

**Main Manuscript for**

Threat extinction without awareness: evidence from

masking and continuous flash suppression

Shani Bachar-Avnieli1\*, Liad Mudrik2, 3,. Daniela Schiller4, Gideon E. Anholt1

1Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

2Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel.

3School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel.

4Department of Psychiatry, Nash Family Department of Neuroscience, and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

\*Corresponding author:

Shani Bachar-Avnieli

Email: avnielis@post.bgu.ac.il

**Classification:**

Social Science: Psychological and Cognitive Sciences

 Keywords: Masking, Unconscious processes, Anxiety.

**Author Contributions**

**This PDF file includes:**

Main Text

Figures 1 to 5

**Abstract**

Anxiety disorders are the most common mental disorders. The typical treatment for these disorders is exposure to the anxiety inducer. However, many patients are reluctant to confront feared objects or situations and are consequently left untreated. To date, studies showed that threat could be acquired without awareness, though there is no conclusive laboratory evidence that threat can also be extinguished unconsciously. The present research aims to evaluate the feasibility and robustness of extinction evoked by unconsciously perceived stimuli, using two common methodologies: Continuous Flash Suppression (CSF) and Visual Masking (VM). We describe two experiments utilizing these two methods (N=38 and N=72, respectively). In each experiment, healthy participants were first threat-conditioned, and later randomly divided into three groups: The first underwent conscious exposure to the threat-conditioned stimuli in order to evoke extinction; the second was unconsciously exposed to these stimuli; and the third served as a control group that did not undergo extinction. We found effective extinction and successful extinction retention (indexed by a skin conductance response) under both conscious and unconscious exposure conditions only under VM, while in CFS the results are inconclusive and justify further investigation. These results demonstrate the effectiveness of unconscious extinction, even when strict measures of awareness are taken. Although CFS and VM have been well described in the literature, this study is the first comprehensive study that directly compares these two methods with respect to unconscious exposure. Our findings bear theoretical implications for the understanding of exposure therapy and may pave a path for the potential clinical utility of unconscious extinction.

**Significance Statement**

Exposure is an effective psychological intervention for the treatment of anxiety. However, this method is associated with considerable treatment refusal and high drop-out rates. A possible advancement towards solving this problem may be found in the field of consciousness studies. The current research goal was to test for unconscious extinction while assessing conscious experience of the suppressed stimuli, using two different suppression techniques. Our results suggest that exposure to phobic images can reduce fear and produce a general decrease of SCR. These findings could lay the foundation for establishing a therapeutic protocol, which relies on unconscious exposure.

**Introduction**

The most common intervention for anxiety disorders is treatment with exposure to a feared situation or object (1). The goal of exposure is to facilitate extinction – that is, to reduce the conditioned threat response to the triggering stimuli (2). Although effective psychological and pharmacological treatments exist for anxiety disorders (3), most people with such disorders never seek treatment (4). A possible explanation for these low rates of treatment-seeking is that patients consider confronting feared objects or situations as overly demanding. Accordingly, novel interventions that will help patients minimize their encounter with aversive stimuli are needed (5). A possible advancement towards this goal may be found in the field of consciousness studies, where different techniques have been developed to present stimuli so that they will be processed without conscious awareness. The current research aimed at exploring the feasibility of unconscious extinction.

In recent years, there has been increasing evidence suggesting that threat responses can be acquired or evoked by stimuli that are rendered invisible (6-10). For example, one study reported that fearful facial expressions emerge from suppression into awareness more quickly relative to images of neutral or happy expressions, suggesting that emotional expressions might be unconsciously processed (11). These results are supported by findings demonstrating increased amygdala activation in response to fearful, compared with happy, masked faces (but see (12)). These findings are further in line with LeDoux’s (13) suggestion of a direct path between the thalamus and the amygdala, such that information may evoke defensive responses, even without activation of the visual cortex.

If threat can be acquired and experimentally evoked outside of awareness, it is plausible that conditioned threat could also undergo extinction under similar conditions. However, previous literature does not provide clear and compelling evidence for extinction evoked by invisible stimuli. Several behavioral studies tested whether subliminal exposure to images of spiders affects one’s willingness to approach a spider among people who are afraid of spiders (14, 15). In these studies, participants completed a behavioral avoidance test (BAT) one week after a masked exposure to a spider to examine long-term effects of this form of exposure. They were then presented with images of spiders, either masked or unmasked. Participants in the masked condition were more willing to approach a spider than those who were consciously exposed to pictures of spiders. These findings were replicated with two-week and one-year follow-up measurement (15, 16) . Importantly however, in these studies (14-16) no online measures of awareness (neither subjective or objective; see (17)) were taken. Instead, the premise that participants were not aware of the stimuli was based on a preliminary masking experiment, with a different sample (14, 15), in which participants were unable to identify the masked images. In another study, Participants were required to complete an identification questionnaire evaluating awareness only at the end of the experiment, so no online tracking of Participants’ awareness of the stimuli was performed (18) . In addition, these studies focused on behavioral effects, which have yet to be corroborated by physiological responses. Only two studies measured participants’ skin conductance in response to exposure; one concluded that masked exposure is not associated with increased physiological responses in the extinction process: although participants in the masked condition did succeed more in the BAT, no evidence for reduced physiological responses was obtained (18). The other showed the potential benefit of unconscious exposure using Continuous Flash Suppression (19). In that study, fear reduction was manifested by a measure of threat-potentiated startle responses, but not in Skin Conductance Response (SCR). Furthermore, that study lacked a control group where no extinction took place.

Given these gaps, our goal was to investigate the effectiveness of unconscious exposure to aversive stimuli, when awareness is properly controlled. Various techniques have been developed to suppress stimuli from awareness and assess unconscious processing. These techniques were used to measure the impact of specific stimuli on participants’ thoughts, feelings, actions and learning processes (20, 21). We focused on two prominent techniques: Visual Masking (22, 23) and CFS (24). In VM, a stimulus (“target”) is presented for a short duration of a few dozens of milliseconds or less, and is immediately preceded/followed by masks, which render it invisible (22, 23, 25). CFS relies on dichoptic vision: A target stimulus is presented to one eye, while the other eye is consistently exposed to a changing pattern of different shapes. This technique prevents participants from seeing the constant target image for a relatively long period of time (up to several seconds) (24). Importantly, the two techniques may involve different underlying mechanisms and may evoke different types of unconscious processing (23, 26, 27). Surprisingly, very few studies have compared CFS to VM using the same task and stimuli (28-31). Thus, doing so might also shed light on their underlying mechanisms. Furthermore, employing both methodologies is advantageous when studying unconscious processes, as it provides means to assess the robustness and generalizability of the effect.

 To assess the threat response, we relied on a commonly used autonomic measure of threat conditioning – SCR(32). Studies have demonstrated that this measure contributes to the understanding of anxiety disorders when used in experiments of fear conditioning (10). The present research aimed at evaluating the feasibility and robustness of extinction evoked by unconsciously perceived stimuli by (a) testing for extinction while carefully assessing conscious experience of the suppressed stimuli; and (b) testing for the robustness and generalizability of non-conscious extinction across both CFS (Experiment 1), where a stimulus is presented to the non-dominant eye and rendered invisible by presenting colored dynamic patterns to the dominant eye (24) and VM (Experiment 2), where conscious perception of a briefly presented stimulus is prevented due to masks that immediately follow it, preceded it, or both(23). Both experiments included measurements of changes in SCR, while viewing a pre-installed computer presentation on a monitor. The studies started with an acquisition phase, where the participants were presented with the Conditioned Stimuli (CS+ and CS-): a scared face of a man or a woman. While a CS was presented, participants received mild electric shocks at a level which they determine to be “aversive” and “uncomfortable, but not painful”. Subsequently, participants underwent an extinction phase, where they were presented with the same stimuli again, without receiving the electric shock. Then, for the extinction phase, participants were divided into 3 groups: (1) “Unaware group”, (2) “Aware group”, and (3) “No Extinction group”. The Unaware group was presented with the CS+/CS- stimuli under CFS (Experiment 1) or VM (Experiment 2), the Aware group was presented with the CS+/CS- stimuli without CFS or VM, and the No Extinction control group was presented with scrambled versions of the CS+/CS- stimuli, under CFS or VM. In both the unaware and the No Extinction groups, participants’ awareness of the suppressed stimuli was carefully assessed using both objective and subjective measures. Finally, in the testing phase, all participants were presented with the CS+/CS- stimuli, to assess the effects of conscious and unconscious extinction relative to the No Extinction group.

**Results**

**EXPERIMENT 1: unconscious threat extinction using CSF**

*Assessment of Awareness*

Participants’ subjective ratings of the visibility of the stimuli (using the Perceptual Awareness Scale; (33) showed that in 80.2% of the trials, Participants reported not seeing anything (PAS=1), and in 14.6% of the trials, they reported having only a vague perception of something (PAS=2). Only in 4.89% of the trials they reported seeing either a clear part of the image or seeing the picture clearly (PAS=3 and 4, respectively). Accordingly, we only retained participants whose PAS average ranged from 1 to 1.125 for all subsequent analyses. For these participants’ objective performance in determining whether the suppressed face was a man or a woman was at chance (M=47.9%, SD=7%); (t(10) = 0.379, p = 0.920). This null result was supported by a Bayesian paired-sample t-test which revealed that, given our data, the null hypothesis was 3.15 times more likely than the alternate hypothesis. Taken together, awareness assessments confirmed that the masking procedures were effective and that participants were unaware of the faces.

**Conditioned threat acquisition**

To evaluate the acquisition process in the three groups, we conducted an ANOVA with group (Unaware, Aware, No-extinction) and stimulus (CS+, CS−) as factors. There was no evidence for group differences as indicated by the non-significant main effect of group (F(1,38) = 2.64, p=0.084, partial η2 = 0.12) or interaction between group and stimulus (F(2,38) = 1.22, p = 0.304, partial η2 = 0.06 ). SCR responses were higher to the CS+ compared to the CS− [F (1,38) =311.8, p < 0.001, partial η2 = 0.89]. As expected, these results confirm sufficient and similar acquisition across all groups.

**Threat extinction**

To evaluate extinction learning in the three groups, we focused on CS difference scores averaged over the extinction phase and checked if they differed between groups and between the early and late phases of extinction, defined as 12 first trials and 12 last trials. Accordingly, CS difference scores were entered into a 3 × 2 repeated measures ANOVA with group (Unaware; Aware; no-extinction) and time (Early; Late) as factors. A main effect of time [F (1, 35) = 9.479, p = 0.004, partial *η*2 = 0.21] was observed. However, no main effects of group [F (2, 35) = 1.60, p = 0.216, partial *η*2 = 0.08], or interaction of group and time [F (2, 35) = 0.58, p =0.565, partial *η*2 = 0.03] were found (Figure 2).

Pairwise comparison analysis for each group separately showed a marginally significant effect (p = 0.07) for the unaware group, whereby responses decreased between the Early and Late Extinction Phases, suggesting that extinction can occur also when the stimuli are rendered invisible. As expected, extinction was found also in the aware group, where responses declined between Early and Late Extinction (p =0.01). As opposed to the aware group, responses in the no-extinction group did not differ between early and late Extinction (p = 0.328), as expected.

**Extinction retention**

To evaluate between-group differences in the extinction process, a Recovery Index (RI) was calculated by deducting the CS-elicited threat response at the last four trials of extinction from the trials of the Test Phase. Univariate analysis confirmed a group difference on the RI, yielding an effect size (partial η2) of 0.34 (Figure 3). The no-extinction group (M = 0.116; p=0.001) showed clear indications of recovery, while no recovery was observed in the aware group (M = -0.077; p=0.02). Critically however, although there was a decrease in SCR between the extinction and the test phase, this difference was far from reaching significance (M = -0.02, p=0.501).

Taken together, Experiment 1 yielded no evidence for unconscious extinction, with decreases in SCR found only among participants in the aware group. This might cast doubt on the mere existence of unconscious extinction – when proper measures are taken to assure the stimuli were indeed invisible. Alternatively, this might suggest that CFS does not allow for sufficient processing needed for unconscious extinction. To arbitrate between these two explanations, Experiment 2 was conducted using a larger sample and the VM paradigm, for which previous studies did find an effect (14-16, 18, 34, 35)

**EXPERIMENT 2: unconscious threat extinction using VM**

*Assessment of Awareness*

Akin to the CFS experiment, here too, the vast majority of trials were rated as 1 (83.9% “I did not see anything”), and 2 (13.1%, “I had a vague perception of something”). Only 2.89% were rated as either 3 (“I saw a clear part of the image”) or 4 (“I saw the picture clearly”). Objective performance confirmed that in the included trials (PAS=1), participants’ accuracy in the gender judgment task was again at chance level (M=48%, SD=12%) [ t (23) = -1.04, p =0.152]. Bayesian paired-sample t-test suggested that the null hypothesis was 2.86 times more likely than the alternate hypothesis. Again, Even-Odd analysis suggested a high reliability of the objective measures (r (23) = 0.15, p = .547). Thus, much like in Exp. 1, the awareness assessments in Exp. 2 suggested that masking was effective in suppressing the stimuli from awareness.

*Conditioned threat acquisition*

An ANOVA on the CS difference scores with group (Unaware, Aware, No-extinction) and stimulus (CS+, CS−) as factors yielded no evidence for group differences (F(1,69) = 0.34, p= 0.71, partial η2 = 0.01), nor an interaction between group and stimulus (F(2,69) = 0.85, p = 0.43, partial η2 = 0.02), yet SCR responses were higher to the CS+ compared to the CS− [F (1,69) =139.92, p < 0.001, partial η2 = 0.67]. As expected, these results confirm sufficient and similar acquisition across all groups.

*Threat extinction*

A 3 × 2 repeated measures ANOVA with group (Unaware; Aware; no-extinction) and Phase (Early; Late) as factors did not reveal a main effect for group [F (1,69) =0.06, p =9.35, partial η2 = 0.02]. However, main effects of time [F (2, 69) = 89.53, p < 0.001, partial *η*2 = 0.56], as well as the interaction of group and time, were found [F (2, 69) = 25.01, p < 0.001, partial *η*2 = 0.42] (Figure 4).

 Pairwise comparison analysis for each group separately went beyond the results of Experiment 1 by indicating a clear extinction effect for the unaware group. There, responses declined between Early and Late Extinction Phases (p < 0.0001). The same pattern was found in the aware group (p < 0.0001). As expected, in contrast to the unaware and the aware groups, the responses in the no-extinction group did not differ between early and late Extinction (p = 0.515).

*Extinction Retention*

The results showed that with respect to the RI, the no-extinction group (M = .0608; p=0.014) demonstrated clear indications of recovery, whereas no recovery was observed in the unaware (M = -.0546; p=0.059) and the aware groups (M = -.0467; p<0.001), yielding an effect size (partial η2) of 0.42. The RI differences scores for the three groups (Figure 5) showed clear indications of recovery for the control group, whereas no recovery was observed in the aware and unaware groups.

 Decreases in SCR were demonstrated among participants in the aware and unaware exposure groups, but not those in the no-extinction group. These results strengthen the findings of Siegal and colleagues (18), where a decrease in SCR was found following exposure to masked stimuli

**General Discussion**

Using two different methods to suppress stimuli from awareness, CFS and VM, we demonstrated that extinction can occur using VM but not using CFS, even when the CS are presented unconsciously. As expected, all experimental groups showed a comparable threat-response in the acquisition process, with higher SCR to the CS+ relative to the CS- stimulus. In CFS, in the late extinction the aware group showed decreased SCR, while the unaware group showed only a marginally significant decrease in SCR. Furthermore, in the RI measure although in the unaware group, there was a decrease in SCR, it did not reach significance. In VM, in the late extinction phase only the aware and the unaware groups showed significant decreased SCR, while in the no-extinction group the responses did not differ between early and late Extinction. This result was strengthened by the RI measure, demonstrating that recovery occurred only among participants that did not undergo extinction; there was no recovery evident among participants in the unaware group or the aware groups. Thus, this constitutes a clear demonstration of unconscious extinction, obtained in the VM paradigm with stringent trial-by-trial measures of awareness. In the CFS paradigm, no unconscious exposure effect was found.

Findings from the VM technique suggest that both conscious and unconscious exposure are effective in reducing the conditioned response via non-reinforced presentations. These findings could pave the way for establishing a new therapeutic protocol, which relies on unconscious exposure. Such a protocol has never been clinically implemented; however, some studies suggest that extinction through unconscious exposure may be effective, in line with our findings. There is a growing body of research on a novel brain imaging approach called decoded fMRI neurofeedback (36-38). This technique depends on rewarding unconscious neural representations of feared stimuli in order to counter-condition the feared representation. This approach has demonstrated promise in decreasing fear responses to laboratory-conditioned fears (39). Both approaches are similar in showing that fear extinction can occur without consciousness. Note however, that the approach utilized in the current research is much easier to implement and does not require the complex infrastructure needed for the decoded fMRI neurofeedback procedure.

Our finding that extinction might be independent from awareness has interesting theoretical implications. One of the predominant theories of exposure therapy is the inhibitory learning model (40). This model suggests that the relationship between the CS and the aversive stimulus is not eradicated during extinction. Rather, a new inhibitory connection is created, whereby the conditioned stimulus no longer predicts the aversive stimulus (thereby inhibiting the fear response). The inhibitory connection then “competes” with the previous fear learning. One of the core processes suggested to underly inhibitory learning is expectancy violation. It is based on the premise that a gap between expectations and actual outcomes is critical for acquiring new inhibitory expectations, that would compete with existing expectations. However, given that extinction learning is based on the formation of non-coincidental relationships between conditioned and unconditioned stimuli, awareness of the stimuli as well as the non-occurrence of the unconditioned stimulus is deemed essential (41). The current findings indicate that extinction can also occur outside of conscious awareness, which goes against this assumption. Future research may indicate whether unconscious exposure involves expectancy violation or other mechanisms (e.g., habituation). Another venue for future studies would be to utilize a 3-day study design, in which the acquisition phase would be separated from the extinction phase by at least 24 h. That would complete the picture obtained here, since used immediate exposure, so that consolidation of the previously acquired fear memory would not be possible.

In the current study, we chose to use two techniques to render stimuli invisible: CSF and VM. In the CFS technique, each eye is presented with a different stimulus. One eye is presented with a series of flashing high contrast images, while the other eye is presented with a stationary, often low-contrast target (24). In the VM technique, the target stimulus is displayed for a short time period of several dozens of milliseconds and is immediately preceded and followed by a masked stimulus (23).The CFS technique is based on the fact that the visual system is not able to handle incompatible input to both eyes. As a result, only one stimulus is able to reach awareness, while the other stimulus remains invisible (until it breaks suppression; (11)). In contrast, in the VM technique, the mask stimuli tamper with the feedback sweep that typically follows feedforward processing of the target stimulus (23). As such, the differences between the two techniques also influence the amount of information that gets through and is being processed (42). In our study VM technique outperformed the CFS, highlighting the differences between the techniques (27, 43), and strengthening claims that CFS might not allow for higher-level processing (44) This is in line with previous studies, reporting that unconscious processing under CFS is more limited than VM in presenting unconscious semantic effects (45, 46). Accordingly, some suggested that backward masking might be more sensitive for measuring unconscious high-level processing than interocular suppression (20). Our findings join that claim, though more research is needed to determine is this is indeed the case.

In conclusion, the present study demonstrated that unconscious exposure using CFS and VM. Results may facilitate the development of novel treatments integrating unconscious exposure with various psychopathologies, populations, therapeutic doses and combinations between unconscious and traditional in vivo treatment regiments.

**Method**

**Experiment 1: unconscious threat extinction using CFS**

*Participants*

Thirty-eight healthy participants with normal or corrected to normal visual acuity from the department of Psychology participated in the experiment for course credit (30 women, 35 right-handed, mean age = 23.82, SD = 1.13). Seven additional participants who did not have measurable responses to the shocks were not included in the data analysis (SCR Score > 0.2). The experiment was approved by the ethics Committee of the Ben-Gurion University. Participants with self-reported psychiatric or neurologic history were excluded. All participants signed an informed consent form before partaking in the experiment.

*Stimuli and apparatus*

Two grayscale photographs of a woman and a man with identical contrast and luminance degree, each expressing fear, served as the target stimuli in all groups. They were selected from the NIMSTIM Database (47). The choice to present fearful facial expressions was based on previous studies (8, 10, 12) showing that an angry or scared face can be identified faster than a smiling face, and that a fearful face might indicate a potential danger and is accordingly more easily conditioned.

The stimuli were presented on a black background. The pictures were blurred at the tips by using Photoshop software and were surrounded by black-and-white rectangle frames, as depicted in Figure 1, to facilitate fusion. The experiment was computerized, performed on a 19-inch color Samsung screen with 60 HZ refresh rate and 1024 × 768 resolution, using E-Prime software version 2.0. Participants performed the experiment in a slightly darkened room. Their heads were maintained 61 cm from the screen using a chin-rest. The stereoscope was produced by Stereo Aids (Western Australia), and allowed the presentation of a separate image to each eye. Electric shocks were administered to the participant via STIMSOLA system of Biopac Company. The system includes a STMISOLA slider and a USB component which enabled the communication between the shocker appliance and the E-Prime software. The power of the electric shock was in the range of 0–50 v and the shock’s duration was 200 ms. A snap electrode with isotonic gel was attached to the participants’ arms.

Fear arousal was measured through SCR, using the 150 MP system of the Biopac GSR100C Company. Samples were collected with the Acknowledge system of Biopac Company.

*Procedure*

All participants underwent three phases during the experiment: acquisition, extinction, and testing within one day (see Figure 1). The experiment started with an acquisition phase, which was done using a threat conditioning paradigm with partial reinforcement. A male/female face served as the conditioned stimulus (CS+; counterbalanced between Participants), and an electric shock was used as the unconditioned stimulus (US). The acquisition phase included 30 trials: four practice trials, 12 CS+ trials (e.g., female face presented alone), 12 CS- trials (male face presented alone), six CS-US trials (female face presented with a shock). The order of stimuli appearance was pseudo-randomized: the first trial was always reinforced and no more than two of the same trial types ever occurred consecutively. Subsequently, participants underwent an extinction phase, in which they were presented with the same stimuli again, but without electric shock (12 CS+ trials, 12 CS- trials). There, participants were divided into three groups: (1) The *unaware group* )n=11) was presented with the CS+/CS- stimuli under CFS, so the stimuli were presented to one of their eyes, while a flickering stimulus of colored squares (Mondrian) was presented to their other eye, (2) the *aware group* )n=13) was presented with the CS+/CS- stimuli without CFS, (3) the no-extinction group )n=14) was presented with scrambled versions of the CS+/CS- stimuli under CFS. In all groups, the stimuli were presented for 4 s. Participants in the unaware group and the no-extinction group, in which CFS was used, were asked two questions at the end of each trial in order to confirm their level of awareness to the stimulus. These questions were used as objective and subjective measures to assess the level of the participants’ awareness. The objective question was: “Was the person in the picture a man or a woman?” The subjective question was: “How visible was the picture?”, and Participants were asked to respond using the Perceptual Awareness Scale (PAS:(33)), where 1 denotes “I did not see anything”; 2 represents “I had a vague perception of something”; 3 signifies “I saw a clear part of the image”; and 4 indexes “I saw the picture clearly”. Finally, in testing phase to assess the effects of conscious and unconscious extinction were assessed relative to the no-extinction group. In the testing phase, all three groups were presented with the CS+/CS- stimuli, in order to measure the recovery process in all three groups.

*Psychophysiological stimulation and assessment*

Participants viewed a pre-installed computer presentation on a monitor while changes in their SCR were measured. For SCR recordings, electrodes were attached to the forefinger and the forearm on each participant’s left hand While SCR was measured, participants received mild electric shocks at a level that they determined to be “aversive” (undesired and unpleasant) and “uncomfortable, but not painful” (48). The electric shock appeared randomly about 0.5–4.5 s from the moment the stimulus was presented. Between one stimulus and another, there was a time gap of 8–12 s. The electric shock was delivered when participants were presented with a CS+ stimulus, but never when presented with a CS- stimulus.

*SCR Analysis*

SCR waveforms were analyzed offline, using the Acknowledge 3.9 software (Biopac Systems Inc.). SCR amplitudes to the conditioned and unconditioned stimuli were the dependent measures of conditioned and unconditioned responses, respectively. The level of SCR was determined by taking the base-to-peak difference for the first waveform (in microSiemens ,µS) in the 0.5–4.5 s window after stimulus onset. The minimal response criterion was 0.02 ms. The raw SCR scores were square-root-transformed to normalize distributions. These normalized scores were scaled according to each participant’s unconditioned response by dividing each response by the mean square-root-transformed unconditioned stimulus response.

**Experiment 2: unconscious threat extinction using VM**

The methods of experiment 2 were identical to those described in experiment 1, with the following exceptions:First, sample size was calculated using *t*he GPower software version 3.0.5 (29). We aimed at obtaining an effect size (ES) estimate of 0.34, based on the results of the first experiment. The projected sample size needed for an ES of 0.34 with an alpha of 0.05, power (1 − β) of 0.95, and three groups, was approximately 60. As such, we decided to recruit 72 participants for the current study (53 women, 65 right-handed, mean age = 25, SD = 1.93). Eleven additional participants also completed the experiment but were excluded for the following reasons: three participants were excluded due to technical problems with data recording, eight participants were classified as non-responders because they lacked a measurable SCR (defined as a response higher than 0.02 ms) on >75% of the trials and, thus, were excluded from analyses.

Second, in the extinction phase, participants again were randomly assigned to one of three extinction groups, yet here awareness manipulation was achieved via masking. Accordingly, in the unaware and the no-extinction groups the intact/scrambled CS was visually masked. Each trial began with a white fixation cue (+) in the center of a black screen for 300 ms. Next, CS was presented for 33 ms, followed by a scrambled CS that was immediately masked for 176 ms. Participants were presented with two questions to assess the level of their awareness, as described in experiment 1. Participants were told they are not required to make a speeded response, and can answer within 6 s. In the aware group, the stimulus was presented without any masking for a duration of 4 s.

**Acknowledgments**

**References**

1. B. J. Deacon, J. S. Abramowitz, Cognitive and behavioral treatments for anxiety disorders: A review of meta‐analytic findings. *Journal of clinical psychology* **60**, 429-441 (2004).

2. J. S. Abramowitz, The practice of exposure therapy: relevance of cognitive-behavioral theory and extinction theory. *Behavior therapy* **44**, 548-558 (2013).

3. E. M. Investigators *et al.*, Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta psychiatrica scandinavica* **109**, 21-27 (2004).

4. H.-U. Wittchen *et al.*, The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European neuropsychopharmacology* **21**, 655-679 (2011).

5. J. Pearson, Associative learning: Pavlovian conditioning without awareness. *Current Biology* **22**, R495-R496 (2012).

6. C. M. Raio, D. Carmel, M. Carrasco, E. A. Phelps, Nonconscious fear is quickly acquired but swiftly forgotten. *Current Biology* **22**, R477-R479 (2012).

7. P. Homan *et al.*, Evidence for a minimal role of stimulus awareness in reversal of threat learning. *Learning & Memory* **28**, 95-103 (2021).

8. A. Öhman, S. Mineka, Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological review* **108**, 483 (2001).

9. A. Öhman, Fear and anxiety as emotional phenomena: clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. (1993).

10. F. Esteves, C. Parra, U. Dimberg, A. Öhman, Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear‐relevant facial stimuli. *Psychophysiology* **31**, 375-385 (1994).

11. T. Stein, M. V. Peelen, Dissociating conscious and unconscious influences on visual detection effects. *Nature Human Behaviour*, 1-13 (2021).

12. P. J. Whalen *et al.*, Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience* **18**, 411-418 (1998).

13. J. LeDoux, The emotional brain: The mysterious underpinnings of emotional life. *World and I* **12**, 281-285 (1997).

14. J. Weinberger, P. Siegel, C. Siefert, J. Drwal, What you cannot see can help you: The effect of exposure to unreportable stimuli on approach behavior. *Consciousness and cognition* **20**, 173-180 (2011).

15. P. Siegel, J. Weinberger, Very brief exposure: The effects of unreportable stimuli on fearful behavior. *Consciousness and Cognition* **18**, 939-951 (2009).

16. P. Siegel, R. Warren, Less is still more: Maintenance of the very brief exposure effect 1 year later. *Emotion* **13**, 338 (2013).

17. E. M. Reingold, P. M. Merikle, Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics* **44**, 563-575 (1988).

18. P. Siegel, R. Warren, G. Jacobson, E. Merritt, Masking exposure to phobic stimuli reduces fear without inducing electrodermal activity. *Psychophysiology* **55**, e13045 (2018).

19. J. P. Oyarzún, E. Càmara, S. Kouider, L. Fuentemilla, R. de Diego‐Balaguer, Implicit but not explicit extinction to threat‐conditioned stimulus prevents spontaneous recovery of threat‐potentiated startle responses in humans. *Brain and behavior* **9**, e01157 (2019).

20. P. Sterzer, T. Stein, K. Ludwig, M. Rothkirch, G. Hesselmann, Neural processing of visual information under interocular suppression: a critical review. *Frontiers in psychology* **5**, 453 (2014).

21. S. Kouider, S. Dehaene, Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**, 857-875 (2007).

22. D. Kahneman, Method, findings, and theory in studies of visual masking. *Psychological Bulletin* **70**, 404 (1968).

23. B. G. Breitmeyer, H. Ogmen, Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & psychophysics* **62**, 1572-1595 (2000).

24. N. Tsuchiya, C. Koch, Continuous flash suppression reduces negative afterimages. *Nature neuroscience* **8**, 1096-1101 (2005).

25. B. G. Breitmeyer, T. Ro, H. Ogmen, A comparison of masking by visual and transcranial magnetic stimulation: implications for the study of conscious and unconscious visual processing. *Consciousness and Cognition* **13**, 829-843 (2004).

26. C.-Y. Kim, R. Blake, Psychophysical magic: rendering the visible ‘invisible’. *Trends in cognitive sciences* **9**, 381-388 (2005).

27. S. V. Fogelson, P. J. Kohler, K. J. Miller, R. Granger, P. U. Tse, Unconscious neural processing differs with method used to render stimuli invisible. *Frontiers in psychology* **5**, 601 (2014).

28. G. Izatt, J. Dubois, N. Faivre, C. Koch, A direct comparison of unconscious face processing under masking and interocular suppression. *Frontiers in psychology* **5**, 659 (2014).

29. Z. Peremen, D. Lamy, Comparing unconscious processing during continuous flash suppression and meta-contrast masking just under the limen of consciousness. *Frontiers in psychology* **5**, 969-969 (2014).

30. J. Almeida, P. E. Pajtas, B. Z. Mahon, K. Nakayama, A. Caramazza, Affect of the unconscious: visually suppressed angry faces modulate our decisions. *Cognitive, Affective, & Behavioral Neuroscience* **13**, 94-101 (2013).

31. N. Faivre, V. Berthet, S. Kouider, Nonconscious influences from emotional faces: a comparison of visual crowding, masking, and continuous flash suppression. *Frontiers in psychology* **3**, 129 (2012).

32. W. Boucsein, *Electrodermal activity* (Springer Science & Business Media, 2012).

33. T. Z. Ramsøy, M. Overgaard, Introspection and subliminal perception. *Phenomenology and the cognitive sciences* **3**, 1-23 (2004).

34. P. Siegel, J. Weinberger, Less is more: The effects of very brief versus clearly visible exposure. *Emotion* **12**, 394 (2012).

35. P. Siegel, J. F. Anderson, E. Han, Very brief exposure II: The effects of unreportable stimuli on reducing phobic behavior. *Consciousness and cognition* **20**, 181-190 (2011).

36. V. Taschereau-Dumouchel *et al.*, Towards an unconscious neurotherapy for common fears. *bioRxiv*, 170183 (2017).

37. R. Sitaram *et al.*, Closed-loop brain training: the science of neurofeedback. *Nature Reviews Neuroscience* **18**, 86-100 (2017).

38. K. Shibata, T. Watanabe, Y. Sasaki, M. Kawato, Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *science* **334**, 1413-1415 (2011).

39. A. Koizumi *et al.*, Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. *Nature human behaviour* **1**, 1-7 (2016).

40. M. E. Bouton, Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological bulletin* **114**, 80 (1993).

41. M. G. Craske, M. Treanor, C. C. Conway, T. Zbozinek, B. Vervliet, Maximizing exposure therapy: An inhibitory learning approach. *Behaviour research and therapy* **58**, 10-23 (2014).

42. B. G. Breitmeyer, Psychophysical “blinding” methods reveal a functional hierarchy of unconscious visual processing. *Consciousness and Cognition* **35**, 234-250 (2015).

43. E. J. Cox, I. Sperandio, R. Laycock, P. A. Chouinard, Conscious awareness is required for the perceptual discrimination of threatening animal stimuli: a visual masking and continuous flash suppression study. *Consciousness and cognition* **65**, 280-292 (2018).

44. S. Yuval-Greenberg, D. J. Heeger, Continuous flash suppression modulates cortical activity in early visual cortex. *Journal of Neuroscience* **33**, 9635-9643 (2013).

45. Y.-H. Yang, Y.-H. Tien, P.-L. Yang, S.-L. Yeh, Role of consciousness in temporal integration of semantic information. *Cognitive, Affective, & Behavioral Neuroscience* **17**, 954-972 (2017).

46. M.-S. Kang, R. Blake, G. F. Woodman, Semantic analysis does not occur in the absence of awareness induced by interocular suppression. *Journal of Neuroscience* **31**, 13535-13545 (2011).

47. N. Tottenham *et al.*, The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research* **168**, 242-249 (2009).

48. A. Öhman, G. Erixon, I. Löfberg, Phobias and preparedness: Phobic versus neutral pictures as conditioned stimuli for human autonomic responses. *Journal of Abnormal Psychology* **84**, 41 (1975).

**Figures and Tables**

|  |
| --- |
| **Diagram  Description automatically generated** |
| **Diagram  Description automatically generated** |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | **Figure 1. Experimental procedure and stimuli**. All participants underwent three phases in the experiment: acquisition, extinction, and test phases. Acquisition phase (**A**): participants were presented with a male/female face, while receiving an electric shock for some of the trials. Extinction phase (**B**): participants were again presented with the same stimuli that were scrambled (in the no extinction group), without receiving the electric shock. In the aware group, the stimuli were visible, while in the unaware group, depicted here, they were rendered invisible using CFS. Participants were asked two questions at the end of each trial. These questions were used as objective and subjective measures to assess the level of the participants’ awareness. Finally, all participants underwent a testing phase to assess the effects of conscious and unconscious extinction forms relative to the no-extinction group.

|  |  |  |
| --- | --- | --- |
|

|  |
| --- |
|  |
| **Figure 2. Experiment 1 (CFS): Evidence for extinction across the groups. Normalized Skin Conductance Responses for early and late extinction displayed by boxplot**. For the aware condition (pink portion) responses to the CS declined significantly between early and late Extinction, and marginally significant declined in unaware condition (blue portion). However, responses in the No Extinction condition (green portion) did not differ throughout the extinction session**.**   |

 |

|  |  |
| --- | --- |
| Chart, box and whisker chart  Description automatically generated |  |
|  |  |
| **Figure 3. Experiment 1- CFS-** **indications of recovery illustrates by boxplot.** Recovery index (RI) was defined by deducting the CS elicited threat response at the last four trial of extinction from the trials of the Test Phase. The Aware condition has lowest mean values and one outlier. The No Extinction condition has the highest mean values, while the Unaware condition has the smallest range Interquartile and the lowest standard deviation. The figure demonstrates that there was no recovery only among participants in the Aware condition.. |

|  |  |
| --- | --- |
|  |  |
| **Figure 4. Experiment 2 (Visual Masking): Evidence for extinction across the groups.** **Normalized SCRs for early and late extinction displayed by boxplot**. For the aware (pink portion) and the Unaware conditions (blue portion) responses to the CS declined significantly between early and late Extinction. However, responses in the No Extinction condition (green portion) did not differ throughout the extinction session**.**

|  |  |
| --- | --- |
|  |  |
| **Figure 5. Experiment 2- Visual Masking-** **Indication of Recovery illustrates by boxplot.** Recovery index (RI) was defined by deducting the CS elicited threat response at the last four trial of extinction from the trials of the Test Phase. The Unaware condition has lowest mean values and the smallest range Interquartile and the lowest standard deviation. The Aware condition has the broader range interquartile and a few outliers. However, the No Extinction condition has the highest mean values. The figure demonstrates that recovery occurred only among participants in the No Extinction condition, while no recovery in both condition: Unaware and Aware.  |

  |  |

 |