Frightened by the perpetrator's voice: Startle responsivity and cognitive processing predict earwitness speaker identification

Maartje J. Schreuder[a] *, Thomas Meyer[a, b, c] *†, Alana C. Krix[a]

[a] Clinical Psychological Science, Maastricht University, The Netherlands

[b] Behavioural Science Institute, Radboud University Nijmegen, The Netherlands

[c] Clinical Psychology and Psychotherapy, University of Münster, Germany

Author note

*The first two authors contributed equally to this work. † Correspondence concerning this article should be addressed to Thomas Meyer, Clinical Psychological Science, Maastricht University, P.O. Box 616, 6200MD Maastricht, The Netherlands. Present address: Clinical Psychology and Psychotherapy, University of Münster, Fliednerstrasse 21, 48149 Münster, Germany. E-mail: t.meyer@uni-muenster.de, Telephone: +49 (0)251-8334198.

Abstract

This study was inspired by the case of a robbery victim who was startled and reminded of the crime upon hearing a stranger's voice, while not clearly recognizing the speaker. To investigate whether specific voices can modulate startle reactions and thereby predict speaker identification, we presented an audio hijack scenario to 84 participants and afterwards asked them to identify the perpetrator among neutral and negative speech fragments, while measuring flash-evoked eye-blink startle responses. Furthermore, we addressed data-driven cognitive processing during the audio scenario as a potential moderator in voice discrimination. Negative speech and the perpetrator's voice led to potentiated startle. Enhanced startle was positively associated with voice discrimination, but only in neutral speech fragments. In negative fragments, this association was weakened as a function of self-reported levels of data-driven processing during encoding. Thus, startle responses can generally predict accurate voice recognition, but speech emotionality and cognitive processing moderate this relationship.

Key words: speaker identification; startle reflex; skin conductance levels; cognitive processing style; traumatic memory

Frightened by the perpetrator's voice: Startle responsivity and cognitive processing predict earwitness speaker identification

A few years ago, an armed robbery took place in a small Dutch supermarket. Several perpetrators forced the manager to open the safe of the market, while the other staff members were held hostage, handcuffed, and threatened with firearms. The staff members did not see the perpetrators but could hear them speak in a language they thought to be Arabic. At the time, the police were unable to arrest any suspects. Curiously, about one year later, a new lead emerged. One of the staff members heard a stranger talk and suddenly had a strong bodily reaction, including goose bumps all over her body, accompanied by intrusive memories of the robbery. Even though she did not clearly recognize the voice or understand what the person said, she notified the police. Following up on this new lead, the police consulted our forensic speech analysis department (the Maastricht Forensic Institute, the Netherlands), inquiring whether the woman's physical reaction to the stranger's voice could be a sign of actual voice recognition. To date, no experimental data are available to provide a satisfactory answer to this question. The present study set out to address this question empirically.

Earwitness identification performance is generally considered to be relatively poor compared to eyewitness identification, and highly susceptible to interference (Bull & Clifford, 1999; Öhman, Eriksson, & Granhag, 2013). However, in cases like the supermarket robbery, it is possible that memory for voices is considerably improved due to the memory-enhancing effects of stress and stress hormones (Wolf, 2009). In addition, a bodily startle reaction to hearing someone's voice – as reported by the employee in the aftermath of the supermarket robbery – could be regarded as a bodily sign of implicit memory, preceding voice recognition and identification (i.e., analogue to a somatic marker in decision making; Bechara, Damasio, Tranel, & Damasio, 1997).

However, matters are complicated by the fact that exaggerated startle responses also represent a core symptom of post-traumatic stress disorder (PTSD; American Psychiatric Association, 2013), a condition characterized by recurrent intrusive trauma memories and physiological hyper-responsivity. That is, PTSD patients often display exaggerated physiological responses when confronted with trauma-related stimuli, indicating heightened sensitivity to negative valence and defensive reactions (Orr, McNally, Rosen, & Shalev, 2004; Orr & Roth, 2000). Critically, in many cases, this modulation can be evoked by a large array of environmental triggers, even if the sensory similarity to the traumatic experience is only vague (Brewin, Gregory, Lipton, & Burgess, 2010; Ehlers, 2010; Ehlers & Clark, 2000). Thus, heightened physiological reactions do not always occur in response to stimuli that were actually present during a traumatic event, and are not necessarily accompanied by accurate memory recollection (e.g., McNally et al., 2004; Zoellner, Foa, Brigidi, & Przeworski, 2000).

This observation may be explained by cognitive processes during traumatic experiences that moderate the relationship between physiological reactivity and voice recognition. Information processing theories of PTSD (Brewin et al., 2010; Ehlers & Clark, 2000) propose that a maladaptive processing style coined *data-driven processing* (Roediger, 1990) plays an essential role in the (hyper-)accessibility of trauma memories. Data-driven processing refers to preferential encoding of superficial perceptual features (e.g., colours, shapes, sounds), which may interfere with encoding of the meaning and context of the situation (labelled *conceptually-driven processing*), thereby impairing later intentional recollection of the memories (Ehlers & Clark, 2000). Indeed, data-driven processing has been linked to intrusive trauma memories (Halligan, Michael, Clark, & Ehlers, 2003; Kindt, van den Hout, Arntz, & Drost, 2008), and may facilitate perceptual priming of trauma-related stimuli (cf. Ehlers, Michael, Chen, Payne, & Shan, 2006; Sündermann, Hauschildt, & Ehlers, 2013).

Remarkably, data-driven processing during an aversive situation may impact physiological responses and voice identification in two ways. First, focusing on superficial perceptual features enhances attention to detail, which might result in a stronger formation of verbatim traces (i.e., memory for surface details, such as voices; Brainerd, Reyna, & Ceci, 2008). Accordingly, both (selective) startle responsivity and voice discrimination performance would be expected to increase as a function of data-driven processing, amplifying the association between physiological responses and discrimination performance. Second and conversely, data-driven processing can be argued to *impair* stimulus discrimination. In particular, it has been suggested that recognition memory depends on the degree to which perceptual features are bound into a coherent trace during encoding (Schacter, Norman, & Koutstaal, 1998). Since data-driven processing is associated with feeling overwhelmed by perceptual impressions and an increase in intrusive memories (e.g., Kindt et al., 2008), it is likely to disrupt effective feature binding in memory (also see Huntjens et al., 2015). Accordingly, this processing style might generally impair stimulus discrimination, both in terms of physiological reactivity and voice discrimination.

With these considerations in mind, the present study aimed to elucidate the relationship between physiological responses to unexpected startle probes accompanying specific voices from an aversive event and voice identification performance, as well as the moderating role of data-driven processing during the aversive event. For this purpose, we subjected participants to an aversive audio scenario involving a violent perpetrator. This allowed us to subsequently test eye-blink startle responses and discrimination performance during an auditory voice lineup, using neutral and negative speech fragments (Bradley & Lang, 2000; Meyer et al., 2014). In order to test the moderating influence of processing style, half of the participants were instructed to focus on perceptual details (i.e., data-driven processing; Kindt et al., 2008) prior to listening to the aversive scenario, whereas the others were instructed to focus on the storyline and meaning (i.e., conceptually-driven processing).

Our main expectations were that voice probes from the perpetrator would lead to potentiation of the startle reflexes, relative to probes from unknown foil speakers. In addition, and replicating prior findings with non-voice auditory stimuli (Bradley & Lang, 2000), speech probes with negative valence were expected to amplify startle responses compared to neutral ones. Next, we expected that participants with stronger voice-induced potentiation of physiological responses would also be better able to discriminate the voices when prompted to identify the perpetrator's voice (both for neutral and negative probes). Finally, we explored the role of data-driven processing style during the aversive audio scenario, with two contrasting hypotheses in mind. In particular, data-driven processing might increase or decrease the association between startle modulation and discrimination, assuming that it (1) increases verbatim encoding due to more attention to detail, or (2) impairs recognition memory due to ineffective feature binding (see above).

In order to examine whether possible associations between startle and voice identification would be paralleled by similar effects in other types of episodic memory, we additionally included a free recall test of the aversive scenario. Based on the assumption that data-driven processing (as opposed to conceptually-driven processing) is associated with a less elaborated depth of processing and encoding of meaning (Craik & Lockhart, 1972), we expected data-driven processing to negatively impact the accuracy for memory for meaning and categories (i.e., gist memory; Brainerd & Reyna, 2002) in favour of memory for specific stimulus information (i.e., verbatim memory). Finally, and analogue to the analyses for voice identification, we explored whether processing styles moderated the association between startle and memory accuracy.

Method

Participants

Eighty-four participants (54 female) enrolled in this study and were assigned to one of two processing conditions: data-driven processing (*n* = 42) or conceptual processing (*n* = 42). All participants were undergraduate psychology students of Maastricht University and native speakers of Dutch, with an age ranging from 18 to 31 years (*M* = 21.3, *SD* = 2.8). They were recruited at the university campus via poster advertisements, announcements after lectures, and via an online research participation system. Upon response, they were screened for inclusion and exclusion criteria by means of a self-report checklist. Exclusion criteria were: hearing problems, psychological or psychiatric complaints or treatment in the past two years, neurological diseases such as epilepsy, psychoactive medication use such as anti-depressants, alcohol consumption of more than 15 units a week, cannabis or other drugs use more than once a week, severe traumatic experiences, or having been victim of serious emotional or physical abuse. Participants received research participation credits or a 15 € voucher in return for completing the study. This study was approved by the standing ethical committee of the Faculty of Psychology and Neuroscience, Maastricht University.

Aversive Audio Scenario

Plot and procedure. To expose participants to a reasonably ecologically valid aversive experience involving voices in the absence of visual stimulation, we used an audio-recorded staged bus hijacking lasting approximately 8 min. To introduce participants to the scenario, they were initially shown a live news article on a constructed website resembling a widely read Dutch news portal (www.nu.nl), surrounded by authentic links and advertisements. The article informed about an ongoing bus hijacking, many aspects of which would still be unknown. Next, participants listened to the recordings portraying the following plot.

A young woman (hereafter: *the victim*) enters a bus and takes a seat. A male stranger (hereafter: *the perpetrator*) sits down next to her and starts conversing in an obtrusive manner. He elaborates that he is going to visit his brother in prison, trivialising that his brother had robbed an elderly couple. The victim is initially reluctant to reply, and eventually tells him that she does not want to talk to him. The perpetrator starts insulting her, and the situation escalates when the bus driver intervenes, triggering the perpetrator to shout angrily. The bus driver calls the police, but the man grabs the telephone and demands the police to release his brother. After ending the call, he shouts at the passengers, pulls a gun, threatens the passengers, and takes the woman hostage. The bus driver almost loses control of the bus. The perpetrator calls the police again, threatening to shoot the passengers if his brother is not released within the hour. He whispers in the ear of the woman that he would shoot her first. The perpetrator keeps on swearing and threatening the passengers, walks around restlessly, kicking against objects, and shooting in the air. Finally, approaching police sirens are heard.

Recordings. The scenario was recorded on a Marantz PMD661 Professional solid state digital audio recorder, with the internal microphone, and saved as high-quality WAVE files (44100 Hz sample rate, 16 bit, stereo), and downsampled to 22050 Hz. Some background sounds (e.g. gunshot, glass breaking, sirens, etc.) were taken from online databases (www.freesound.org). The dramaturgy of the scenario was supported by various sound effects, including noises recorded in a local bus, unintelligible and unrecognizable background speech, heavy breathing and crying of the victim and other passengers, shrieking breaks, car horns, a breaking window, and sirens. Throughout the scenario, all sounds were recorded within centimetres of the victim and/or edited such that the listener experiences the audible events from her perspective, thereby amplifying the intensity of breathing and crying sounds.

Voice actors. Three versions of the scenario were used (counterbalanced across participants), differing only in the voice actor for the perpetrator role. This was done to ensure that later voice recognition effects are not relatable to a specific perpetrator voice. The actors were provided with a general script of the storyline and were free to improvise in order to make the scenario sound natural. Importantly, each actor spoke in a variety of styles (i.e., normal, agitated, angry, shouting, muffled, and whispered speech), recorded in different audio quality (i.e., good and distorted through an intercom), in order to enable the listeners to become accustomed to a broad diversity of the speaker's style (Brungart, Simpson, Ericson, & Scott, 2001; Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2004). We selected the three male students (all in their 20s) as perpetrator voice actors based on a small online pilot study ($N = 21$; see supplementary materials). The victim and the bus driver were played by a 22-year old female student and by a 48-year old male actor, respectively. All actors originated from the province of Limburg (in the Netherlands), and spoke in a light accent of that area. None of the actors except one (perpetrator role) had prior professional acting experience.

Encoding condition instructions: Data-driven and conceptually driven processing

The aversive audio scenario was accompanied by encoding instructions that corresponded either with a data-driven or a conceptually driven processing style (Kindt et al., 2008), depending on the participant's encoding condition. In particular, we adapted the processing style instructions described in Kindt et al. (2008; Study 2) to the aversive audio scenario in Dutch. Participants in the *data-driven processing condition* were instructed to act as a live witness involved in the events, and to focus on as many perceptual details as they could perceive, including sounds, emotions, and other details. Moreover, participants were told that we would later ask questions about perceptual aspects of the scenario, and that the rationale behind the events and the storyline would not be important. In the *conceptually-driven processing condition*, participants were asked to act as an uninvolved, external observer of the events, and to focus on the storyline, as well as on the rationale and intentions behind all events. They were told that questions would later be asked about the storyline and rationale of the events, while perceptual characteristics would not be important.

In both conditions, no additional instruction was given to focus specifically on the voice or to remember the voice, meaning that we prompted incidental learning of the perpetrator voice rather than intentional learning. Without preparation for speaker recognition, people are found to perform less well on speaker recognition tasks than with preparation (Hammersley & Read, 1996; Saslove & Yarmey, 1980).

To assess the degree to which participants engaged in data-driven and conceptually driven processing of the aversive audio scenario, we administered a Dutch translation of the Cognitive Processing Questionnaire (Halligan, Clark, & Ehlers, 2002). The data-driven processing subscale consists of eight items (e.g., *My mind was full of impressions and my reactions to them*; α = .72), whereas the conceptually driven scale consists of six items (e.g., *I tried to figure out what would happen next*; α = .50), each item requiring a response on a 5-point Likert scale (1 = *not at all*, 5 = *very strongly*).

Voice recognition and eye-blink startle reactivity

Based on existing paradigms to assess recognition memory and physiological responses to emotional stimuli and memory cues (Jackson et al., 2003; Meyer et al., 2014; Meyer et al., 2013), we devised a voice recognition paradigm that allows the simultaneous measurement of startle reactivity. In line with the majority of startle paradigms in the literature, we elicited startle reflexes cross-modally using photoflashes (e.g., Bradley & Lang, 2000). Participants were presented with 28 audio probes from each of the three perpetrator voice actors via headphones (i.e., 84 probes in total). Thus, 28 probes were from the target voice (i.e., the perpetrator voice heard by the participant), and 56 served as foils (i.e., the two other voices not heard by the participant). The content of the probes was unrelated to the aversive audio scenario, half being emotionally neutral (e.g., detailed description of pictures), and half having a negative valence (e.g., an aggressive insult). Participants were informed that they would listen to 84 audio fragments with male voices, and that we were interested in their ability to identify the perpetrator from the aversive audio scenario. Importantly, similar to eyewitness identification studies, it was stated that the selection of voices included ‘innocent civilians' and that the perpetrator might not actually be among the presented voices. Thus, we explicitly asked participants to answer 'no' in case of uncertainty. Furthermore, we warned participants to not get distracted by the contents of the fragments, and that they could ignore the photoflashes that would occur while listening.

Each audio probe lasted approximately 10 s, followed by a 10 s pause, after which the participant had to respond whether the voice belonged to the perpetrator (yes/no), and to provide a confidence rating on a scale ranging from 0 (*very uncertain*) to 10 (*very certain*). To elicit eye-blink startle reflexes, the audio fragments were accompanied by a white photoflash lasting 1/1042 s, occurring with a random jitter between 4 and 6 s after audio probe onset. The flashes were generated by two Nikon flash units (SB-700 Autofocus Speedlight) facing the participant from the left and right sides of the computer screen (approximate angle of 10 degrees, 80 cm distance), covered in transparent foam-isolated casings to minimize the flash noises. To reduce predictability and habituation, 12 probes were presented without photoflash (evenly distributed across speakers and valence). Trial order was randomized for each participant with the restriction that the same speaker and probe valence was not presented more than three times consecutively. To prevent fatigue, we inserted three pauses that could be terminated by the participant. The speech probes were recorded in the same way as the aversive audio scenario. Valence and arousal differences were established in a small online pilot study, as well as speaker voice distinctiveness and similarity (N = 21; see supplementary materials).

Script memory

Free recall. To assess the participants' memory of the events in the aversive audio scenario, a free recall task was employed. In particular, we instructed participants to describe as many details of the scenario as they could remember, including the people involved, actions, events, sounds, as well as details the participants might deem unimportant, but to refrain from guessing. Further, we instructed them to write everything the way it would come to mind, even if the details would not be in the exact order in which they happened. Participants were required to spend at least 5 min writing the free recall, and were told to imagine that their testimony was of critical value for the clarification of the events.

Specific memory questionnaire. In addition to the free recall, we tested specific detail memory using a questionnaire consisting of 30 statements about various details of the aversive audio scenario. Each item required participants to indicate whether the statement was accurate. Half of the items consisted of incorrect details, and half of accurate details that were present in all three versions of the scenario.

Physiological recordings

We recorded electro-dermal activity (EDA) during the aversive audio scenario to determine skin conductance levels. In line with Giesbrecht, Merckelbach, van Oorsouw, & Simeon (2010), EDA was recorded with a low voltage (0.6 V) using a galvanic skin response (GSR) module and a bipolar BrainAmp ExG amplifier (Brain Products, Germany). The two electrodes were placed on the middle phalanx of the index and the middle fingers of the non-dominant hand. Signals were sampled continuously at 1000 Hz, low-pass filtered at 250 Hz, and stored. Beforehand, EDA was additionally sampled during a 5 min resting phase in order to obtain baseline skin conductance levels.

During the voice recognition paradigm, we measured electromyography (EMG) in order to quantify eye-blink startle responses to the light flashes. Following published guidelines (Blumenthal et al., 2005), EMG was sampled continuously at 1000 Hz, high-pass filtered at 0.1 Hz, and stored. Two Ag/AgCl electrodes were attached below the eye opposite to the participant’s dominant hand to measure contractions of the orbicularis oculi muscle. A third electrode on the forehead served as the signal ground. Signals were amplified (separately from the GSR module) using the BrainAmp ExG amplifier. Electrode impedances were kept below 5 kOhm.

Self-report data

We collected personal data regarding the participants' gender, age, and linguistic background. To assess individual differences in frequent symptoms of depression, anxiety, and stress, we used the Dutch 21-item version of the Depression Anxiety and Stress Scales (DASS-21; Lovibond & Lovibond, 1995). The total sum score ($\alpha = .89$) served as an indicator of general psychopathology. Affective responses to the aversive audio scenario were assessed with the Dutch state version of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). The PANAS consists of two ten-item subscales measuring current positive affect (PA; αs > .73) and negative affect (NA; $\alpha > .72$). In addition, we administered four 100 mm Visual Analogue Scales (VAS; anchors: 0 = *not at all*, 100 = *very much*) to measure specific negative emotions (i.e., anxious, shocked, angry, sad).

Procedure

Participants were invited to a single session lasting about 140 min. They first gave informed consent and completed a set of baseline questionnaires. Next, the electrodes for physiological measurements were attached, after cleansing the hands with water (for EDA) and preparing the skin with abrasive gel (for EMG). Participants were then provided with headphones and left alone in a sound-isolated testing booth with dimmed lights. They could communicate via a CCTV system with the experimenters, who were sitting in an adjacent control room. Following 5 min of baseline SC measurement, they completed the PANAS and emotion VAS. They then read the encoding instructions (data-driven or conceptually driven processing) and were exposed to the aversive audio scenario. Afterwards, they underwent a retention interval of 30 min. In this period, they completed a second PANAS and emotion VAS, the Cognitive Processing Questionnaire, and performed a filler task (Attention Network

Test; Fan, McCandliss, Sommer, Raz, & Posner, 2002). Finally, participants performed the voice recognition task, followed by the free recall and the specific memory questionnaire.

Data reduction

Skin conductance level. In line with prior studies (Giesbrecht et al., 2010), EDA recordings during the baseline phase were averaged to derive baseline skin conductance levels. Next, the activity during the aversive audio scenario was averaged into fifteen bins of 30 sec and then transformed by subtracting baseline skin conductance levels, followed by square-root transformation to reduce heteroscedasticity.

Eye-blink startle. In line with published guidelines (Blumenthal et al., 2005), EMG signals were high-pass filtered at 28 Hz (12 dB/octave), rectified, and smoothed with a 40 Hz low-pass filter (48 dB/octave), using BrainVision Analyzer (Version 2.0.2.5859, Brain Products, Germany). Segments were extracted for each photoflash trial from -50 to 250 ms relative to flash onset. The 50 ms pre-flash interval was used for baseline correction. Trials with artefacts (i.e., signal changes > 20 μV during baseline, reflex onset during baseline) were identified and excluded semi-automatically. Reflex onset values were determined within windows from 20 – 150 ms post-flash, as well as the subsequent peak values, using an automated procedure (onset criterion: 4 *SD* with respect to baseline) followed by manual reassessment. Startle magnitudes were derived by subtracting onset from peak values, *z*-transformed, and then averaged per stimulus category (perpetrator or foils; neutral or negative speech). In addition, we calculated startle potentiation scores by subtracting the average startle magnitude in foil trials from perpetrator trials, separately for trials with neutral and negative speech.

Accuracy and response bias. In line with prior recognition memory studies (Quaedflieg, Schwabe, Meyer, & Smeets, 2013), voice recognition performance was addressed in terms of hit rate (i.e., proportion of correct and missed identifications of the perpetrator) and false alarm rate (i.e., proportion of false identifications and correct rejections), resulting in a discrimination index P_r, i.e. (# Hits + 0.5) / (# Hits + # Misses + 1) – (# False alarms + 0.5) / (# False alarms + # Correct rejections + 1), according to the two-high threshold model (Snodgrass & Corwin, 1988). In addition, a bias index B_r was determined as [(# False alarms + 0.5) / (# False alarms + # Correct rejections + 1)] / (1 - P_r). A higher P_r indicates better discrimination of the perpetrator voice compared to foils. B_r scores < .50 are indicative of a conservative response bias, whereas scores > .50 indicate the tendency to respond more liberally. Since the task involved more foils than targets and participants were instructed to respond conservatively, low B_r scores were expected.

Free recall coding. To get a comprehensive amount of variables, the (transcribed) audio scenarios were unitised by idea. According to Miller, deWinstanley, and Carey (1996), an idea unit is the smallest linguistic utterance containing both subject and predicate. For each unit, the presence of gist, verbatim, recall of a sound, false recall, fabrication, and emotional inference was coded. Verbatim recall was defined as the accurate recall of an idea unit. For instance, this required a sentence of a conversation to be written in exactly the same words as in the conversation. Gist recall was defined as the accurate recall of the intention of an idea unit, such as when participants paraphrased a sentence of a conversation using their own words. Sound recall was defined as the accurate description of non-speech noises, such as the ringing of a cell phone. False recall is an inaccurate recall of an original unit. An example of this is remembering hearing the bus driver say something, when actually the perpetrator said it. A fabrication is the recall of something that did not occur at all in the tapes (Campos & Alonso-Quecuty, 2008). Emotional inferences were defined as any mention of an emotional state of one of the subjects in the tape (e.g., “All passengers panicked.”).

Coding was done by two different raters who were blind to the participants’ encoding condition. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters. The inter-rater reliability for the two raters was substantial, Kappa = 0.77 ($p < .001$), allowing valid conclusions to be drawn from our data (Landis & Koch, 1977). After coding, the accuracy for each participant was calculated by dividing the total number of accurate verbatim, gist, and sound details by the total number of details (including false recall and fabrications) that were mentioned.

Statistical analyses

Within-subject effects and group differences were addressed using repeated measures analyses of variance (ANOVAs) and *t*-tests. In addition, we tested for potential (unintended) differences between the three versions of the aversive audio scenario by repeating the analyses with voice actor as additional between-subject factor. Below, we report these analyses in footnotes only when there were statistical effects of voice actor. When sphericity assumptions for ANOVA were violated, Greenhouse-Geisser corrected *p*-values are reported along with the respective epsilon and uncorrected degrees of freedom. Finally, linear associations were assessed using Pearson correlation analyses. Alpha was set at .05 (two-tailed) for all analyses.

Results

Cognitive and affective responses

At first, we tested whether the two different cognitive processing instructions yielded different self-reported levels of cognitive processing after the aversive audio scenario. *T*-tests showed that despite the different instructions, participants in both conditions reported similar levels of data-driven processing, t (82) = 0.25, $p = .80$, and conceptually-driven processing, t = 0.10, $p = .92$.

Next, we assessed mood effects in a 2 (Time: pre, post) × 2 (Instruction: data-driven, conceptually driven) repeated measures ANOVA on PANAS-NA scores. This revealed a main effect of Time, F (1,82) = 74.96, $p < .001$, $\eta^2_p = .48$, scores increasing from 11.6 (*SE* =

0.2) to 16.0 (*SE* = 0.6), in the absence of main or interaction effects involving Instruction, *p*s > .68. Similarly, there were main effects of time for all four emotion VAS scores (all *p*s < .001) that did not interact with Instruction (all *p*s > .29), all scores increasing in response to the aversive scenario (anxious: $M_{Difference}$ = 10.8, *SD* = 20.1; shocked: $M_{Difference}$ = 22.2, *SD* = 21.8; angry: $M_{Difference}$ = 18.4, *SD* = 23.3; sad: $M_{Difference}$ = 12.4, *SD* = 18.7).

Skin conductance levels

For SC levels, we assessed condition differences at baseline and during the aversive audio scenario. Despite the allocation to the two conditions based on the order of participation (i.e., comparable to random allocation), participants in the data-driven processing condition had higher baseline SC levels (*M* = 2.95 sqrt(μS), *SD* = 0.81) than in the conceptually-driven processing condition (*M* = 2.53 sqrt(μS), *SD* = 0.82), *t* (82) = 2.40, *p* = .018. A 15 (Time) × 2 (Instruction) repeated measures ANOVA on baseline corrected SC levels during the scenario revealed a significant effect of Time, *F* (14,1148) = 7.95, ε = .18, *p* < .001, η^2_p = .09, in the absence of main or interaction effects of instruction, all *p*s > .74 (see Figure 1). Throughout the entire scenario, SC levels were and remained clearly higher than at baseline, intercept *F* (1,82) = 58.00, *p* < .001, η^2_p = .41 (see Figure 1).

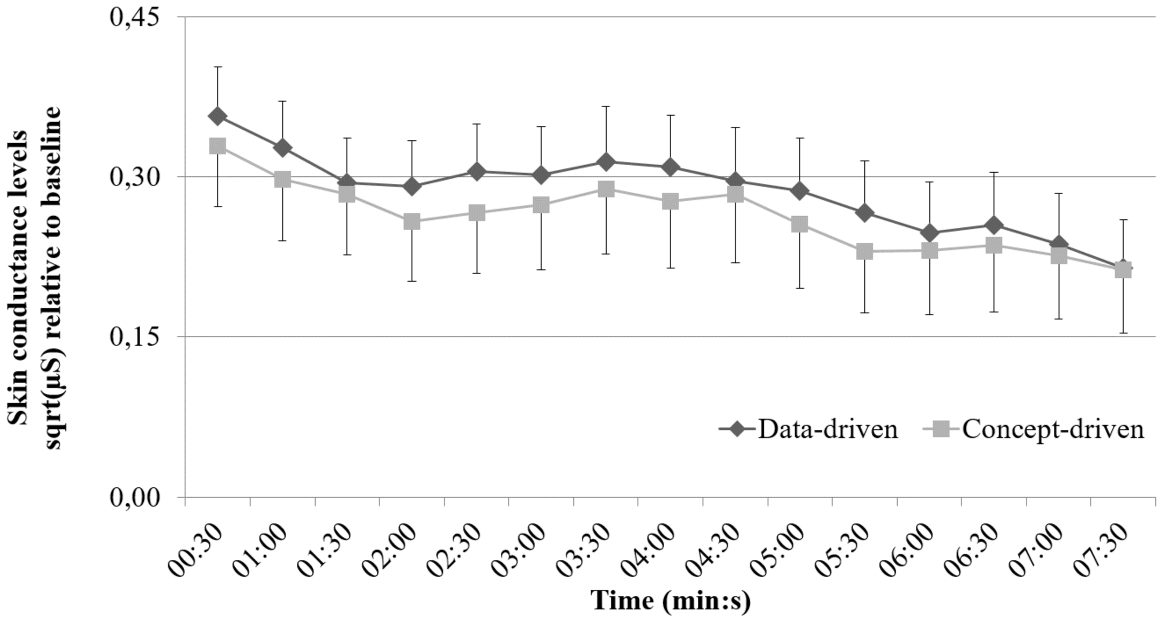


Figure 1. Skin conductance levels during the aversive audio scenario. Error bars indicate standard errors.

Voice recognition

A 2 (Valence: neutral, negative) by 2 (Instruction) repeated measures ANOVA on hit rates revealed a main effect of Valence, F (1,82) = 80.83, $p < .001$, $\eta^2_p = .50$, with relatively more hits on negative compared to neutral trials, but no main or interaction effect of Instruction, ps > .56. Similarly, we found higher FA rates in negative compared to neutral trials, F (1,82) = 16.26, $p < .001$, $\eta^2_p = .17$. This effect did not interact with Instruction, F (1,82) = 0.70, $p = .41$, $\eta^2_p = .01$. The perceptually-driven processing condition only tended to display higher false alarm rates (trend-level) than the conceptually driven condition, F (1,82) = 3.30, $p = .073$, $\eta^2_p = .04$. Higher hit and FA rates for negative compared to neutral trials were also reflected in better discrimination performance (Pr), F (1,82) = 74.71, $p < .001$, $\eta^2_p = .47$, and a more liberal response bias (Br), F (1,82) = 31.73, $p < .001$, $\eta^2_p = .28$, both in the absence of effects involving Instruction, all ps > .17. The average recognition performance scores can be inspected in Table 1.[1]

--- insert Table 1 about here ---

Script memory

Independent samples *t*-tests on the number of correctly reported verbatim details (free recall) showed no difference between the two conditions. Meanwhile, the conceptually-driven processing instruction yielded significantly more gist details, as well as fewer false details than the data-driven processing instruction. This pattern was further reflected in higher accuracy rates (see Table 2). When Actor was added as a factor to these analyses (ANOVA), these findings remained practically unchanged.

T-tests for the script memory test showed no significant group differences. However, participants in both conditions displayed very high hit rates (> 80%) and low false alarm rates (< 10 %) on average, indicating the possibility of a ceiling effect masking group effects with this method.

---

[1] The analyses with Voice actor (1, 2, 3) as an additional between-subject factor indicated that one of the voice actors was easier to discriminate than the other two, especially in probes with negative valence. That is, for FA rates, there was a Voice actor by Valence interaction, F (2,78) = 7.23, p = .001, η^2_p = .16. Negative probes of one actor yielding lower FA rates compared to the other two (*p*s < .001), who did not differ amongst each other (p > .99). This effect was less pronounced in neutral probes, where only one of the pairwise comparisons remained significant (p = .009). These effects were mirrored in the Pr discrimination scores (Voice actor main effect, p = .005; Valence interaction, p = .072), and in significant main and interaction effects on the bias score BR (*p*s < .001), suggesting that one of the actors led to more conservative and accurate identification than the other two, especially during negative voice probes. We repeated the voice recognition analyses after removing participants who listened to the voice actor with deviant FA rates, and found principally the same results. There again was a trend for participants in the data-driven processing condition to display higher FA rates, F (1,54) = 3.90, p = .053, η^2_p = .07, and a more liberal response bias BR, F (1,54) = 3.39, p = .071, η^2_p = .06, compared to the conceptually driven processing group.

--- insert Table 2 about here ---

Physiological reactivity

A 2 (Speaker: Perpetrator, Foils) by 2 (Valence: neutral, negative) by 2 (Instruction) repeated measures ANOVA revealed that as expected, flash stimuli in trials with negative valence induced higher startle magnitudes than in trials with neutral valence, F (1,82) = 21.34, $p < .001$, $\eta^2_p = .21$. Furthermore, we found a main effect of Speaker, F (1,82) = 12.10, $p = .001$, $\eta^2_p = .13$, with higher startle magnitudes during trials with the perpetrator, as compared with foil trials. There were no interaction effects or effects involving Condition, all *p*s > .33.

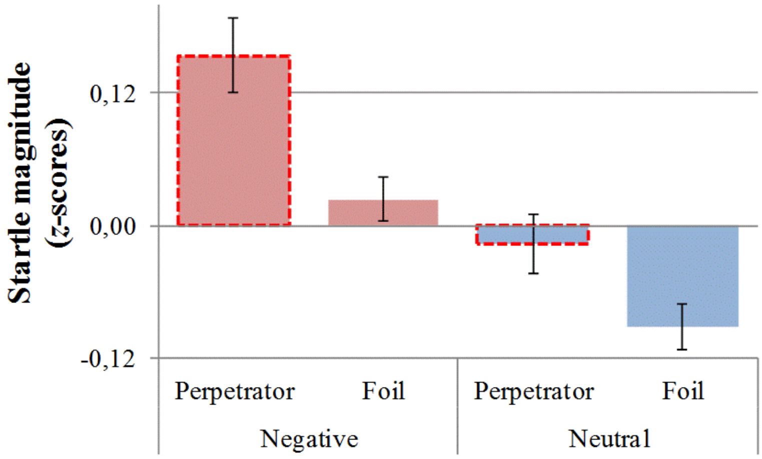


Figure 2. Average startle magnitudes per voice probe condition (*z*-transformed within subjects). Error bars indicate standard errors.

The relation between startle reactivity and memory performance

Speaker identification. We performed correlation analyses to examine whether stronger startle potentiation in response to a flash while hearing the perpetrator voice was associated with enhanced discrimination performance (Pr scores), separately for neutral and negative trials. We further explored whether potential effects would be driven by hit or FA rates, or by response bias. Since startle reactivity was unaffected by the encoding instructions, we performed these analyses in the entire sample, across the two instruction groups. As can be seen in Table 3, a positive correlation between startle and Pr scores emerged only for neutral trials, and was not observed in negative trials.[2]

--- Insert Table 3 about here ---

Memory of the scenario. Next, we explored whether enhanced startle would be associated with performance in the free recall and the script memory test. As the results in Table 4 show, startle potentiation scores in neutral trials were unrelated to script memory performance, while startle reactivity in negative trials tended to be associated with more reported verbatim details during free recall, and with better performance on the script memory test.

--- Insert Table 4 about here ---

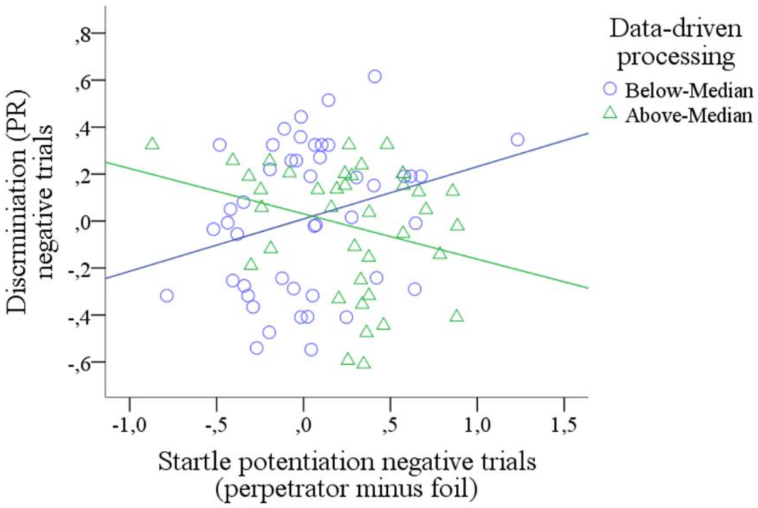
Moderating effects of cognitive processing

To further explore the potential role of cognitive processing in subsequent voice recognition, we performed multiple regression analyses on voice recognition performance scores (Pr; mean-centred per voice actor), separately for neutral and negative trials. Our particular interest was to test whether self-reported data-driven processing (regardless of encoding instruction) would influence the relationship between startle potentiation and subsequent voice recognition performance. Using step-wise multiple regression analyses, we first entered each individual's data-driven processing scores and startle potentiation scores, followed by the interaction term of data-driven processing and startle potentiation scores in

[2] In order to check whether participants who had listened to the voice actor with deviant FA rates were responsible for these effects (despite mean-centring per voice actor group, e.g., due to a restricted range in FA rates and Pr scores) we repeated these analyses after omitting these participants. In the remaining sample (n = 56), we found somewhat stronger positive correlations in neutral trials with Pr scores (r = .35, p = .008) and hit rates (r = .32, p = .015). Positive correlations were also present in negative trials, though small and non-significant, for Pr scores (r = .23, p = .085) and hit rates (r = .16, p = .251). Thus, the findings presented in Table 3 do not appear to be carried or dampened by this subgroup of participants.

step two. For Pr scores on neutral trials, the main effect of startle potentiation on neutral trials emerged again (see Table 3), $\beta = .25$, $t = 2.21$, $p = .030$, but there were no other main or interaction effects (all *p*s > .38). In contrast, for negative trials, a significant interaction term (data-driven processing by startle potentiation) emerged, $\beta = -.31$, $t = 2.88$, $p = .005$, in the absence of other main effects, all *p*s > .86. The nature of this interaction is visualized in Figure 3; the association between startle potentiation and voice discrimination tended to be more positive for participants with a less data-driven processing style during encoding. The above-mentioned results remained unchanged when affective and physiological responses to the aversive scenario (i.e., PANAS-NA change scores, baseline-corrected SCL) were entered as additional covariates. Similarly, the results remained unchanged, and no additional effects emerged when individual scores for conceptual-driven processing were entered as additional covariate.

Figure 3. Illustration of the interaction between startle potentiation and data-driven processing style during encoding in predicting voice discrimination performance in trials with



negative speech.

Discussion

Inspired by the case of a robbery victim who was startled by hearing a stranger's voice, we investigated whether stronger physiological startle responses to a sudden light flash accompanying voices from an aversive audio scenario correspond with better voice identification performance. We furthermore explored whether this link would be moderated by data-driven cognitive processing of the aversive scenario. Relative to unknown voices and neutral speech, we found potentiated physiological startle responses during recordings of the perpetrator's voice and negative speech. Voice discrimination performance was poor for neutral speech fragments and somewhat better for negative fragments. Importantly, startle potentiation in perpetrator trials correlated with better voice discrimination performance, but only for neutral speech fragments. These associations were weaker and non-significant in trials with negative speech. Meanwhile, startle potentiation and voice discrimination were largely uninfluenced by our cognitive processing instructions. However, higher self-reported levels of data-driven processing were associated with a decreased link between startle potentiation and discrimination performance.

Our data confirm the expectation that fear states associated with aversive memories can indeed be triggered by hearing a specific voice. This replicates and extends prior research showing similar effects when affective memories or imagery are triggered through external cue stimuli (e.g., Lang, Bradley, & Cuthbert, 1990; Meyer et al., 2013). Moreover, speech probes with negative valence generally amplified startle responses compared to neutral ones, in line with earlier findings with non-voice auditory stimuli (Bradley & Lang, 2000). Meanwhile, the relatively low voice discrimination scores for neutral probes are in line with the view that earwitness performance is generally poor (Bull & Clifford, 1999; Öhman et al., 2013). Discrimination performance was better for negative probes, possibly due to a higher affective congruence of the negative probes with the aversive scenario, facilitating both encoding and recognition (Bower, Gilligan, & Monteiro, 1981). Importantly, however, other factors may drive these effects, because the probes differed on more dimensions than valence and arousal. For instance, the differences in emotional intensity are necessarily coupled to differences in attention, interest, and speech characteristics (e.g., intensity, intonation) influencing the distinctiveness of voices (Robers, 2011). This limitation notwithstanding, our data indicate that our laboratory paradigms are well suited for studying the relationship between physiological reactivity and voice recognition performance. As we discuss in the following, our data suggest that this relationship may be moderated by the emotionality of the voice probes on the one hand, and by cognitive processing during encoding on the other.

The expectation that stronger physiological reactivity during hearing the perpetrator's voice predicts superior speaker discrimination was only supported for neutral recognition probes and not for negative probes. This is surprising, because the startle magnitudes were similarly influenced by the perpetrator's voice in neutral and negative probes (i.e., there was no interaction effect). As a possible explanation, the effects of negative valence on startle reactivity may have overshadowed an association between physiological responses and voice recognition. For instance, this relationship may have been obscured by a shift towards a more liberal response bias on negative trials. That is, participants were better at discriminating the perpetrator's voice from foils in negative compared to neutral speech fragments, which came at the cost of higher FA rates. Alternatively, automatic bodily fear responses might indeed precede and facilitate voice recognition and identification (e.g., similar to a somatic marker; Bechara et al., 1997), but only if the modulation of the response cannot be attributed to stimuli that are inherently fear-provoking. In addition, our results suggest that enhanced startle reactions to a light flash during negative speech fragments of the perpetrator are associated with memory characteristics other than voice discrimination, including better recall of verbatim details during free recall (but not overall accuracy) and recognition of script details.

Regarding the role of a data-driven processing style in memory formation and voice recognition, our data provide a number of novel insights that merit further investigation. Participants in our data-driven condition were less accurate in their free recall, due to more gist and fewer false details. This is in line with the view that data-driven processing leads to less deep and elaborated encoding of meaning and gist (Brainerd & Reyna, 2002; Craik & Lockhart, 1972). Notably, however, we found no immediate effect on skin conductance levels and self-reported data-driven processing. Using similar instructions before encoding traumatic films, Kindt et al. (2008; Experiment 2) found that their data-driven condition engaged more in this processing style (measured using a single-item VAS) than their conceptual condition, but not compared to a neutral instruction condition. Together with our findings, it appears that data-driven processing cannot be reliably enhanced through encoding instructions, while our conceptual condition failed to reduce it. In other words, participants may have relied much more on their automatic, non-instructed thinking styles than on our instructions in both groups. This might explain why the processing instructions had no impact on startle responses and voice recognition, except for a non-significant trend towards higher FA rates in the data-driven condition. Future studies may opt for a more potent manipulation of processing style, such as a writing assignment used by Kindt et al. (2008; Experiment 1), or manipulating memory load using a concurrent task.

Meanwhile, self-reported levels of data-driven processing moderated the link between startle potentiation and voice discrimination performance in speech fragments with negative valence. This suggests that data-driven processing does not affect startle reactivity or voice discrimination per se. In other words, our data contravene the hypothesis that focusing on superficial perceptual features during encoding directly enhances memory for a perpetrator's voice, for instance through enhanced verbatim traces (Brainerd, Reyna, & Ceci, 2008). Similarly, we found no support for the alternative proposal that data-driven processing impairs stimulus discrimination in a straightforward manner. Rather, individuals who engage more strongly in data-driven processing might be less able to rely on the automatic activation of fear states to identify a voice, if the speaker engages in negative arousing speech. This interpretation aligns with the idea that data-driven processing disrupts the encoding process of binding perceptual features into a coherent memory trace (Schacter et al., 1998). As a result, triggering relevant memory traces can cause fearful reactions without co-activating memory for features necessary for accurate stimulus discrimination. However, it is similarly possible that the effects occur at retrieval rather than during encoding. Thus, the precise mechanism by which their physiological and conscious discrimination of voices becomes uncoupled remains to be investigated more closely.

A few limitations of the present study merit particular attention. To begin with, encoding of the aversive audio scenario and memory testing took place within one session, separated by a relatively short retention interval of 30 min. As a consequence, memory performances on the free recall and the recognition questionnaire were very good, possibly causing ceiling effects for accuracy. Moreover, both physiological reactivity and voice recognition are likely to change in a longer retention interval, especially in individuals who develop intrusive memories and thus rehearse perceptual elements of the scenario. Similarly, a longer retention interval would enable a more systematic investigation of PTSD symptoms, as these take time to develop. Furthermore, our findings are based on a trauma-analogue paradigm with high-functioning students who were not actually threatened, and might not translate directly to trauma-exposed individuals or to standard earwitness paradigms. Thus, future studies are required to replicate and extend our findings with clinical samples, as well as with a longer time interval between encoding and memory testing. Finally, it is worth noting that we focused on the modulation of startle reflexes by neutral and negative arousing voice stimuli as an index of physiological reactivity. While startle was used as an index of negative valence and defensive motivation, we cannot rule out that the effects are driven by emotional arousal, or attention to emotionally significant stimuli, since we did not include stimuli with positive valence. Moreover, our findings may not be generalizable to other indices of physiological arousal.

Conclusions

Our study generally supports the view that modulation of startle responses upon confrontation with a perpetrator voice can inform about memory for an emotional event. However, the direct link between selective startle reactivity and accurate voice discrimination is moderate and limited to recognition in speech fragments that are emotionally neutral. In negative speech fragments that elicit heightened startle on their own, voice discrimination was generally better, but a positive association with selective startle reactivity emerged only in participants who reported lower levels of data-driven processing during encoding of the aversive scenario. Thus, speech valence and individual differences in cognitive processing of aversive experiences need to be considered when interpreting a startle reaction predictor of accurate speaker identification.

Funding and competing interests

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare no conflict of interest.

Acknowledgements

We wish to thank our students Ashley Vuurstaek, Shannen Remans, and Michèle Bal for their help in recording the audio scenarios and in collecting the data, Bo Strijker, and Femke Kitzen for the data collection, Richard Benning for adding extra sound effects, and we are grateful to our voice actors.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*(5304), 1293-1295. doi:10.1126/science.275.5304.1293

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology, 42*(1), 1-15. doi:10.1111/j.1469-8986.2005.00271.x

Bower, G. H., Gilligan, S. G., & Monteiro, K. P. (1981). Selectivity of learning caused by affective states. *Journal of Experimental Psychology: General, 110*(4), 451. doi:10.1037/0096-3445.110.4.451

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, *37*(2), 204-215. doi:10.1111/1469-8986.3720204

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164-169. doi:10.1111/1467-8721.00192

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*(3), 343-382. doi: 10.1037/0033-2909.134.3.343

Brewin, C. R., Gregory, J. D., Lipton, M., & Burgess, N. (2010). Intrusive images in psychological disorders: Characteristics, neural mechanisms, and treatment implications. *Psychological review, 117*(1), 210-232. doi:10.1037/a0018113

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5), 2527-2538. doi:10.1121/1.1408946

Bull, R., & Clifford, B. R. (1999). Earwitness testimony. *Medicine, science and the law, 39*(2), 120-127. doi:10.1177/002580249903900206

Campos, L., & Alonso Quecuty, M. L. (2008). Language crimes and the cognitive interview: Testing its efficacy in retrieving a conversational event. *Applied Cognitive Psychology, 22*(9), 1211-1227. doi:10.1002/acp.1430

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684. doi:10.1016/S0022-5371(72)80001-X

Ehlers, A. (2010). Understanding and treating unwanted trauma memories in posttraumatic stress disorder. *Zeitschrift Fur Psychologie-Journal of Psychology, 218*(2), 141-145. doi:10.1027/0044-3409/a000021

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy, 38*(4), 319-345. doi:10.1016/S0005-7967(99)00123-0

Ehlers, A., Michael, T., Chen, Y. P., Payne, E., & Shan, S. (2006). Enhanced perceptual priming for neutral stimuli in a traumatic context: A pathway to intrusive memories? *Memory, 14*(3), 316-328. doi:10.1080/09658210500305876

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340-347. doi:10.1162/089892902317361886

Giesbrecht, T., Merckelbach, H., van Oorsouw, K., & Simeon, D. (2010). Skin conductance and memory fragmentation after exposure to an emotional film clip in depersonalization disorder. *Psychiatry Research, 177*(3), 342-349. doi:10.1016/j.psychres.2010.03.010

Halligan, S. L., Clark, D. M., & Ehlers, A. (2002). Cognitive processing, memory, and the development of PTSD symptoms: two experimental analogue studies. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(2), 73-89.

Halligan, S. L., Michael, T., Clark, D. M., & Ehlers, A. (2003). Posttraumatic stress disorder following assault: The role of cognitive processing, trauma memory, and appraisals. *Journal of Consulting and Clinical Psychology, 71*(3), 419. doi:10.1037/0022-006X.71.3.419

Hammersley, R., & Read, J. D. (1996). Voice identification by humans and computers. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 117-152). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Huntjens, R. J., Wessel, I., Postma, A., van Wees-Cieraad, R., & de Jong, P. J. (2015). Binding temporal context in memory: Impact of emotional arousal as a function of state anxiety and state dissociation. *The Journal of Nervous and Mental Disease, 203*(7), 545-550. doi:10.1097/NMD.0000000000000325

Jackson, D. C., Mueller, C. J., Dolski, I., Dalton, K. M., Nitschke, J. B., Urry, H. L., . . . Davidson, R. J. (2003). Now you feel it, now you don't: Frontal brain electrical asymmetry and individual differences in emotion regulation. *Psychological Science, 14*(6), 612-617.

Kerstholt, J. H., Jansen, N. J., Van Amelsvoort, A. G., & Broeders, A. (2004). Earwitnesses: Effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology, 18*(3), 327-336. doi:10.1002/acp.974

Kindt, M., van den Hout, M., Arntz, A., & Drost, J. (2008). The influence of data-driven versus conceptually-driven processing on the development of PTSD-like symptoms. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(4), 546-557. doi:10.1016/j.jbtep.2007.12.003

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, *33*(1), 159-174. doi:10.2307/2529310

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological review, 97*(3), 377-395.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sidney: Psychology Foundation.

McNally, R. J., Lasko, N. B., Clancy, S. A., Macklin, M. L., Pitman, R. K., & Orr, S. P. (2004). Psychophysiological responding during script-driven imagery in people reporting abduction by space aliens. *Psychological Science, 15*(7), 493-497. doi:10.1111/j.0956-7976.2004.00707.x

Meyer, T., Quaedflieg, C. W. E. M., Giesbrecht, T., Meijer, E., Abiad, S., & Smeets, T. (2014). Frontal EEG asymmetry as predictor of physiological responses to aversive memories. *Psychophysiology, 51*, 853–865.

Meyer, T., Smeets, T., Giesbrecht, T., Quaedflieg, C. W. E. M., Girardelli, M. M., Mackay, G. R. N., & Merckelbach, H. (2013). Individual differences in spatial configuration learning predict the occurrence of intrusive memories. *Cognitive, Affective, & Behavioral Neuroscience, 13*(1), 186-196. doi:10.3758/s13415-012-0123-9

Miller, J. B., deWinstanley, P., & Carey, P. (1996). Memory for conversation. *Memory, 4*(6), 615-631. doi:10.1080/741940999

Öhman, L., Eriksson, A., & Granhag, P. A. (2013). Enhancing adults' and children’s earwitness memory: Examining three types of interviews. *Psychiatry, Psychology and Law, 20*(2), 216-229. doi:10.1080/13218719.2012.658205

Orr, S. P., McNally, R. J., Rosen, G. M., & Shalev, A. Y. (2004). Psychophysiologic reactivity: Implications for conceptualizing PTSD. In G. Rosen (Ed.), *Posttraumatic stress disorder: Issues and controversies* (pp. 101-126). Chichester: John Wiley & Sons.

Orr, S. P., & Roth, W. T. (2000). Psychophysiological assessment: Clinical applications for PTSD. *Journal of Affective Disorders*, *61*(3), 225-240.

Quaedflieg, C. W. E. M., Schwabe, L., Meyer, T., & Smeets, T. (2013). Time dependent effects of stress prior to encoding on event-related potentials and 24h delayed retrieval. *Psychoneuroendocrinology, 38*(12), 3057-3069. doi:10.1016/j.psyneuen.2013.09.002

Roberts, L. (2011). Acoustic effects of authentic and acted distress on fundamental frequency and vowel quality. *Proceedings of the International Congress of Phonetic Sciences, 17*, 1694-1697.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*(9), 1043-1056. doi:10.1037/0003-066X.45.9.1043

Saslove, H., & Yarmey, A. D. (1980). Long-term auditory memory: Speaker identification. *Journal of Applied Psychology, 65*(1), 111-116. doi:10.1037/0021-9010.65.1.111

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*(1), 289-318. doi:10.1146/annurev.psych.49.1.289

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34. doi:10.1037/0096-3445.117.1.34

Sündermann, O., Hauschildt, M., & Ehlers, A. (2013). Perceptual processing during trauma, priming and the development of intrusive memories. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(2), 213-220. doi:10.1016/j.jbtep.2012.10.001

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063-1070. doi:10.1037/0022-3514.54.6.1063

Wolf, O. T. (2009). Stress and memory in humans: Twelve years of progress? *Brain Research, 1293,* 142-154. doi:10.1016/j.brainres.2009.04.013

Zoellner, L. A., Foa, E. B., Brigidi, B. D., & Przeworski, A. (2000). Are trauma victims susceptible to "false memories?". *Journal of Abnormal Psychology, 109*(3), 517. doi:10.1037/0021-843X.109.3.517

Table 1. Means (± standard errors) for voice recognition performance.

	Encoding instruction			
	Data-driven		Conceptually-driven	
	Neutral	Negative	Neutral	Negative
Recognition (% hits)	31.0 (5.1)	65.8 (5.8)	28.4 (4.6)	68.0 (5.6)
False alarms (% FA)	8.3 (2.2)	17.5 (3.4)	4.5 (1.6)	10.5 (2.4)
Discrimination (Pr)	.22 (.04)	.46 (.05)	.24 (.04)	.55 (.05)
Bias (Br)	.17 (.04)	.41 (.06)	.11 (.03)	.33 (.05)

Table 2. Means (± standard errors) for free recall and script memory performance.

	Encoding instruction				
	Data-driven	Conceptually-driven	*t* (*df* = 82)	*d*	*p*
Free recall					
Verbatim details	7.8 (0.5)	9.0 (0.7)	1.26	0.28	.211
Gist details	20.3 (1.2)	23.9 (1.1)	2.24	0.49	.028
Sounds	2.6 (0.2)	2.4 (0.2)	-1.19	-0.26	.238
Fabrications	1.1 (0.2)	0.8 (0.2)	-1.02	-0.22	.312
False details	0.9 (0.2)	0.5 (0.1)	-2.01	-0.44	.048
Accuracy	94.1 (0.7)	96.6 (0.5)	2.72	0.59	.008
Script memory test					
Recognition (% Hits)	82.2 (1.9)	83.3 (1.9)	0.41	0.09	.682
False alarms (% FA)	9.4 (1.3)	10.0 (1.1)	0.40	0.09	.692
Discrimination (Pr)	.68 (.02)	.69 (.02)	0.15	0.03	.885
Bias (Br)	.40 (.04)	.43 (.03)	0.52	0.11	.605

Table 3. Correlations between startle responses and speaker recognition

	Startle potentiation (perpetrator minus foil)	
	Neutral trials	Negative trials
Neutral trials		
Discrimination (Pr)	.24*	
Recognition (% hits)	.24*	
False alarms (% FA)	.07	
Bias (Br)	.18	
Negative trials		
Discrimination (Pr)		.03
Recognition (% hits)		.00
False alarms (% FA)		-.04
Bias (Br)		-.06

Note. The recognition performance scores were mean-centred per voice actor prior to the analyses. * $p < .05$.

Table 4. Correlations between startle responses and script memory

	Startle potentiation (perpetrator minus foil)	
	Neutral trials	Negative trials
Free recall		
Verbatim details	-.11	.31**
Gist details	.02	.20
Sounds	-.08	.17
Fabrications	.12	-.12
False details	-.02	.11
Accuracy	-.09	.12
Script memory test		
Recognition (% Hits)	-.07	.25*
False alarms (% FA)	-.16	-.06
Discrimination (Pr)	.03	.26*
Bias (Br)	-.13	.14

Note. The recognition performance scores were mean-centred per voice actor prior to the analyses. * $p < .05$; ** $p < .01$.