopharmacology, 17*(3), 310–316. https://doi.org/10.1177/02698811030173013

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466–478. https://doi.org/10.1037/0096-3445.130.3.466

Peremen, Z., Hilo, R., & Lamy, D. (2013). Visual consciousness and intertrial feature priming. *Journal of Vision, 13*(5), 1–12. https://doi.org/10.1167/13.5.1

Peters, M. A. K., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *ELife, 4*. https://doi.org/10.7554/eLife.09651

Phaf, R. H., & Kan, K. J. (2007). The automaticity of emotional Stroop: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 38*(2), 184–199. https://doi.org/10.1016/j.jbtep.2006.10.008

Purcell, D. G., Stewart, A. L., & Stanovich, K. E. (1983). Another look at semantic priming without awareness. *Perception and Psychophysics, 34,* 65–71. https://doi.org/10.3758/BF03205897

Pury, C. L. S. (2002). Information-processing predictors of emotional response to stress. *Cognition & Emotion, 16*(5), 667–683. https://doi.org/10.1080/02699930143000400

Putman, P., Hermans, E., & van Honk, J. (2004). Emotional Stroop performance for masked angry faces: It's BAS, not BIS. *Emotion, 4*(3), 305–311. https://doi.org/10.1037/1528-3542.4.3.305

Sackville, T., Schotte, D., Touyz, S., Griffiths, R., & Beumont, P. (1998). Conscious and preconscious processing of food, body weight and shape, and emotion-related words in women with anorexia nervosa. *International Journal of Eating Disorders, 23*(1), 77–82. https://doi.org/10.1002/(SICI)1098-108X(199801)23:1<77::AID-EAT9>3.0.CO;2-Z

Sand, A. (2016). Reversed priming effects may be driven by misperception rather than subliminal processing. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00198

Sand, A., & Nilsson, M. E. (2016). Subliminal or not? Comparing null-hypothesis and Bayesian methods for testing subliminal priming. *Consciousness and Cognition, 44,* 29–40. https://doi.org/10.1016/j.concog.2016.06.012

Severance, L. J., & Dyer, F. N. (1973). Failure of subliminal word presentations to generate interference to color naming. *Journal of Experimental Psychology, 101*(1), 186–189. https://doi.org/10.1037/h0035768

Snodgrass, M., Bernat, E., & Shevrin, H. (2004). Unconscious perception: A model-based approach to method and evidence. *Perception and Psychophysics, 66,* 846–867. https://doi.org/10.3758/BF03194978

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Tzelgov, J., Porat, Z., & Henik, A. (1997). Automaticity and consciousness: Is perceiving the word necessary for reading it? *The American Journal of Psychology, 110*(3), 429–448. https://doi.org/10.2307/1423567

Uttal, W. R. (2000). *The war between mentalism and behaviorism: On the accessibility of mental processes*. Lawrence Erlbaum Associates.

Uzzaman, S. (2017). *The Stroop task redefined: The multiple-item Stroop task* (University of Toronto). University of Toronto, Toronto. Retrieved from https://tspace.library.utoronto.ca/handle/1807/79534

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin, 135*(3), 452–477. https://doi.org/10.1037/a0015329

Van den Bussche, E., Vermeiren, A., Desender, K., Gevers, W., Hughes, G., Verguts, T., & Reynvoet, B. (2013). Disentangling conscious and unconscious processing: A subjective trial-based assessment approach. *Frontiers in Human Neuroscience, 7,* 1–8. https://doi.org/10.3389/fnhum.2013.00769

Van Den Hout, M., Tenney, N., Huygens, K., Merckelbach, H., & Kindt, M. (1995). Responding to subliminal threat cues is related to trait anxiety and emotional vulnerability: A successful replication of MacLeod and Hagan (1992). *Behaviour Research and Therapy, 33*(4), 451–454. https://doi.org/10.1016/0005-7967(94)00062-O

Van Honk, J., Tuiten, A., de Haan, E., van den Hout, M., & Stam, H. (2001). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition and Emotion, 15*(3), 279–297. https://doi.org/10.1080/0269993004200222

van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., de Haan, E., & Verbaten, R. (1998). Baseline salivary cortisol levels and preconscious selective attention for threat. *Psychoneuroendocrinology, 23*(7), 741–747. https://doi.org/10.1016/S0306-4530(98)00047-X

Vasey, M. W., Daleiden, E. L., Williams, L. L., & Brown, L. M. (1995). Biased attention in childhood anxiety disorders: A preliminary study. *Journal of Abnormal Child Psychology, 23*(2), 267–279. https://doi.org/10.1007/BF01447092

Wikström, J., Lundh, L. G., & Westerlund, J. (2003). Stroop effects for masked threat words: Pre-attentive bias or selective awareness? *Cognition & Emotion, 17*(6), 827–842. https://doi.org/10.1080/02699930244000219

Wikström, J., Lundh, L., Westerlund, J., & Högman, L. (2004). Preattentive bias for snake words in snake phobia? *Behaviour Research and Therapy, 42*(8), 949–970. https://doi.org/10.1016/j.brat.2003.07.002

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120*(1), 3–24. https://doi.org/10.1037/0033-2909.120.1.3

Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion, 7*(4), 882–886. https://doi.org/10.1037/1528-3542.7.4.882

Ypsilanti, A., Mullings, E., Hawkins, O., & Lazuras, L. (2021). Feelings of fear, sadness, and loneliness during the COVID-19 pandemic: Findings from two studies in the UK. *Journal of Affective Disorders, 295,* 1012–1023. https://doi.org/10.1016/j.jad.2021.08.031

Submitted: 2.15.2022
Revised: 5.25.2022
Accepted: 5.27.2022