Learning and Retrieval of Crossmodally Associated Valence in Face Perception

Dissertation

Zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen
im Promotionsprogramm Behavior and Cognition (BeCog)
der Georg-August University School of Science (GAUSS)

vorgelegt von

Annika Carina Ziereis

aus Kempten (im Allgäu)

Göttingen, 2023

doi: 10.53846/goediss-9841

Betreuungsausschuss

Prof. Dr. Anne Schacht, Affektive Neurowissenschaft und Psychophysiologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Arezoo Pooresmaeili, MD, PhD, Forschungsgruppe Perception and Cognition, Europäisches Neurowissenschaftliches Institut, Göttingen

Prof. Dr. Alexander Gail, Kognitive Neurowissenschaften, Forschungsgruppe Sensomotorik, Universität Göttingen

Mitglieder der Prüfungskommission

Referentin: Prof. Dr. Anne Schacht, Affektive Neurowissenschaft und Psychophysiologie, Georg-Elias-Müller-Institut für Psycholgie, Universität Göttingen

Korreferentin: Arezoo Pooresmaeili, MD, PhD, Forschungsgruppe Perception and Cognition, Europäisches Neurowissenschaftliches Institut, Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Alexander Gail, Kognitive Neurowissenschaften, Forschungsgruppe Sensomotorik, Universität Göttingen

Dr. Tanya Behne, Kognitive Entwicklungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. York Hagmayer, Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie, Universität Göttingen

Prof. Dr. Hansjörg Scherberger, Neurobiologie, Deutsches Primatenzentrum, Göttingen

Tag der mündlichen Prüfung: 16.03.2023

doi: 10.53846/goediss-9841

Es genügt nicht bloss die bewusste Absicht dazu, jetzt mit dem einen Auge zu sehen, dann mit dem anderen, sondern man muss sich eine möglichst deutliche sinnliche Vorstellung hervorrufen von dem, was man zu sehen wünscht. Dann tritt dies auch in der Erscheinung hervor.

— Hermann von Helmholtz über die Gesichtswahrnehmungen (1871)

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Acknowledgments

First, I would like to thank Prof. Dr. Anne Schacht. Your charismatic nature and passion for research, your guidance through crises with optimism and confidence, and your strength and approachability have inspired and motivated me since the beginning of this journey. You gave me the freedom to follow my interests and yet managed to keep me on the right track. You have supported me during all phases of the dissertation but most importantly, your trust in me and appreciation of my work encouraged me to also believe in myself. There is so much I can learn from you.

Second, I would like to thank Dr. Arezoo Pooresmaeili who immediately agreed to be on my thesis committee and to review this thesis. Your reliability and quick reactions were much appreciated in the (sometimes) stressful life of a PhD student. But more importantly, it was for your passion for basic experimental research that truly changed my way of thinking and made me a bit humbler, which I highly value.

Third, I would like to thank Prof. Dr. Alexander Gail for his willingness to be in my thesis committee. You supported my PhD project by asking the right questions at the right time and giving helpful hints on all levels (content, organization, and pragmatics). I would also like to thank Prof. York Hagmayer, Dr. Tanya Behne, and Prof. Dr. Hansjörg Scherberger for their willingness to be members of my examination board.

A huge thanks to Fra, Yasi, Adi, Ronja, Anna, and Sri, the best colleagues one can imagine. Without you, I would have been lost! With you, I had a lot of fun which made everything feel less heavy. With you, I had valuable and (sometimes very) intense discussions. And the best of all is that in you, I have found real friends!

Thanks to all former and current members of the ANaP Lab, to the incredibly smart, reliable, dedicated, and nice students and student assistants: Jantje and Ronja, from whom I learned a lot about the topic and the role as a supervisor. Mareike, Esther, Alex, Marlene, Max, Jan Ole, Patrick, Lina, Lena, Caro, Flo, Emma, Hanna, Yasmin, and Bediz for helping in collecting high-quality data. I truly enjoyed working with you!

Thank you, Dr. Rebecca Jürgens and Ute Dorra, for solving so many problems (e.g., the Vienna case), being kind and patient, and helping me with the sometimes overwhelming administrative issues. You are exceptional!

Special thanks to Dr. Peter Zezula and Marc Reichhardt. This project would not have been possible without your quick and competent technical support. (Thanks also for the fantastic coffee and the fun soldering session)

doi: 10.53846/goediss-9841

Thank you to my graduate RTG 2070/GEMI friends and colleagues, for “Schwimmmontag” and Plätzchenbacken, Picknicks, and Roadtrips. It felt good to be in this together. Ben, you are a true friend and a fantastic person. Thank you for the guitar sessions, the ”Grannika song”, Laredo, and for being there during a particularly delicate phase of my PhD.

Thanks to my family and friends outside the university, for their understanding and unconditional support. And from the bottom of my heart, I want to say thank you, Christoph, even though words cannot express how grateful I am. You have supported me, endured my ups and downs, encouraged me, practiced math with me, picked me up from work late at night, listened to me, and (thankfully) sometimes disagreed with me. You were there when I needed you the most.

doi: 10.53846/goediss-9841

How someone's face is perceived is not only influenced by their facial expression but also largely by the situational context as well as by previous experiences with that person. Although emotional expressions of faces and voices are relevant social cues that have been suggested to enjoy a prioritized role in attentional selection processes, the connection between processing inherent emotional and associated valence is yet not well understood. In this PhD project, a set of behavioral and event-related potential (ERP) studies were conducted to examine the temporal dynamics and degree to which affective social cues are prioritized over neutral social cues and learned under different task constraints. In *Studies 1* to *3*, faces were cross-modally associated with affective vocalizations. Associated valence in learning and retrieval were tested in a valence-implicit Pavlovian conditioning paradigm (*Studies 1* and *2*) in which only gender information was task-relevant. In *Study 3*, the retrieval of faces previously associated with valence was contrasted for a valence-implicit and valence-explicit task. The influence of physical stimulus properties, e.g., frequency spectra and size, on valence effects was addressed in *Studies 4* and *5*. Although the neural (*Study 1*) and behavioral (*Study 2*) results suggested sensitivity for the voices' valence, there was little evidence for the acquired valence effects in the conditioned faces. In contrast, by relaxing task constraints during the learning of the face-voice pairs (*Study 3*), effects of associated valence were observable in both valence-implicit and -explicit tasks during retrieval. Effects on early visual processing were not observable for emotional stimuli but shown for the extensively trained stimulus features (gender) in *Study 1*. At mid-latencies, both positive and negative facial expressions affected ERPs (*Study 5*), whereas the effects of associated valence were restricted to negative associations. Valence effects on later processing appeared particularly sensitive to task requirements (*Studies 3* and *5*). In summary, the findings suggest a differential prioritization of valence in emotional expressions of the face and voice compared to associated valence. Moreover, tasks steering attention away from the stimulus' valence might strongly impair the acquisition of valence-based associations, whereas the retrieval of already acquired associations, similar to inherent emotional expressions, can influence attentional processes also in valence-implicit contexts.

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Zusammenfassung

Wie das Gesicht einer Person wahrgenommen wird, hängt nicht nur von ihrem Gesichtsausdruck ab, sondern auch weitgehend vom situativen Kontext sowie von früheren Erfahrungen mit dieser Person. Obwohl Emotionsausdrücke in Gesicht und Stimme relevante soziale Reize sind, von denen angenommen wurde, dass sie eine vorrangige Rolle bei selektiven Aufmerksamkeitsprozessen spielen, ist der Zusammenhang zwischen der Verarbeitung inhärenter Emotion und assoziierter Valenz noch nicht gut verstanden. Im Rahmen dieser Dissertation wurden eine Reihe von Verhaltensstudien und Studien zu ereigniskorrelierten Potentialen (EKPs) durchgeführt, welche die zeitliche Dynamik und das Ausmaß untersuchen, in dem affektive gegenüber neutralen, sozialen Reizen unter verschiedenen Aufgabenbedingungen priorisiert und gelernt werden. In den *Studien 1* bis *3* wurden Gesichter modalitätsübergreifend mit affektiven Vokalisationen assoziiert. Die assoziierte Valenz beim Lernen und Abrufen wurde in einem valenz-impliziten, klassischen Konditionierungsparadigma getestet (*Studien 1* und *2*), in welchem nur die Geschlechterinformation aufgabenrelevant war. In *Studie 3* wurde der Abruf von Gesichtern, die zuvor mit Valenz assoziiert wurden, in einer valenz-impliziten und einer valenz-expliziten Aufgabe verglichen. In den *Studien 4* und *5* wurde der Einfluss physikalischer Stimuluseigenschaften, z.B. Frequenzspektrum und Größe, auf Valenz-Effekte untersucht. Obwohl die neuronalen (*Studie 1*) und verhaltensbezogenen (*Studie 2*) Ergebnisse auf eine Sensitivität für Valenz in der Stimmverarbeitung hindeuteten, gab es kaum Hinweise auf erworbene Valenz-Effekte für konditionierte Gesichter. Im Gegensatz dazu waren die Effekte von assoziierter Valenz sowohl bei valenz-impliziten als auch -expliziten Aufgaben während des Abrufs zu beobachten, wenn keine Fokussierung auf spezifische Stimuluseigenschaften während des Lernens der Gesichter-Stimmen-Paare durch die Aufgabe vorgegeben wurde (*Studie 3*). Frühe visuelle Verarbeitungseffekte waren für emotionale Stimuli nicht zu beobachten, zeigten sich aber für die intensiv trainierten Stimulusmerkmale (Geschlechterinformation) in *Studie 1*. EKPs mittlerer Latenz wurden sowohl durch positive als auch negative Gesichtsausdrücke moduliert (*Studie 5*), während die Effekte assoziierter Valenz auf negative Assoziationen beschränkt waren. Valenzeffekte auf spätere Verarbeitungsstufen schienen besonders empfindlich auf die spezifischen Aufgabenanforderungen zu reagieren (*Studien 3* und *5*). Zusammenfassend deuten die Ergebnisse auf eine unterschiedliche Priorisierung von Valenz bei Emotionsausdrücken in Gesicht und Stimme im Vergleich zu assoziierter Valenz hin. Darüber hinaus scheinen Aufgaben, die die Aufmerksamkeit von der Valenz des Stimulus ablenken, den Erwerb valenzbasierter Assoziationen stark zu beeinträchtigen, während der Abruf bereits erworbener Assoziationen, ähnlich wie bei inhärenten emotionalen Ausdrücken, auch in valenz-impliziten Kontexten Aufmerksamkeitsprozesse beeinflussen kann.

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Chapter 1

Introduction

Learning about others is essential for the coordination of social interactions, and thus it is a vital mechanism for all social species, including humans. Knowledge about regularities in others' behavior and outcomes of previous interactions can help us to predict their future behavior and to respond to it accordingly. An important source of social information and signals is other people's faces (Haxby et al., 2000). If we are unsure whether the person walking in front of us is our friend, we usually want to confirm our hypothesis by seeing their face. Recognizing familiar faces, considering their complexity and similarity, requires comparatively little time and is only one of several fast cognitive processes associated with face perception. Every time we encounter other people, a cascade of cognitive processes is running, of which many are extremely fast and based on heuristics, e.g., people can decide about facial attributes like trustworthiness and aggressiveness at very short presentation and, remarkably, these quick assessments do not change significantly with longer presentation times (Todorov et al., 2015). Some information even seems automatically activated when seeing a face, e.g., with only little effort, we process and infer attributes that are inherent in the other person's face, including social dimensions such as gender and age (Dobs et al., 2019) and variable ones such as their facial expression (for a review, see Palermo & Rhodes, 2007). However, other information needs to be retrieved more intentionally, e.g., retrieving a person's name (Borghesani et al., 2019).

An influential cognitive model of face perception was brought forward by Vicki Bruce and Andy Young (1986). The authors proposed independent processing of identity and emotional features based on findings of cognitive and neuropsychological studies of healthy individuals and prosopagnosic individuals. Briefly summarized, the model predicts hierarchical processing, starting with the structural encoding of the face, generating a view-centered representation which is the basis for the two separate tasks, namely the analysis of changing properties (as expression and speech analysis) and an abstract representation of the face (independent of pose and expression)

doi: 10.53846/goediss-9841

for identity recognition. Identity recognition can be separated into three different chunks, with the lowest level comparing face structure with face recognition units. If face recognition units match the structure, person-related knowledge will be activated, which can lastly lead to higher-order retrieval, such as the generation of the person's name (for a review, see Young, 2018). With the advances in neuroimaging research, the findings about the neuroanatomy underlying face processing became remarkably close to the proposed model of Bruce and Young, although newer findings suggest that identity and facial expressions are not processed completely independently (e.g., Atkinson et al., 2005; Kaufmann & Schweinberger, 2004). There have been several face-sensitive areas defined: the fusiform face area and a face-selective area in the posterior part of the superior temporal sulcus (Kanwisher et al., 1997), the occipital face area located in the inferior occipital gyrus (Haxby et al., 1999), and, more recently, also anterior areas, e.g., the anterior temporal lobe (e.g., Tsao et al., 2008), the anterior superior temporal sulcus (Pitcher et al., 2011) which responded strongly to dynamic faces, and the inferior frontal gyrus (Fox et al., 2009). For a review of face-selective areas and their functional role, see Duchaine and Yovel (2015).

However, factors outside the face are largely ignored in this model, such as the respective situational context and our expectations which can modulate both face recognition (Watkins et al., 1976) and face perception. For example, one might not immediately recognize colleagues from work when running into them while on vacation (Brown, 2020) or if they have a new haircut (Frowd et al., 2012). Context has important implications not only on the recognition of faces but also on *how* faces are perceived (for a review, see M. J. Wieser & Brosch, 2012), including the perceiver's "internal context", such as their mood (e.g., An & Hsiao, 2021; Bouhuys et al., 1995; Curby et al., 2012; Forgas & East, 2008; Hills et al., 2011; Hills & Lewis, 2011), attentional capacity (for a review, see Palermo & Rhodes, 2007) and task-goals (e.g., Rellecke et al., 2012a; Yan et al., 2019). Although context effects on perception have also been established for other types of stimuli (e.g., Gerdes et al., 2013; Jamal et al., 2017; Leo et al., 2011; Shams et al., 2000; Todorović, 2010), arbitrary facial expressions, such as non-expressive neutral faces and faces expressing surprise seem particularly susceptible for contextual information (Cooney et al., 2006; Hester, 2019; Yoon & Zinbarg, 2008), possibly due to the high social relevance of inferring about other people's intentions.

In line with this, person perception outside the laboratory is usually not restricted to the visual domain but occurs in a multimodal and dynamic fashion (Jessen & Kotz, 2013; Schirmer & Adolphs, 2017). There is evidence for efficient and supra-additive multi-sensory integration, particularly for facial and vocal information (Pan et al., 2017; Pourtois & Dhar, 2012; Watson et al., 2014), although it seems not independent of competing attentional effects (Dessel & Vogt, 2012;

doi: 10.53846/goediss-9841

Ho et al., 2015). The integration of vocal information is tightly bonded to, e.g., the perception of gender (Masuda et al., 2005; Smith et al., 2007) and the processing of facial expressions (e.g., Dolan et al., 2001; Gelder & Vroomen, 2000; Kokinous et al., 2014; V. I. Müller et al., 2011; Skuk & Schweinberger, 2013; Watson et al., 2013).

Given the many levels of social and emotional information related to a person's face, voice, and situational context, and since our cognitive system is limited, one might wonder how information is prioritized and what information is learned. Theories of attention aim to describe and explain selection processes for information to become available to sensation (e.g., Lachter et al., 2004; Treisman, 1964). Which social information is preferentially processed and learned can be moderated by different factors, such as stimulus-driven effects (e.g., biologically "hard-wired" features that don't require learning or experimental manipulation to capture attention, e.g., Öhman et al., 2001; Pool et al., 2016) and goal-directed effects (e.g., directing the focus to an anticipated stimulus location; Eimer, 2000; Hillyard & Anllo-Vento, 1998; Russo, 2003; or a specific stimulus feature; Hillyard & Anllo-Vento, 1998; Summerfield & Egner, 2016; Zani & Proverbio, 1995). The dichotomous account of stimulus-driven and goal-directed effects has been frequently criticized, partly for the difficulty in truly separating both processes (as highlighted by B. Anderson, 2011; Theeuwes, 2013; Yantis & Jonides, 1990) and partly for not fully capturing phenomena like memory-driven effects of previously learned selection-strategies ("attentional sets," e.g., Kim & Anderson, 2019; Leber et al., 2009), or value-driven effects (e.g., B. A. Anderson et al., 2011; B. A. Anderson & Halpern, 2017; Antono et al., 2022; Hammerschmidt et al., 2017; Lunghi & Pooresmaeili, 2023; Pooresmaeili et al., 2014; Rossi et al., 2017; Schacht et al., 2012). A third class of attentional processes, experience-driven attention, was instated by B. A. Anderson et al. (2011), describing the mechanism for previous experience exerting direct influence on attentional control (e.g., through selection history and associations with rewarding or aversive outcomes).

To understand attentional processes in social cognition, independent of what drives attention, it is also essential to identify the time scales of the attentional processes, i.e., whether they reflect early selection through sensory gain processes (e.g., Hillyard et al., 1998) or later selection through post-perceptual processes (e.g., Fockert et al., 2001). Event-related potentials (ERPs), with their high temporal resolution, are an excellent tool for measuring the dynamics of larger-scaled changes in cortical activity. ERPs are time-locked and averaged voltage changes measured at the scalp's surface, predominately reflecting the postsynaptic, summed activity of pyramidal neurons in the cortex (S. J. Luck, 2005). Typical ERP components related to face perception are the P1, N170, EPN, and LPC. The P1, a bilateral occipital positivity generated from the extrastriate cortex (Hillyard & Anllo-Vento, 1998; Russo, 2003), peaks around 100 ms after the onset of a

doi: 10.53846/goediss-9841

visual stimulus and has shown enhanced amplitudes for stimuli presented in an attended location (Clark & Hillyard, 1996; S. J. Luck et al., 2000; Pourtois, 2004). Some studies reported P1 effects for emotional compared to neutral facial expressions (e.g., Batty & Taylor, 2003; Bublatzky et al., 2014; Foti et al., 2010; Hammerschmidt et al., 2017; Müller-Bardorff et al., 2018; Rellecke et al., 2011; Valdés-Conroy et al., 2014), and faces compared to other objects (Neumann et al., 2011). Since its enhancement was inconsistently found for faces with emotional expressions (for a review, see Schindler & Bublatzky, 2020) and it is known to be particularly sensitive to physical stimulus properties (e.g., Allison, 1999; De Cesarei et al., 2013), the underlying causes for P1 differences between physically different stimuli (such as emotional compared to neutral expressions) have been seen as controversial (e.g., Schindler, Bruchmann, Gathmann, et al., 2021). The P1 is followed by the N170 component, a negative deflection over occipito-temporal regions, peaking around 170 ms. Unlike its well-replicated sensitivity for faces compared to other objects (Bentin et al., 1996; Bentin & Deouell, 2000; Rossion et al., 2000), also N170 modulations by face familiarity (Caharel & Rossion, 2021) have been reported, suggesting that structural processing and the face identification of well-known faces are temporally close to each other. Although comparatively more often than effects on the P1, modulations of the N170 by emotional expressions have only been inconsistently found (see, e.g., Hinojosa et al., 2015; Rellecke et al., 2012b). In contrast, the early posterior negativity (EPN; Schupp et al., 2006), a relative negativity over occipitotemporal regions, has been related to the selective differentiation between affective and neutral stimuli and has been found robustly across several stimulus domains (e.g., Bayer & Schacht, 2014; Schacht & Sommer, 2009). For faces with emotional expressions, it is most pronounced around 200–300 ms (e.g., Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017; Weidner et al., 2022). Longer latency ERPs such as the late positive complex/potential (LPC/LPP), starting from 300 ms with a more parietal distribution, have been demonstrated to be augmented by facial expressions, particularly for aversive expressions (e.g., Bayer & Schacht, 2014; Recio et al., 2014; Schacht & Sommer, 2009; Schupp et al., 2004). However, later processing seems more strongly affected by specific task requirements (e.g., Rellecke et al., 2012a; Schacht et al., 2008).

Despite the temporal variation in the unfolding of emotional information in expressions of the voice, attentional effects between affective and neutral sounds on auditory processing have been reported for semantic speech (e.g., Kotz & Paulmann, 2007) and particularly for rapidly unfolding expressions like affect bursts (e.g., Pell et al., 2015; Scherer, 2013) and prosody (e.g., Paulmann et al., 2013). Similar to the visual domain, earlier auditory ERPs, e.g., the N1, are held to be rather stimulus-driven, whereas the auditory P2 has been related to early relevance-detection. In contrast to the P2 (e.g., T. Liu, Pinheiro, Deng, et al., 2012; Paulmann et al., 2013; Sauter & Eimer, 2010;

doi: 10.53846/goediss-9841

Schirmer et al., 2012; Spreckelmeyer et al., 2009), emotion effects have been less consistently reported for the N1 (e.g., Iredale et al., 2013; T. Liu, Pinheiro, Deng, et al., 2012; Pell et al., 2015).

Especially the question about unintentional or automatic processing of emotional expressions of the face and the voice compared to other kinds of stimuli has remained controversial (e.g., M. M. Müller et al., 2007; Pessoa, McKenna, et al., 2002; Pourtois et al., 2013; Straube et al., 2011; Vuilleumier, 2005; Whalen et al., 1998), possibly due to conceptual questions about the underlying mechanism behind early effects, e.g., whether certain low-level features such as spatial frequency of pitch characteristics might serve as cues for early but not for late selection. However, the general potential of emotional stimuli to "capture attention" has led some researchers to consider that emotional attention might form an extra class of attention processes that can occur on several levels, i.e., stimulus-driven and goal-directed, although it was proposed that they are more similar to early, stimulus-driven attentional processes (for a review, see Pourtois et al., 2013). Whereas it might be methodologically difficult to resolve to which degree physical stimulus characteristics modulate early effects of emotional stimuli, research on associative learning has shown that also early processing is influenced by stimulus valence. Beyond methodological advantages, also content-related questions about the similarities and differences between the processing of inherent hedonic stimuli and stimuli that acquired valence through learning have been addressed: Importantly, not only do expressions of emotions carry relevant social information but also knowledge and previous experience with a person do. ERP research has provided evidence for modulations of face processing for various types of context information associated to faces, ranging from descriptions about a person (e.g., Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017; Luo et al., 2016; Suess et al., 2014; Xu et al., 2016), and verbal statements of a person (M. J. Wieser, Gerdes, et al., 2014), over more abstract hedonic value such as monetary reward and loss (Hammerschmidt, Kagan, et al., 2018; Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017), to stimuli presented in different sensory channels such as unpleasant odor (Steinberg et al., 2012), aversive noise bursts (Watters et al., 2018) and fear-conditioning research (Camfield et al., 2016; Rehbein et al., 2014; Schellhaas et al., 2020; Sperl et al., 2021; Wiemer et al., 2021; for a review see Miskovic & Keil, 2012).

Compared to the relatively large amount of research on the interplay between attentional processes and inherent expressions of emotion (for reviews, see Pourtois et al., 2013; Schindler & Bublatzky, 2020), only a few studies included expressions of emotion as unconditioned stimuli in associative learning paradigms (e.g., Haddad et al., 2013). This is remarkable, given the reports on early and involuntary processing of emotional expressions suggesting them to be effective unconditioned stimuli (US). Moreover, since faces and voices co-occur with high probability in natural

doi: 10.53846/goediss-9841

communicative situations, they should form stimulus pairs with higher biologically shared relevance ("belongingness"; Garcia & Koelling, 1966; Rescorla, 2008), which has been shown to facilitate associative learning processes (Seligman, 1970).

The first ERP evidence for successful associations of faces with emotional expressions was provided by a study by Aguado and colleagues (2012), in which non-expressive faces (CS) were associated with images of the same identities expressing happiness or anger. Compared to neutrally-associated faces, anger- and happiness-associated faces showed enhanced P1 amplitudes, followed by a modulation of the N170 for anger-associated faces, i.e., both relatively early processing stages. In contrast, cross-modal associations of fearful screams on non-expressing faces modulated ERPs beginning at the N170, EPN, and LPC, as shown in a recent study by Bruchmann et al. (2021).

Unlike the overall evidence for experience-driven attention effects for associated faces, it is less clear what drives the effects on early selection and sensory gain, and post-perceptual selection. Although the first intuition might suggest an intramodal advantage for early selection by increased resource allocation to the same modality in order to optimize the processing of the following US, it fails to explain findings on very early effects of cross-modal associations (e.g., E. M. Mueller & Pizzagalli, 2015; Sperl et al., 2021; Steinberg et al., 2012; Steinberg, Bröckelmann, Dobel, et al., 2013; Steinberg, Bröckelmann, Rehbein, et al., 2013) and associations with verbal semantics (Morel et al., 2012) or monetary reward (Hammerschmidt et al., 2017). On the contrary, intra-modally presented stimuli might, under certain circumstances, compete stronger for attention compared to cross-modally presented stimuli (e.g., Duncan et al., 1997; Schupp et al., 2008; Soto-Faraco & Spence, 2002) and impair learning. Moreover, findings about multimodal integration of faces and voices (for a review, see Klasen et al., 2012) and affective priming studies (e.g., Garrido-Vásquez et al., 2018) suggest that information of the face and voice are easily, if not to some degree automatically, integrated and show strong cross-modal interactions, possibly because vocal expression and facial expression usually co-occur with high contiguity and have overlapping functional domains (e.g., Gelder & Vroomen, 2000; Shams et al., 2000; Wang et al., 2016).

Since competition between stimuli can attenuate learning, it seems reasonable to assume that also situational factors, such as the task(-demands), affect learning. There have only been a few associated learning studies with tasks in which valence or the conditioned category was irrelevant for learning (Abdel Rahman, 2011; Hammerschmidt, Kagan, et al., 2018) or retrieval (Hammerschmidt et al., 2017; Luo et al., 2016; Pooresmaeili et al., 2014; Rossi et al., 2017) or in which task-relevance was systematically manipulated (Bruchmann et al., 2021). A further important factor for learning is stimulus intensity (e.g., Bevins, 1997; Bradfield & McNally, 2008; Odling-Smee, 1975). Electric shocks or loud noise burst resemble strong aversive stimuli, which should

doi: 10.53846/goediss-9841

be learned with high priority. It is less clear whether the biological and social relevance suffices to associate valence of medium-intense emotional expressions. Moreover, there are functional properties of expressions of emotion in the face and voice that might prevent associative learning: One of the characteristic properties of emotional expressions is their inter- and intraindividual variability (people rarely always smile, nor do they always frown). In particular, the variability makes them a communicative tool, which is also reflected in the capability of the receiver to detect these changes, e.g., the sensitivity to the change of emotional expressions has been shown to be higher compared to invariant attributes such as identity (Taubert et al., 2016). It was a priori unknown whether faces could be associated with unconditioned stimuli which are known to be variable. Some contemporary theories of learning propose that attention is increased to uncertain stimuli with the internal goal of learning about yet unpredictable events in the environment, while others would predict increased attention to the most predictive cues. (for a discussion, see Pelley et al., 2012).

Aim and overview of this thesis

The present thesis aims at addressing these gaps by investigating the acquisition and retrieval of valence-based associations in faces, specifically in the context of cross-modal associative learning of medium-intense emotional stimuli, building upon and connecting findings on face perception and attentional effects of emotion and intention. Addressing the question of (in-)voluntary processing of associated valence, a focus of this thesis is the temporal dynamics of face processing, to elude at which point prioritization of emotional information might occur. The method of choice was event-related potentials (ERPs), but behavioral measures (response times and accuracy if applicable) and pupil size as a physiological measure were included to obtain a more global perspective on the effects of associated valence.

To test whether faces conditioned with emotional expressions of the voice would gain additional relevance even to the point where emotion was not relevant for the task, we pre-registered and conducted an *associative learning* study (*Study 1*) in which faces with neutral expressions were cross-modally conditioned with affect bursts of emotional or neutral valence. During learning and delayed test, only gender-related decisions on the presented stimuli had to be made. In both sessions, we recorded typical face- and emotion-sensitive ERP components (P1, N170, EPN, and LPC), pupil diameter, and behavior (response times, accuracy, and likability ratings of faces).

Study 2 was conducted to gain insights into the *learning dynamics* of the task-irrelevant cross-modal associations of emotion in faces. In this study, response times and accuracy of two inde-

doi: 10.53846/goediss-9841

pendent but similarly designed associative learning studies were compared. We used both a distributional and a mechanistic (drift-diffusion) approach to explore differences between emotion and task-relevant influences on behavior over the course of learning and extinction.

The focus of *Study 3* was on attention effects during the *retrieval* of faces associated with emotional valence. In this pre-registered study, a novel, flexible learning paradigm was applied in which participants studied face-voice pairs with the aim of correctly matching the faces and their corresponding neutral or emotional voices. In a delayed ERP test session, an emotion-implicit and an emotion-explicit task were performed on the previously learned faces to compare at which processing stages intentional and automatic effects of associated valence occur.

In addition to the three main studies, two studies were conducted not directly related to associative learning but either served as a supplementary study (Study 4) or emerged from the main studies (Study 5). The aim of *Study 4* was to validate different auditory scrambling methods applied to affect bursts in order to create reference stimuli with comparable low-level features but eliminating semantic or valence information, similar to scrambled visual stimuli. The potential benefits and pitfalls of using these stimuli were discussed. In *Study 5*, we tested the interaction of stimulus size and facial expression on early, mid-, and late face-sensitive ERP components. Since we did not find consistent modulations of early visual processing for emotion-based associations in Study 1 and Study 3 (cf. Aguado et al., 2012; Hammerschmidt et al., 2017; Muench et al., 2016; Schindler et al., 2022) we wanted to exclude the alternative explanation that the comparatively small size of the presented stimuli made stimuli less effective. To investigate at which processing stage size of face stimuli interacts with emotional information, as a first step, we tested faces with emotional expressions since they have been robustly shown ERP modulations, especially at mid-latency components (EPN, e.g., Hammerschmidt, Kagan, et al., 2018; Hammerschmidt, Kulke, et al., 2018; for a review, see Schindler & Bublatzky, 2020).

Chapters 2 and 3 were devoted to the investigation of learning of associated emotion in faces in an emotion-implicit setting. Chapter 4 addresses the question about motivated attention during the retrieval of associated valence in faces. Chapters 5 and 6 include the two supplementary studies of this project. The general discussion of the findings, implications, and concluding remarks are summarized in Chapter 7.

A brief note on terminology. Dipping into the research fields of emotion, attention, and learning, one challenge is the heterogeneous use of common terms and constructs or even an axiomatic assumption of their existence, as pointed out by researchers in the field (B. Anderson, 2021, 2011; B. A. Anderson, 2021; Cabanac, 2002; Duffy, 1941). If it was possible to overcome

doi: 10.53846/goediss-9841

this issue, it would clearly be beyond the scope of this thesis. Nevertheless, I attempted to provide working definitions or exemplary descriptions for clarification of what was investigated as part of this PhD project. For example, the working definition for *prioritized processing* for visually presented stimuli in this thesis is the increased allocation of resources, measured as *enhanced* neural or physiological activation, based on the early selection hypothesis (e.g., Treisman, 1964), which predicts an enhancement for attended stimuli. This becomes relevant at the point where automatic, involuntary processing, and capacity-free processing make different predictions about the measured neural signals, although, in the literature, all have been interpreted as a prioritization of emotional compared to neutral stimuli. Moreover, the operationalization of prioritization as differences in neuro-physiological activation for emotional vs. neutral conditions can include, but does not automatically imply, observable behavioral effects (e.g., response time differences or differences in accuracy). A dissociation between behavioral, neural, and physiological responses is not uncommon (e.g., Antov et al., 2020; Apergis-Schoute et al., 2014; Bublatzky et al., 2019; C. C. Luck & Lipp, 2015, 2016). Notably, the prioritization of emotional stimuli can have opposing outcomes on behavior, depending on the current (task-defined) goals of an individual and the outcome measure (Bradley et al., 2012): For instance, prioritization of stimuli classified as being emotional can lead to faster/more accurate responses in detection tasks (e.g., Öhman et al., 2001), but also result in the opposite, i.e., slower/inaccurate responses when emotion is not task-relevant (e.g., Schacht & Sommer, 2009).

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Chapter 2

Cross-modal affective learning of vocalizations requires task relevance of emotion: Evidence from electrophysiological and pupillary responses

Abstract

Social and emotional cues from faces and voices are highly relevant and have been reliably demonstrated to attract attention involuntarily. However, there are mixed findings as to which degree associating emotional valence to faces occurs automatically. In the present study, we tested whether inherently neutral faces gain additional relevance by being conditioned with either positive, negative, or neutral vocal affect bursts. During learning, participants performed a gender-matching task on face-voice pairs without explicit emotion judgments of the voices. In the test session on a subsequent day, only the previously associated faces were presented and had to be categorized regarding gender. We analyzed event-related potentials (ERPs), pupil diameter, and response times (RTs) of $N = 32$ subjects. Emotion effects were found in auditory ERPs and RTs during the learning session, suggesting that task-irrelevant emotion was automatically processed. However, ERPs time-locked to the conditioned faces were mainly modulated by the task-relevant information, i.e., the gender congruence of the face and voice, but not by emotion. Importantly, these ERP and RT effects of learned congruence were not limited to learning but extended to the test session, i.e., after removing the auditory stimuli. These findings indicate successful associative learning in our paradigm, but it did not extend to the task-irrelevant dimension of emotional relevance. Hence, associations of emotional relevance might not be completely automatic but seem to depend on context specifics such as task requirements.[1]

[1]This manuscript has been submitted for publication and is currently in revision.

doi: 10.53846/goediss-9841

2.1 Introduction

In the most everyday encounter with another person, we automatically extract a variety of information from their face (Haxby et al., 2000). To effectively behave in a social situation, it is equally important to recognize the other's emotions and intentions and to consider the current context and previous experience with this person. It has been demonstrated that the affective context of the same and other modalities modulates face perception (Aviezer et al., 2011; Hassin et al., 2013; McCrackin & Itier, 2018; for a review, see M. J. Wieser & Brosch, 2012). However, there are open questions about the conditions under which context is integrated and associated with faces, e.g., how people automatically gain knowledge about others and how this generalizes to other situations. Aiming to fill this research gap, we investigated whether the perception of novel and neutral faces changes when associated with task-irrelevant emotional context and whether these associations transfer to a different test setting. To this aim, we implemented a cross-modal associative learning paradigm and recorded emotion-sensitive event-related brain potentials (ERPs), pupil size changes, and behavioral measures.

Research on Pavlovian aversive conditioning has repeatedly shown that faces as conditioned stimuli (CS^+) can acquire negative valence when being paired with biologically aversive unconditioned stimuli (US), e.g., aversive odor (Steinberg et al., 2012), electric shocks (Rehbein et al., 2014) or loud noise bursts (Watters et al., 2018) as US (for a review: Miskovic & Keil, 2012). Neutral faces have also been shown to acquire positive valence, e.g., when associated with monetary reward (Hammerschmidt, Kagan, et al., 2018; Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017). The CS^+ faces can then evoke physiological reactions (CR) like changes in skin conductance, heart rate, pupil size, and enhanced neural processing, e.g., in evoked steady-state potentials (ssVEP, e.g., M. J. Wieser, Miskovic, et al., 2014), neural oscillations (e.g., Chen et al., 2021), and ERPs: To better understand the mechanisms underlying associative learning, several studies compared ERP modulations of conditioned faces to the typical effects of inherent emotional facial expressions, ranging from early sensory processing to higher cognitive evaluations.

The **P1** reflects early attention and usually peaks around 100 ms after stimulus onset with a bilateral occipital positivity generated from the extrastriate cortex (Hillyard & Anllo-Vento, 1998; Russo, 2003). It was found to be enhanced for emotional compared to neutral facial expressions (Bublatzky et al., 2014; Foti et al., 2010; Hammerschmidt et al., 2017; Rellecke et al., 2011), although other studies reported a lack of modulations of the P1 by emotional expressions (for a review, see Schindler & Bublatzky, 2020). The P1 is followed by the face-sensitive **N170** com-

doi: 10.53846/goediss-9841

ponent, a negative deflection over occipito-temporal regions, peaking around 170 ms, typically enhanced for faces compared to other objects (Bentin et al., 1996; Rossion et al., 2000). Similar to P1 modulations, emotion effects on the N170 have inconsistently been reported, potentially due to the use of different stimuli across studies, and thus variations in low-level visual factors like contrast (Bobak et al., 1987), size (Kornmeier et al., 2011; Yiannikas & Walsh, 1983), or luminance (Bieniek et al., 2013). Since differences in low-level features can be controlled by randomizing the CS-US pairing, both P1 and N170 components allow insights into the early processing of associated relevance without being confounded with perceptual stimulus features related to the face. The early posterior negativity (**EPN**), a relative negativity most pronounced around 200–300 ms over occipitotemporal regions, was reported robustly in several ERP studies for emotional versus neutral stimuli across several stimulus domains (e.g., Bayer & Schacht, 2014; Schacht & Sommer, 2009), and has been assumed to reflect the facilitation of sensory encoding and selective attention mechanisms (Schupp et al., 2006). Later ERP modulations like the **LPP/LPC** seem to be more strongly affected by specific task requirements (e.g., Rellecke et al., 2012a) but have reliably been demonstrated to be augmented by particularly facial expressions of aversive emotions (Schindler & Bublatzky, 2020; Schupp et al., 2004).

ERP findings on faces with associated relevance

There is a long tradition of conditioning research with faces serving as CS^+ and different kinds of (mostly aversive) US stimuli showing that face perception can change at different processing stages (for a review: Miskovic & Keil, 2012). Aside from the described ERP components, few studies reported effects on early processing stages, observable already before 100 ms after the CS^+ onset (Morel et al., 2012; E. M. Mueller & Pizzagalli, 2015; Sperl et al., 2021; Steinberg et al., 2012; Steinberg, Bröckelmann, Dobel, et al., 2013; Steinberg, Bröckelmann, Rehbein, et al., 2013), and, remarkably, for different kinds of unconditioned stimuli (US), like odor, auditory startle, and electric shocks (Steinberg, Bröckelmann, Rehbein, et al., 2013). Typical ERP modulations for Pavlovian or evaluative conditioning and instrumental learning paradigms were reported at latencies from 100 ms on, during short- (P1, N170), mid- (EPN), and long- (LPC) latencies. Enhanced amplitudes for faces with associated relevance have been reported for the **P1** (monetary reward: Hammerschmidt et al., 2017; facial emotional expressions: Aguado et al., 2012), **N170** (fear-conditioning: Camfield et al., 2016; Schellhaas et al., 2020; Sperl et al., 2021; aversive screams: Bruchmann et al., 2021; person knowledge: Luo et al., 2016; Schindler, Bruchmann, Krasowski, et al., 2021; facial emotional expressions: Aguado et al., 2012), **EPN** (fear-conditioning: Bruchmann et al., 2021; Sá et al., 2018; Schellhaas et al., 2020; person knowledge: Abdel Rah-

doi: 10.53846/goediss-9841

man, 2011; Luo et al., 2016; Suess et al., 2014; Xu et al., 2016; affective communication: M. J. Wieser, Gerdes, et al., 2014) and the **LPC** (fear-conditioning: Panitz et al., 2015; Rehbein et al., 2018; Sá et al., 2018; Sperl et al., 2021; Wiemer et al., 2021; aversive screams: Bruchmann et al., 2021; monetary reward: Hammerschmidt, Kulke, et al., 2018; person knowledge: Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017; Schindler, Bruchmann, Krasowski, et al., 2021; Xu et al., 2016). Despite this evidence, there is still uncertainty under which boundary conditions associated effects occur, e.g. regarding the need for explicit awareness of a CS-US contingency for stable associations (Mertens & Engelhard, 2020). Implicit conditioning usually refers to the subliminal presentation of the CS^+ and not the US. Furthermore, many studies use salient aversive stimuli like electric shocks as US or explicit instructions to draw attention to the contingency between the CS^+ and the US, e.g., by picturing actions or encounters of the CS^+ face (Aguado et al., 2012; Verosky et al., 2018). Moreover, very few studies investigated whether and how robustly emotional contextual information becomes (automatically) associated even when this information is not task-relevant. Task-irrelevant emotional stimuli have been supposed to capture attention (J. Armony, 2002; Morris et al., 1998; Öhman et al., 2001), in bottom-up or top-down ways, respectively, which is a prerequisite for automatic associations. In this line, amygdala activations have been reported for emotional visual stimuli, even when emotion was not task-relevant, but only if the task load was not too high (Pessoa, McKenna, et al., 2002; Pessoa, Kastner, et al., 2002). In contrast, emotional auditory stimuli appear to be more robust against distractors as long as the attentional focus stays within the auditory modality (e.g., Ethofer et al., 2006; Quadflieg et al., 2008; Sander et al., 2005; cf. Bach et al., 2008). However, there is not much known about the cross-modal transfer of emotional information, particularly when features other than the emotional content of the US are relevant for the task during associative learning.

Aim of the study

The present study aimed to fill this gap and specifically tested whether inherently neutral faces gain additional relevance when being associated with either positive or negative compared to neutral vocal affect bursts when the emotion of the burst is not task-relevant. We recorded ERPs and pupil size during a learning and a delayed test phase to investigate the temporal dynamics of the acquisition and extinction of the associated reactions. Based on previous research (Hammerschmidt, Kagan, et al., 2018; Hammerschmidt et al., 2017; Schacht et al., 2012; Sperl et al., 2021), we set an approximate 24-hour break between learning and test to allow for memory consolidation (Menz et al., 2016; Pace-Schott et al., 2015; Sopp et al., 2017). We chose to associate neutral faces with vocal affect bursts because faces and voices are considered socially and

doi: 10.53846/goediss-9841

biologically relevant, naturally co-occurring and usually integrated into a whole-person percept (Freeman & Ambady, 2011a). Neutral facial expressions are more likely to be perceived as ambiguous (Schwarz et al., 2012; M. J. Wieser, Gerdes, et al., 2014; Yoon & Zinbarg, 2008) and suit well as CS^+ stimuli (Bublatzky et al., 2020). We presented vocal affect bursts as US, as they do not have the segmental structure of speech or pseudo speech, are relatively short in length, and unfold the emotional information rapidly. To ensure active processing of the auditory US, we presented gender-matching or gender-mismatching face-voice pairs during learning, where participants performed gender-congruence decisions. Since face-voice pairing and thus gender congruence was fixed for every trial, we randomly interspersed no-go trials (beep sound instead of the US) to counteract cross-modal inhibition (Johnson & Zatorre, 2006) and responses solely based on the target face after its repeated presentation.

Hypotheses

Our overall hypothesis was that inherently neutral faces acquire emotional relevance through learned associations with affective (CS^+: ${CS^+}_{pos}$, ${CS^+}_{neg}$) but not neutral (CS_{neu}) vocal bursts. Emotional relevance was operationalized in terms of differential neural responses towards the faces as a function of learning and extinction and different speed and accuracy measures when performing a gender matching (learning session) or gender (test session) decision.

During **learning**, we expected slower responses and lower accuracy for emotional (CS^+: ${CS^+}_{pos}$, ${CS^+}_{neg}$) compared to neutral (CS_{neu}) face-voice pairs according to findings about the attentional binding of emotional information (A. K. Anderson, 2005; Gutiérrez-Cobo et al., 2019) that slow down tasks in which other information has to be processed (Schacht & Sommer, 2009; X. Zhang et al., 2019; cf: Roesch et al., 2010) and due to the emotional incongruence between the neutral faces and emotional voices (Föcker et al., 2011). Higher sensitivity to threatening stimuli (Öhman et al., 2001) should additionally lead to lower accuracy for angry (${CS^+}_{neg}$) compared to happy (${CS^+}_{pos}$) or neutral (CS_{neu}) face-voice pairs. Furthermore, we expected slower RTs and lower accuracy for gender-mismatching compared to -matching face-voice pairs since the conflicting input signals between both modalities would interfere with an automatic and integrated perception of gender (Freeman & Ambady, 2011a). Heightened arousal and attention should increase the pupil size (e.g., Cosme et al., 2021; Kret et al., 2013), triggered by the CS^+ faces predicting an emotional auditory burst during learning (for a review on pupil dilation in conditioning, see Finke et al., 2021). On the neural level, the gain of CS^+ faces should modulate early processing (P1) and subsequent processing stages (i.e., N170, EPN, LPC). We expected enhanced P1 amplitudes for

doi: 10.53846/goediss-9841

CS^+ faces, whereas for the N170, EPN, and LPC[2], we expected a difference between CS^+ and CS faces but were unsure about the direction due to mixed findings in the literature.

In the **test session** after overnight consolidation, we investigated the dynamics of the extinction of the associated effects. Because of a more intense processing, behavioral effects for CS^+ faces should be reversed compared to learning: Gender decisions on CS^+ faces would be faster and more accurate than on CS_{neu} faces, with happy associations ($CS^+{}_{pos}$) boosting task performance more than angry associations ($CS^+{}_{neg}$), based on unpublished pilot data. We included a likability rating at the end of the test session and predicted that emotion-based associations should manifest in the ratings with $CS^+{}_{pos}$ > CS_{neu} > $CS^+{}_{neg}$ (similar to Suess et al., 2014). According to Hammerschmidt, Kagan, et al. (2018), effects of associated emotional relevance should not become evident in pupil size but in the same ERPs that would be modulated during learning, even if we assumed to observe partly extinction over the course of the session. We had no specific ERP-related hypotheses regarding differences between the gender-matching and mismatching conditions.

2.2 Method and Materials

The study was pre-registered prior to data collection (https://osf.io/b3fh2).

Stimuli

We selected sixteen frontal portrait photographs of faces with a neutral expression from the Göttingen Faces Database (Kulke et al., 2017). The faces were presented with their natural color on a light grey background, and edited and combined with a transparency mask covering the hairline, ears, and neck. They had a visual angle of about 3.16 × 5.14 degrees and a 200 × 300 pixels resolution. Images were controlled for luminance (M_{HSV} = 0.47, *SD* = 0.01, $\chi^2(225) = 240$, *p* = .235; Dal Ben, 2019). Vocal stimuli were taken from the Montreal Affective Voices database (Belin et al., 2008). Based on findings by Lausen & Schacht (2018), we chose twelve sounds with the highest recognition of emotion (angry, happy, neutral) and gender (female, male). The duration of the selected sounds ranged from 511 ms to 1831 ms. Their maximum intensity was digitally set equal (Praat; Boersma & Weenink, 2018) and resulted in a mean peak sound level of *M* = 47 dB (*SD* = 1.8 dB) at the participants' head position. The 'beep' tone for no-go trials was a sine wave sound of 630 Hz, 300 ms, with a 30 ms amplitude ramp in the beginning. We presented 12 unique face-voice pairs whereby one face stimulus was contingently paired with one voice stimulus for

[2] For the expected LPC effects, ongoing face processing and the overlapping sound onset might be hard to disentangle and thus require additional caution

doi: 10.53846/goediss-9841

each participant. Half of the face-voice pairs were gender congruent (i.e., face and voice female or face and voice male), and the other half were gender in-congruent pairs (e.g., female face and male voice). We measured an asynchrony between face offset and voice onset, ranging between -9 and +9 ms and occurring independently from certain stimuli or conditions.

Randomization

We counterbalanced the sound stimuli (based on how well emotion and gender were recognized in Lausen & Schacht, 2018) between congruence conditions across participants. The allocation of the face stimuli to the voices was fully randomized: We drew 12 out of 16 face stimuli (assuring six male and six female faces) and combined them with the 12 voices. The four remaining faces were used as new, i.e., not associated, faces only for the likeability rating. Stimuli were presented in 50 blocks, with each block containing a random sequence of the 12 stimuli (learning session: face-sound pairs; test session: faces only). For the memory checks after the learning and the test session, the order of the 12 faces and the emotion category labels' positions were also random.

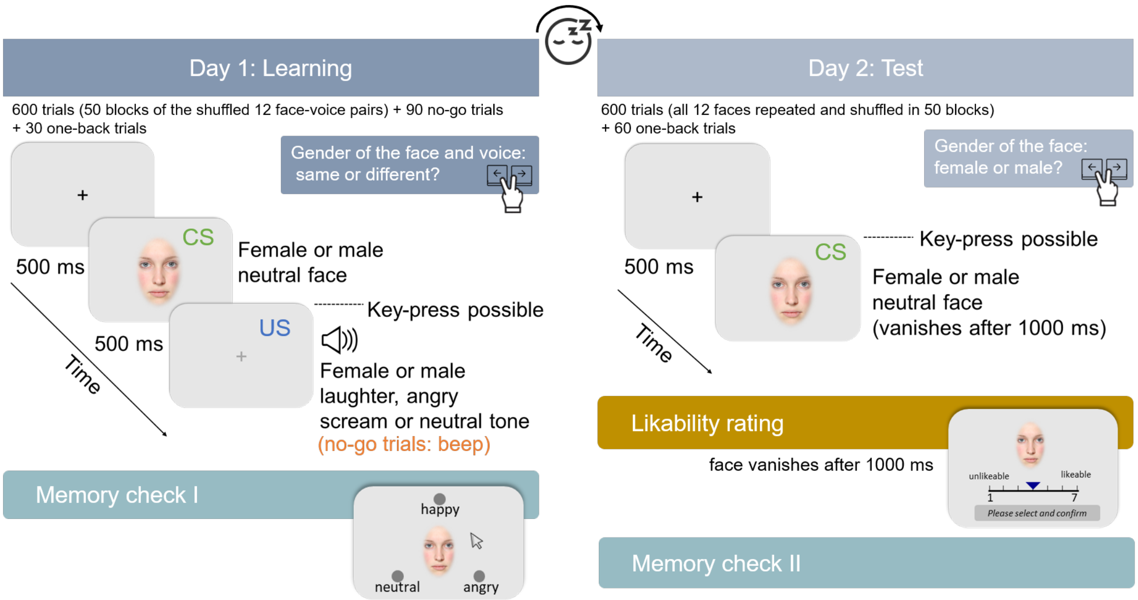
Procedure

Before the experimental sessions, participants filled out an online questionnaire of the German version of the Social Interaction Anxiety Scale (SIAS, Stangier et al., 1999). The sessions in the laboratory took place on two subsequent days, each lasting approximately two hours. At the beginning of each session, participants gave written consent to participate voluntarily in the study. At the beginning and the end of each session, we assessed their current mood with the German version of the positive and negative affect schedule (PANAS, Breyer & Bluemke, 2016, see Table A1). Additionally, we assessed socio-demographic data and handedness (day one) and quality of sleep with a modified version of the Pittsburgh Sleep Quality Index (PSQI, Buysse et al., 1989) (day two). Participants were seated in front of a computer screen in a dimly lit, electrically shielded, and sound-attenuated room at a distance of approx. 78 cm between eyes and presented face stimuli. They positioned their chin in a height-adjustable chin rest to avoid head movements. Two speakers were located on the left and right and at the monitor's height. For the presentation of the experiment, we used Python (2.7), including the modules PsychoPy (Peirce, 2009), PyGame (Shinners, 2011), and PyGaze (Dalmaijer et al., 2013) amongst standard modules. At the beginning of both sessions, we presented detailed instructions about the task on screen, and participants completed six practice trials (incl. feedback). After ensuring that the task was understood and cal-

doi: 10.53846/goediss-9841

ibrating the eye-tracker, the main experimental task began. For a visualization of the procedure, see Figure 2.1.

Figure 2.1
Procedure of the learning and test session.



Learning session. The task was to indicate as fast and accurately as possible via key press whether the gender of the face and voice match. At the center of the screen, first, a fixation cross and then the neutral face stimulus was presented, each for 500 ms. The vocal stimulus (negative, neutral, or happy affect burst) set in with face offset. Participants could respond as soon as the voice set in. The next trial started after a response was given and a variable inter-trial interval (M = 1800 ms; SD = 200 ms) automatically. At random positions, we included filler trials (90 no-go trials, and 30 one-back tasks) to motivate participants to stay focused. The no-go condition, in which a beep sound followed the face and in which participants shall not press any key, was implemented to ensure that attention was paid to the auditory stimulus. Hence it was not sufficient to learn/know the assignment of a face to a response key (which was always consistent within each participant). Every participant completed $50 \times 12 = 600$ trials (+ filler trials). After every 120 trials, there was a short break to rest. *Memory check.* Following the learning session, we assessed whether participants were able to allocate the faces and the emotional categories of the voices, although they were not instructed at any time to memorize the faces or face-voice pairs. We presented each face individually, and participants had to click on one of three labeled buttons (happy, angry, neutral) around the face to indicate the emotional category of the associated voice.

doi: 10.53846/goediss-9841

Test session. The following day, participants performed a gender decision task on the previously associated faces but without any voices or sounds present. Faces were displayed for 1000 ms, and the response was indicated via key press. Again, participants completed 50 blocks of the shuffled 12 faces and 60 additional one-back tasks.

Likability rating and second memory check. After completing the main part of the test session, all previously associated faces, intermixed with four new faces, were presented once each, and participants had to judge the faces concerning their likability on a 7-point Likert scale (1 = “unlikeable”, 7 = “likable”). Subsequently, the same explicit memory check of the first session took place. At the very end, participants were debriefed about the study’s aims and could clarify open questions with the experimenter.

Sample size and power analysis. The study had a 2 (gender congruence: match/mismatch) × 3 (emotion: angry/happy/neutral) within-subject design. As there is no standardized way to do a power analysis and sample size estimation for linear mixed models, we based the power estimation on a within-factors repeated-measures ANOVA (G*Power 3.1.9.2, Erdfelder et al., 1996), assuming a correlation among the repeated factors of .50. However, it served only as a rough estimate because it did not exactly reflect the planned and performed analyses. We stopped data collection after 41 participants completed both sessions successfully and refrained from collecting more due to a very uneven data loss for both sessions and measures. Especially for the learning session, we had more ERP trials with artifacts than we initially anticipated. To make the data from the learning and test session comparable, we also excluded the data for the other measures and sessions to ensure the same group of 32 participants in the learning and testing session.

Participants

Our final sample consisted of 32 participants (22 female, 10 male, 0 diverse; Age: 19-34, Mean = 23.5). All participants were right-handed (according to Oldfield, 1971), fluent in German, and did not self-report any (neuro-)psychiatric disorders. Participants with sight correction of more than plus/minus one diopter or any self-reported hearing difficulties were excluded. Participants were recruited through advertisements on campus and social network groups in Göttingen; hence the sample consisted mainly of students (29 out of 32). Participation was reimbursed by a fixed amount of money or course credit. The study was conducted following the Declaration of Helsinki (WMA, 1964), and all participants signed informed consent prior to both experiment sessions. The

doi: 10.53846/goediss-9841

mean SIAS score of our sample was 23.19 (range: 7 to 44) out of max. 80. Seven [3] participants showed elevated scores (>30, Stangier et al., 1999).

EEG recording and pre-processing

We recorded EEG from 64 (+6 external) electrodes during the learning and test session. Participants wore an electrode cap (Easy-Cap, BioSemi, Amsterdam, Netherlands) according to the extended 10-20 system (Pivik et al., 1993). External electrodes were positioned at the left and right mastoids, at the outer canthi of and below both eyes to record electro-oculograms. Common Mode Sense (CMS; active) and Driven Right Leg (DRL; passive) are special "ground" electrodes serving as an online reference during recording (see www.biosemi.com/faq/cms&drl.htm). Continuous EEG was recorded with a sampling rate of 512 Hz and a bandwidth of 102.4 Hz. Offline, the raw data was pre-processed in MATLAB (2018) with EEGLAB (v2019.0, Delorme & Makeig, 2004). We shifted all event markers by a constant of 24.3 ms to account for the monitor's systematic delay in stimulus appearance. Data was re-referenced to average (whole head) reference excluding external electrodes and filtered with a 0.01 Hz high-pass filter. The plugin "CleanLine" (v1.04, Mullen, 2012) was used to remove 50 Hz line noise. Data was epoched from -500 ms to 2000 ms and corrected to a 200 ms pre-stimulus baseline. We performed Independent Component Analysis (ICA) on a 1 Hz high-pass filtered copy of the dataset and subsequently transferred the resulting ICA weights to the original 0.01 Hz filtered dataset. ICA components were used to detect eye and muscle-related activity in the data. Data was corrected by removing components with a high probability of being labeled as such (muscle >80%, eye-related >90% or channel noise >90%) using "IClabel" (v1.2.4, Pion-Tonachini et al., 2017). Consequently, channels were interpolated if classified as bad. We trimmed epochs to -200 ms to 1000 ms and performed trial-wise rejections: amplitudes exceeding -100/100 μV (Learning: avg. 7.7%; Test: 7.4%) during face presentation, steep amplitude changes (5000μV within epoch; Learning: avg. 3.9%; Test: 2.8%) or improbable activation (>5 deviation of mean distribution for every time point; Learning: avg. 11.9%; Test: 12.8%) were excluded. Overall, there was a mean rejection rate of 17.3% (range: 9.7-36.8%) of trials for the learning session and 17.2% (range: 6.5-36.2%) for the test session due to these artifacts.[4] As eye-blinks were corrected with ICA, we extracted blink information from the pupil data to reject trials in which participants blinked during face presentation. We defined the time windows and regions of interest (ROIs) electrodes for the ERP components of interest based on previous research with similar stimuli and settings (face-locked ERPs: Hammerschmidt et al., 2017; voice-locked

[3] IDs: 6, 7, 21, 28, 23, 36, 40

[4] referring to the reported 32 data sets.

doi: 10.53846/goediss-9841

ERPs: Paquette et al., 2020; Pell et al., 2015) as follows: For the visual (face-locked) components: a) P1: mean and peak amplitudes, 80-120 ms; occipital electrode cluster: O1, O2, and Oz; b) N170: mean and peak amplitudes, 130 and 200 ms; occipito-temporal electrode cluster: P10, P9, PO8, and PO7 [5]; c) EPN: mean amplitudes, 250 - 300 ms; occipito-temporal cluster: O1, O2, P9, P10, PO7, and PO8; d) LPC: mean amplitudes, 400 and 600 ms; occipito-parietal electrode cluster: Pz, POz, PO3, and PO4.In addition to the pre-registered face-locked components, we analyzed voice-locked [6] ERPs with the following ROIs taken from Paquette et al. (2020) as they also used stimuli of the Montreal Affective Voices database (MAV, Belin et al., 2008): N1-P2 complex with N1 (90 - 145 ms) and P2 (165 - 300 ms), both with the identical fronto-central electrode cluster: F3, F1, Fz, F2, F4, FC1, FC3, FC2, FC4, C3, C1, Cz, C2, C4, CP1, CP3, CPz, CP2, and CP4.

Pupil recording and pre-processing

Pupil size was recorded binocularly in pixels with a sampling rate of 500 Hz using the Eye-Link 1000 desktop mount eye tracker (SR Research, Mississauga, ON, Canada). Before the start of the experiment, a nine-point eye-tracking calibration- and validation procedure was performed. We based artifact detection in pupil samples on the guidelines proposed by Kret & Sjak-Shie (2018). The pupil time series was time-locked to the onset of the face stimulus, and all artifact detection was performed sequentially. Samples were classified as blinks or invalid when both eyes were lost. We marked invalid samples and specific trial windows to flag trials in which the participant missed the stimulus (-onset), e.g., during baseline, face presentation, or later. Median absolute deviation (MAD) speed was estimated, and samples with a higher speed than 16 times MAD were flagged as artifacts. A smoothed trend line was calculated, and clusters of samples with a strong deviation from the trend line were flagged (in four iterations). Isolated samples within longer periods of missing data (separated clusters) were dismissed. Samples at the border of a gap were trimmed (50 samples pre- and post-gap, i.e., “extended blinks”) and interpolated but only for gaps that were maximally 125 samples of missing data. The eye with fewer invalid trials was selected, and a baseline adjustment was performed (samples were subtracted by the mean of the baseline period from -200 ms to face onset). The time course of a trial was segmented into 60 bins, and outlier samples within a bin were flagged (>3 standard deviations from the bin mean). A trial was rejected if > 75% of its samples or pre-specified time windows of the trial were invalid. A smoothing 4 Hz filter was applied to the data before averaging by participant and condition.

[5] PO9 and PO10 were replaced by PO8 and P07 as they were not part of the used EEG recording system.

[6] Our setting allowed measuring the exact voice onset with an audio-photo-diode only without the use of the speakers and not during the actual experiment. For the target event to epoch the data, we used the offset of the face stimulus. Due to a slight jitter, the timing of the auditory ERPs is less precise compared to the visual ERPs.

doi: 10.53846/goediss-9841

Statistical analysis

Tables with statistical models (incl. estimates, confidence intervals, stability measures, and likelihood ratio tests) are in the Appendix A. We used linear mixed models to analyze RTs and neurophysiological data, aggregated by participant and condition, using the function “lmer” of the package “lme4” (Bates et al., 2015). All statistical analysis was conducted in R (v 4.0, R Core Team, 2020). For parameter estimation, we chose the maximum likelihood (ML) method. The model predictors are the emotion of the associated sounds (“neutral”, “angry”, “happy”), gender-congruence of the face-voice pairs (“match”, “mismatch”), and a random intercept (participant) to consider the dependency in the data due to the repeated-measures design and variability between participants. We compared models (full model including the interaction, additive models, and leaving out each predictor) with goodness of fit tests, known as likelihood ratio (LR) chi-square difference tests, to identify which predictors add significantly to explaining variance in the data (Snijders & Bosker, 2012). For likelihood ratio tests (LRT), we used the “mixed” function of the package “afex” (Singmann et al., 2020). Regression coefficients (β), Standard Errors (*SE*), 95 % Confidence intervals (*CI*) and stability of the coefficients are reported. To obtain *CIs*, we used a parametric bootstrap (N = 1000). We estimated the stability of model coefficients by fitting the same model on subsets of the data (dropping one random effect at a time). Residuals of the models were inspected visually, and potential collinearity among predictors was determined with Variance Inflation Factors (VIF), which will be reported if model assumptions appeared violated. Reference levels in all models were “match” for congruence and “neutral” for emotion, respectively. The following linear mixed models are sum-contrast-coded, reflecting main effects rather than marginal effects. Here, the intercept corresponds to the (unweighted) grand mean, and lower-level effects are estimated at the level of the grand mean. This coding implies that the reference factor level (“match” or “neutral”, respectively) receives a value of -1 on all contrast variables, whereas all other factor levels are mapped onto exactly one contrast variable with a value of 1. This implies that for every factor with k levels, k-1 parameters are estimated and cannot be directly mapped to the factor levels. Post-hoc Šidák adjusted tests were used to test the difference between levels of factors with more than two levels, using “emmeans” (Lenth, 2020).

For the ERPs (P1, N170, EPN, and LPC) and pupil size, in addition to overall effects, we planned to investigate the dynamics of the acquisition of the association during the learning and a possible extinction during the test session (only for ERPs). To reduce the number of statistical tests and differently from what we preregistered, we only analyzed the dynamics if an overall effect of a predictor was present. To do so, we used an exploratory data-driven approach and fitted GAMLSS-models (generalized additive model of location scale and shape, Rigby & Stasinopou-

doi: 10.53846/goediss-9841

los, 2005) with a Gaussian outcome distribution, including the variables emotion, congruence, and experiment block (i.e., repetition of the face-voice pairs but averaged across conditions) as well as a random intercept for participant ID. Block was included as a non-linear, P-spline (Eilers & Marx, 2010) smoothing function, of which its degree of smoothness (lambda) was estimated automatically via a local Maximum Likelihood method. We used a backward step-wise method to detect the model with the lowest generalized Akaike criterion (GAIC), starting from a full interaction model with the three-way interaction of the fixed effects Emotion $\times$ Congruence $\times$ pb(Block) $+$ random(participant). We reported the predictors of the final model but visualized the fitted values of the full model, allowing for a three-way interaction. Note that we used this approach to explore the data. Due to general issues with step-wise model selection methods, we refrain from reporting any p-values or making statements about the data-generating process.

Exclusion criteria and re-coding. We detected unexpected systematic errors in some of the participant's behavioral data. To preserve as much power as possible, we decided to apply additional criteria for exclusion or re-coding in the following cases: a) Same errors occurring consistently (i.e., at least in two-thirds of trials) for a specific stimulus pair (face + voice): As for the learning session it was not distinguishable whether the participants had difficulties identifying the gender of the face or the gender of the voice, we re-coded only cases in which both, in the learning and the test session, the same faces were constantly misclassified. This happened in four cases: VP 35 (l10_neu_m.png), VP 27 (l5_neu_f.png), VP 7 (l5_neu_f.png) and VP 36 (l2_neu_f.png). In these cases, previously wrong answers were recoded to correct, and the congruence of the stimulus pair was changed (if originally mismatch to match and vice versa). b) All trials with (systematic) errors which occurred only in the learning session but not in the test session were excluded and not recoded. c) One participant (VP 25), initially mixed up the key assignment and answered all trials incorrectly. After re-instructing the participant, all subsequent trials were answered correctly. We did not recode the answers of this participant but excluded the trials before the re-instruction. Unlike preregistered and due to the unexpected systematic error patterns, we only report descriptive statistics of the accuracy data in this study.

Outlier removal and model robustness. Despite the EEG and eye-tracking clean-up, pronounced variability and outlier observations were still present in some components and measures. Instead of excluding a participant if they lost more than 25% of their trials as preregistered, we set the lower limit of 30 valid trials per condition. To enhance model stability and robustness, we excluded observations consistently for all pupil size and ERP models in both sessions for which

doi: 10.53846/goediss-9841

Table 2.1
Means and standard deviations of the main ERP, Pupil and behavioral measures for the learning session

	match						mismatch					
	neutral		happy		angry		neutral		happy		angry	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
P1	3.95	3.16	3.58	3.03	3.50	3.07	3.87	3.05	3.85	2.97	3.87	3.25
P1.peak	6.30	3.59	5.90	3.29	5.76	3.32	6.08	3.34	6.01	3.27	6.17	3.66
N170	-5.80	3.44	-5.52	3.23	-5.37	3.33	-5.51	3.29	-5.26	3.35	-5.44	3.02
N170.peak	-10.27	4.77	-10.07	4.55	-10.02	4.65	-10.04	4.61	-9.77	4.60	-10.00	4.44
EPN	0.27	3.75	0.23	3.95	0.38	3.40	0.84	3.78	0.61	3.62	0.55	3.61
LPC	3.09	2.35	3.23	2.20	3.14	2.46	2.99	2.30	2.95	2.20	2.78	2.37
N1(auditory)	-0.83	0.77	-0.69	0.81	-0.59	0.79	-0.72	0.79	-0.69	0.72	-0.72	0.77
P2(auditory)	1.12	1.35	1.38	1.29	1.29	1.26	0.79	1.16	0.88	1.19	0.97	1.28
Pupil.early	-20.20	40.13	-18.26	40.30	-21.34	48.05	-20.22	38.52	-25.69	40.72	-28.59	45.55
Pupil.late	83.83	75.15	91.02	80.75	81.21	80.91	100.38	88.49	90.33	86.34	79.17	84.82
RT	691.94	165.66	734.95	165.24	713.70	162.32	785.11	164.11	833.31	192.28	791.97	173.74
Accuracy	94.34	11.78	96.41	3.93	93.62	10.76	93.06	13.60	96.06	8.86	88.66	17.48

Notes: Response times are in ms, ERP amplitudes in μV, and Pupil size in pixels.

their cook's distance was larger than 0.5 in any model. Due to these exclusion methods, the total sample size for all ERP measures resulted in 32 participants.

2.3 Results

Implicitness of the association learning

Participants differed in their ability to recall the emotion category of the voices when faces were presented separately from the voices. Some participants performed significantly above chance, i.e., $\geq$ 8 out of 12 faces correct, corresponding to a 2-sided Exact Binomial Test with $p \leq .05$. After the learning session, 5 out of 32 participants and after the test session, 3 out of 32 met this criterion. However, only one participant performed above chance across both session checks and would therefore not be considered an implicit learner.

Learning session

Table 2.1 contains all means and standard deviations of the visual and auditory ERPs, pupil size, and behavioral measures.

Face-locked ERPs. P1: P1 mean amplitudes were not significantly modulated by emotion ($\chi^2(2) = 3.91$, $p = .141$). There was a trend towards an effect of gender-congruence ($\chi^2(1) = 3.41$, $p = .065$) with larger mean amplitudes ($d_{\text{mismatch-match}} = 0.19\ \mu V$) in gender-mismatching than matching trials. The interaction between emotion and congruence was not significant ($\chi^2(2)$

doi: 10.53846/goediss-9841

= 3.6, p = .165). There were no significant effects on P1 peak amplitudes, neither for emotion ($\chi^2(2)$ = 2.86, p = .239), congruence ($\chi^2(1)$ = 0.68, p = .409), nor their interaction ($\chi^2(2)$ = 4.05, p = .132).

N170: There was a trend for N170 mean amplitudes to be modulated by emotion ($\chi^2(2)$ = 5.75, p = .057), with happy (-5.39 μV) and angry (-5.40 μV) descriptively being less negative compared to neutral-associated faces (-5.66 μV) when averaged across gender-congruence conditions. However, post-hoc contrasts between emotion levels were not significant (all *ps* >= .05). There was no significant modulation by congruence ($\chi^2(1)$ = 2.47, p = .116) and no significant interaction between congruence and emotion ($\chi^2(2)$ = 2.6, p = .273). N170 peak amplitudes were not significantly modulated by emotion ($\chi^2(2)$ = 2.24, p = .327), nor congruence ($\chi^2(1)$ = 2.03, p = .155) nor their interaction ($\chi^2(2)$ = 0.82, p = .664).

EPN: The mean amplitudes of the EPN component were not significantly modulated by emotion ($\chi^2(2)$ = 0.73, p = .693) but by congruence ($\chi^2(1)$ = 7.49, p = .006) with less negative amplitudes for gender-mismatching (0.67 μV; β_{mismatch} = 0.19, *SE* = 0.07, t = 2.72) compared to matching trials (0.29 μV). No interaction between congruence and emotion was present ($\chi^2(2)$ = 1.47, p = .480). **EPN during learning:** To investigate the dynamics of the EPN component during learning we fitted a GAMLSS model. The model with the lowest GAIC included the main effects for emotion, congruence, and block, a one-way interaction between emotion and block, and the random factor subject ID. A visualization of the fitted values of the full model (allowing for all interactions) is shown in Figure 2.2D. A non-linear function of the amplitudes for the learning session is observable.

LPC: LPC mean amplitudes were not significantly modulated by emotion ($\chi^2(2)$ = 0.84, p = .658), but there was a main effect of congruence ($\chi^2(1)$ = 4.21, p = .040). Gender-matching trials (3.15 μV) had more positive amplitudes compared to mismatching trials (2.91 μV; β_{mismatch} = -0.12, *SE* = 0.06, t = -2.03). The interaction between congruence and emotion ($\chi^2(2)$ = 0.84, p = .657) was not significant. **LPC during learning:** The model for the LPC over time with the lowest GAIC was a model including congruence and block and the random factor subject ID. Neither emotion nor any interactions were included in the final model. A visualization of the fitted values of the full model (allowing for all interactions) can be found in Figure 2.3D.

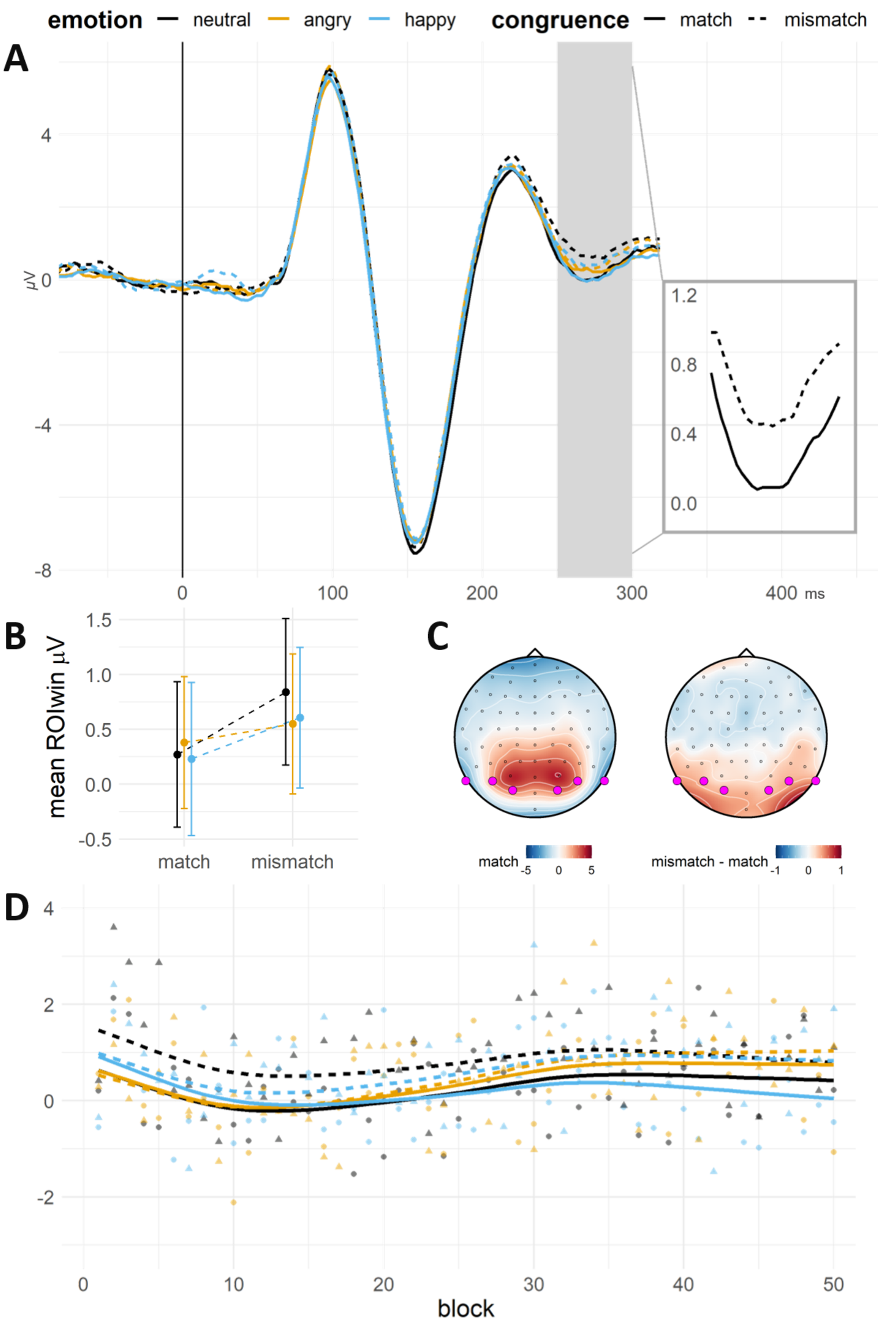
Voice-locked ERPs. **N1:** For the N1 component, neither emotion ($\chi^2(2)$ = 3.35, p = .187) nor congruence ($\chi^2(1)$ = 0.01, p = .918) nor their interaction ($\chi^2(2)$ = 2.98, p = .226) were significant.

P2: The P2 component was modulated both by emotion ($\chi^2(2)$ = 6.3, p = .043) and congru-

doi: 10.53846/goediss-9841

Figure 2.2

Face-locked EPN in the learning session.

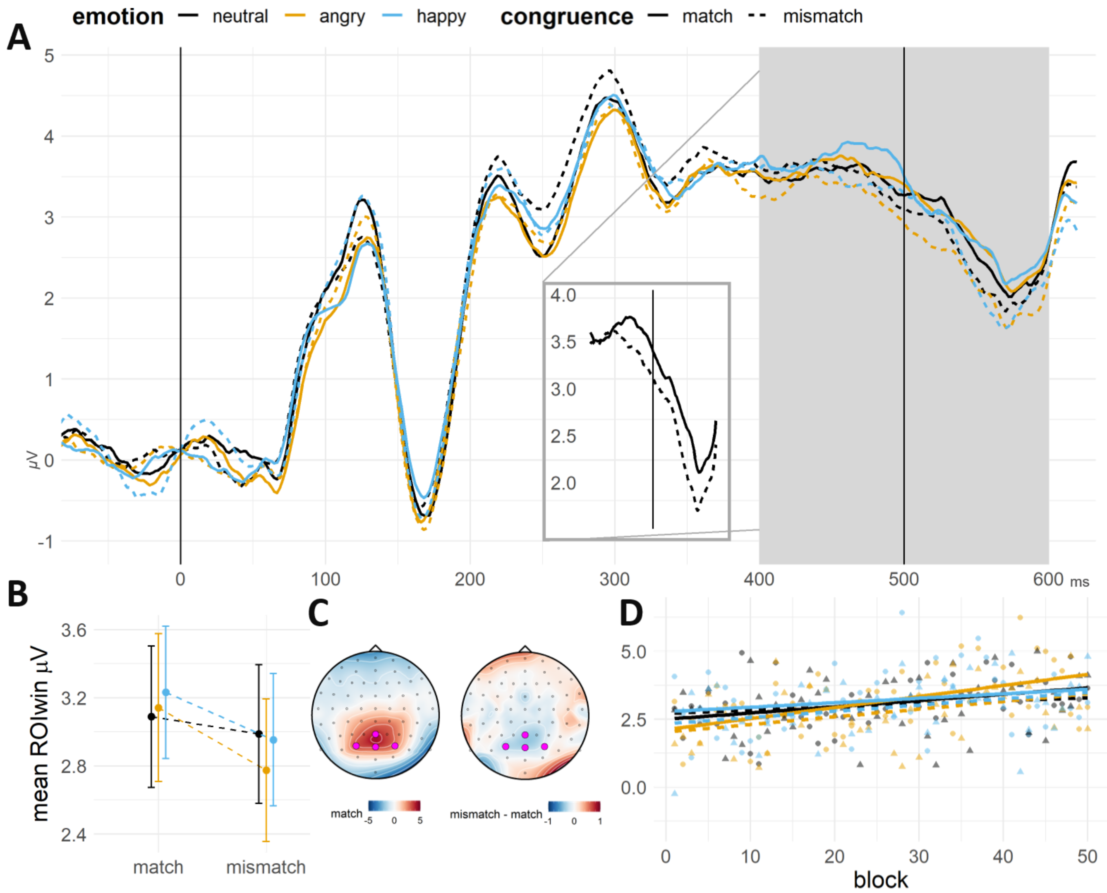


Notes: **A** Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender-congruence, averaged over all emotion conditions. **B** Grand averaged ERP amplitudes of the ROI, contrasted for all conditions. Errorbars indicate +/- 1 *SE* of the mean. **C** Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink. **D** Mean EPN amplitudes over the course of the learning session. Dots represent the grand averages per block and condition. The curves represent the fitted values of the GAMLSS model.

doi: 10.53846/goediss-9841

Figure 2.3

Face-locked LPC in the learning session.



Notes: **A** Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender-congruence, averaged over all emotion conditions. **B** Grand averaged ERP amplitudes of the ROI, contrasted for all conditions. Errorbars indicate +/- 1 *SE* of the mean. **C** Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink.**D** Mean LPC amplitudes over the course of the learning session. Dots represent the grand averages per block and condition. The curves represent the fitted values of the GAMLSS model.

doi: 10.53846/goediss-9841

ence ($\chi^2(1) = 29.51, p <.001$). However, there was no interaction between congruence and emotion ($\chi^2(2)$ = 1.44, p = .486). Gender-mismatching trials (0.88 μV; β_{mismatch} = -0.19, *SE* = 0.03, t = -5.6) had a smaller P2 amplitude compared to matching trials (1.26 μV). Emotional voice stimuli (angry: 1.13 μV; β_{angry} = 0.06, *SE* = 0.05, t = 1.22; happy: 1.13 μV; β_{happy} = 0.06, *SE* = 0.05, t = 1.27) elicited more positive P2 amplitudes compared to the neutral voice stimuli (0.95 μV). When adjusting for multiple comparisons, pairwise comparisons failed significance. Descriptively, the largest difference was between neutral and happy ($d_{\text{hap-neu}}$ = 0.18, p = .090) bursts followed by neutral and angry bursts ($d_{\text{ang-neu}}$ = 0.18, p = .097). Both, happy and angry bursts showed a similar pattern ($d_{\text{hap-ang}}$ = 0.00, p = 1.000). **P2 during learning:** The model for the P2 over time with the lowest GAIC included emotion, congruence, block, and the random factor subject ID. No interactions were in the final model. For the full model (allowing for all interactions), see Figure 2.4.

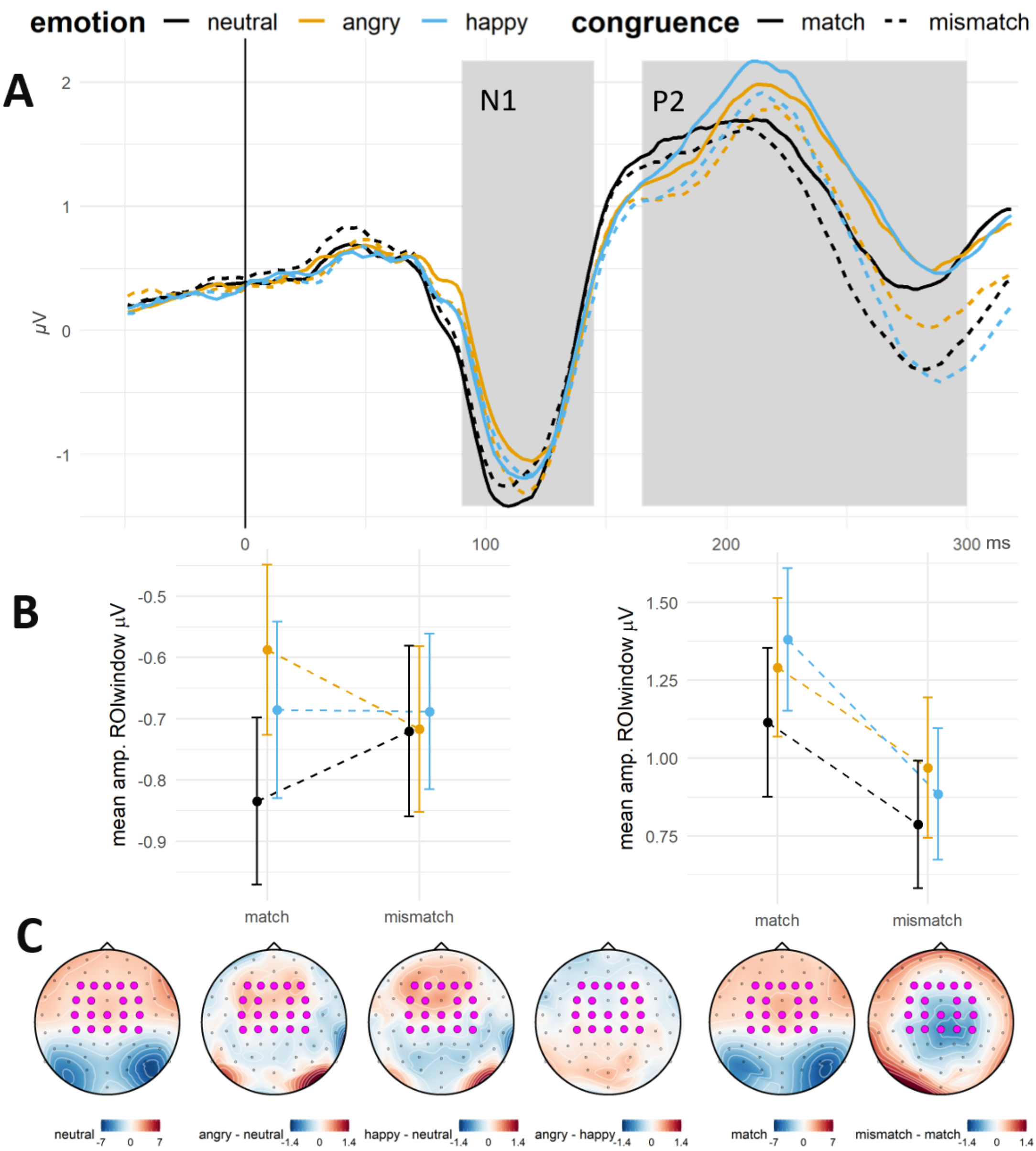
Pupil size modulations. For the early time window (0 - 1000 ms after face onset), which mainly reflects a modulation of the pupil constriction, there was no significant modulation of the pupil size by emotion ($\chi^2(2)$ = 3.69, p = .158) but there was a main effect of congruence ($\chi^2(1)$ = 5.7, p = .017), with a stronger constriction for mismatching stimuli ($d_{\text{mismatch-match}}$ = -4.90, p = .019). No interaction between emotion and congruence was present ($\chi^2(2)$ = 2.86, p = .240). In a later time window (1000 - 2000 ms after face onset), the pupil size was significantly modulated by emotion ($\chi^2(2)$ = 8.07, p = .018). Pairwise comparisons revealed significant differences in pupil size between the angry and neutral condition ($d_{\text{ang-neu}}$ = -11.92, p = .031), but no significant differences between happy and neutral trials ($d_{\text{hap-neu}}$ = -1.43, p = .986) and happy and angry trials ($d_{\text{hap-ang}}$ = 10.49, p = .070). There was no main effect of congruence in the later time window ($\chi^2(1)$ = 1.55, p = .214). Although there was only a trend for interaction between emotion and congruence ($\chi^2(2)$ = 5.16, p = .076), the interpretation of the main effect shall be taken with caution. **Pupil size during learning:** Models with the lowest GAIC for both time windows of pupil size (constriction and dilation) over experiment blocks included the main effect of emotion, congruence, and block and the random factor subject ID. No interactions were in the final model.

Behavioral measures. **Accuracy:** Prior to any re-coding or rejection, the overall accuracy in the gender matching task was 94% (N = 32), with descriptively higher accuracy for match- (95%) than for mismatch trials (93%). The accuracies for each emotion category were 96% for happy, 91% for angry and 94% for neutral trials. None of the subjects fell into the exclusion criteria (>25% incorrect trials). After re-coding systematic error patterns (see sec. Exclusion criteria and re-coding), the overall accuracy was 94%. Again, match trials had a higher accuracy (95%) compared

doi: 10.53846/goediss-9841

Figure 2.4

Voice-locked N1 and P2 in the learning session.

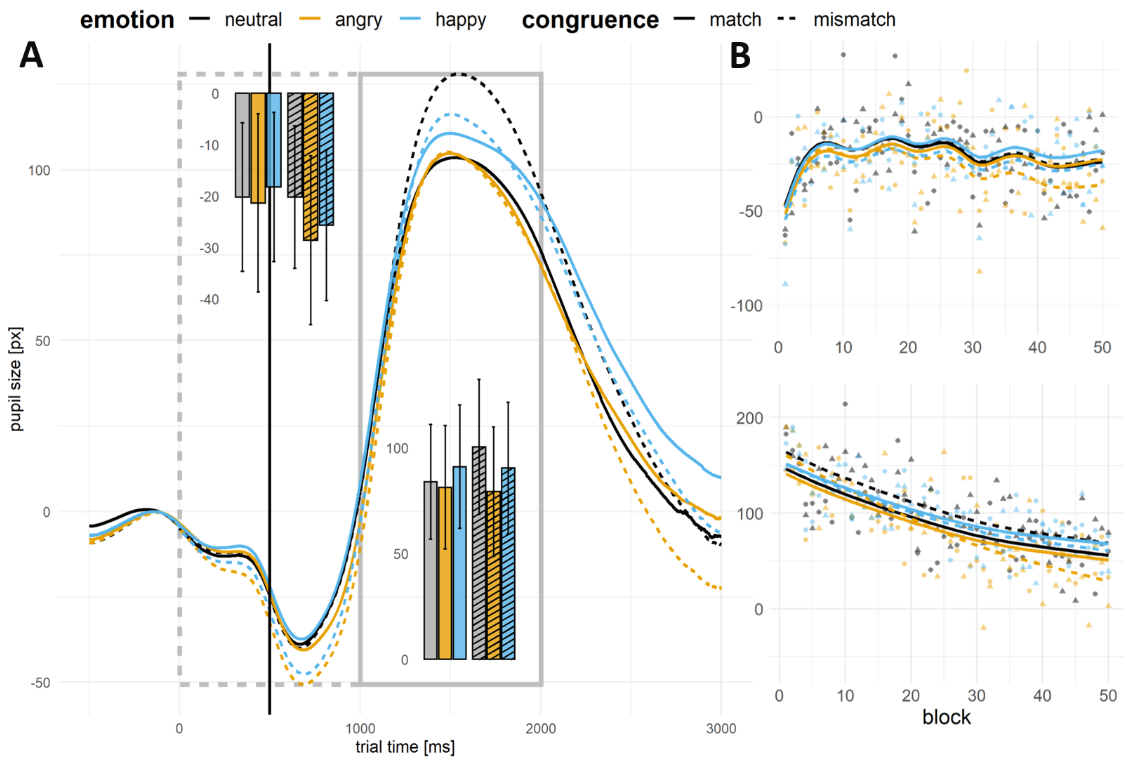


Notes: ROI channels are highlighted in pink and identical for the N1 and P2. **A** Grand average ERP of the averaged ROI channels. The highlighted areas display the ROI time windows of the N1 (left) and P2 (right). **B** Grand-averages of the N1 ROI (left panel) and P2 ROI (right panel), contrasted for all conditions. Errorbars indicate +/- 1 *SE* of the mean. **C** Topographies of the ERP distribution of the P2, depicting the main effects of congruence and emotion: Neutral bursts (collapsed across gender-congruence levels), pairwise differences of the emotion levels, and gender congruent bursts (collapsed across emotion levels) compared to gender-incgonruent bursts. ROI channels are highlighted in pink.

doi: 10.53846/goediss-9841

Figure 2.5

Pupil size results of the learning session.



Notes: **A** Grand average pupil size time series. Bar plots refer to the pupil size in each condition for the respective marked time windows. **B** Pupil size time series across the experiment for the early time window (0 - 1000 ms; upper panel) and the later time window (1000 - 2000 ms; lower panel).

doi: 10.53846/goediss-9841

Table 2.2
Means and standard deviations of the main ERP, Pupil and behavioral measures for the test session

	match						mismatch					
	neutral		happy		angry		neutral		happy		angry	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
P1	5.05	3.54	5.09	3.29	4.85	3.37	5.30	3.29	5.35	3.30	5.17	3.50
P1.peak	7.26	3.91	7.46	3.53	7.18	3.56	7.64	3.51	7.73	3.58	7.43	3.82
N170	-5.86	3.95	-5.82	3.98	-5.73	3.76	-5.47	3.80	-5.69	3.96	-5.85	3.85
N170.peak	-10.65	5.32	-10.56	5.33	-10.47	5.09	-10.31	5.10	-10.49	5.24	-10.59	5.22
EPN	-0.17	3.92	-0.26	4.27	-0.11	4.16	0.33	4.21	0.13	4.31	-0.03	3.96
LPC	4.09	2.95	4.00	2.64	3.76	2.81	3.88	2.53	3.90	2.71	4.00	2.88
Pupil.early	5.04	58.53	3.94	52.73	5.82	54.67	6.94	54.33	4.04	56.83	3.28	57.30
Pupil.late	32.52	73.50	35.18	65.22	37.63	68.88	37.25	67.76	36.26	74.92	29.48	63.64
RT	596.42	56.33	593.67	58.20	598.61	56.85	603.77	62.27	598.65	56.80	601.33	59.00
Accuracy	98.41	1.34	97.80	3.94	96.71	7.71	96.29	9.70	98.32	1.85	96.85	7.78

Notes: Response times are in ms, ERP amplitudes in μV, and Pupil size in pixels.

to mismatch trials (93%), and concerning emotion categories, happy trials (97%) had a higher accuracy compared to neutral (94%) and angry (92%) trials.

Response Time (RT): RT data was analyzed only for correctly answered trials. First, we trimmed RTs using a maximum cut-off of 5000 ms. Then, we applied a skewness-adjusted boxplot method to exclude extreme values separately for every subject, using the function “adjbox” of the package “robustbase” (Maechler et al., 2021; based on: Hubert & Vandervieren, 2008). The overall (non-aggregated) mean RT for the gender-matching task of the learning session was 758 ms (SD = 369 ms). We based the RT-model estimation on aggregated data, taking the mean for each condition (emotion, congruence) and subject.[7] Results showed a modulation by both, congruence ($\chi^2(1) = 71.84$, p <.001) and emotion ($\chi^2(2) = 15.49$, p <.001) but there was no interaction ($\chi^2(2) = 0.81$, p = .667). Matching trials (714 ms) were answered faster compared to mismatching trials (803 ms; $\beta_{\text{mismatch}} = 44.97$, SE = 4.8, t = 9.37). Neutral trials were answered fastest (739 ms), followed by angry (753 ms; $\beta_{\text{angry}} = -5.66$, SE = 6.79, t = -0.83) and happy trials (784 ms; $\beta_{\text{happy}} = 25.63$, SE = 6.79, t = 3.78). Pairwise comparisons revealed that estimated RT differed significantly between neutral (fastest) and happy (slowest) trials ($d_{\text{hap-neu}}$ = 46 ms, p <.001) and between happy and angry trials ($d_{\text{hap-ang}}$ = 31 ms, p = .025).

Test session

Table 2.2 contains all means and standard deviations of the visual ERPs, pupil size, and behavioral measures.

[7]Taking the median instead of the mean did change parameters slightly but not the direction or significance of the effects.

doi: 10.53846/goediss-9841

Face-locked ERPs. **P1:** P1 mean amplitudes were not significantly modulated by emotion ($\chi^2(2)$ = 2.73, p = .255), but by congruence ($\chi^2(1)$ = 6.3, p = .012). No interaction was found ($\chi^2(2)$ = 0.08, p = .963). Similar to the learning session, gender-mismatching trials (5.27 μV; β_{mismatch} = 0.14, SE = 0.06, t = 2.49) had a descriptively higher P1 amplitude than matching trials (5.00 μV), although the sound stimuli were not presented anymore. A similar pattern was found for P1 peak amplitudes, which were not modulated by emotion ($\chi^2(2)$ = 3.52, p = .172), but by congruence ($\chi^2(1)$ = 5.52, p = .019). Again, no interaction between emotion and congruence was found ($\chi^2(2)$ = 0.19, p = .910). **P1 during extinction:** Similar to the learning phase, we were interested in the temporal dynamics during extinction. We used the same model structure as we did for the learning data. The data-driven selected model for the P1 mean amplitudes over time with the lowest GAIC was the full model, including all main and interaction effects. Panel D of Figure 2.6 shows the fitted values. All model results can be found in Appendix A.

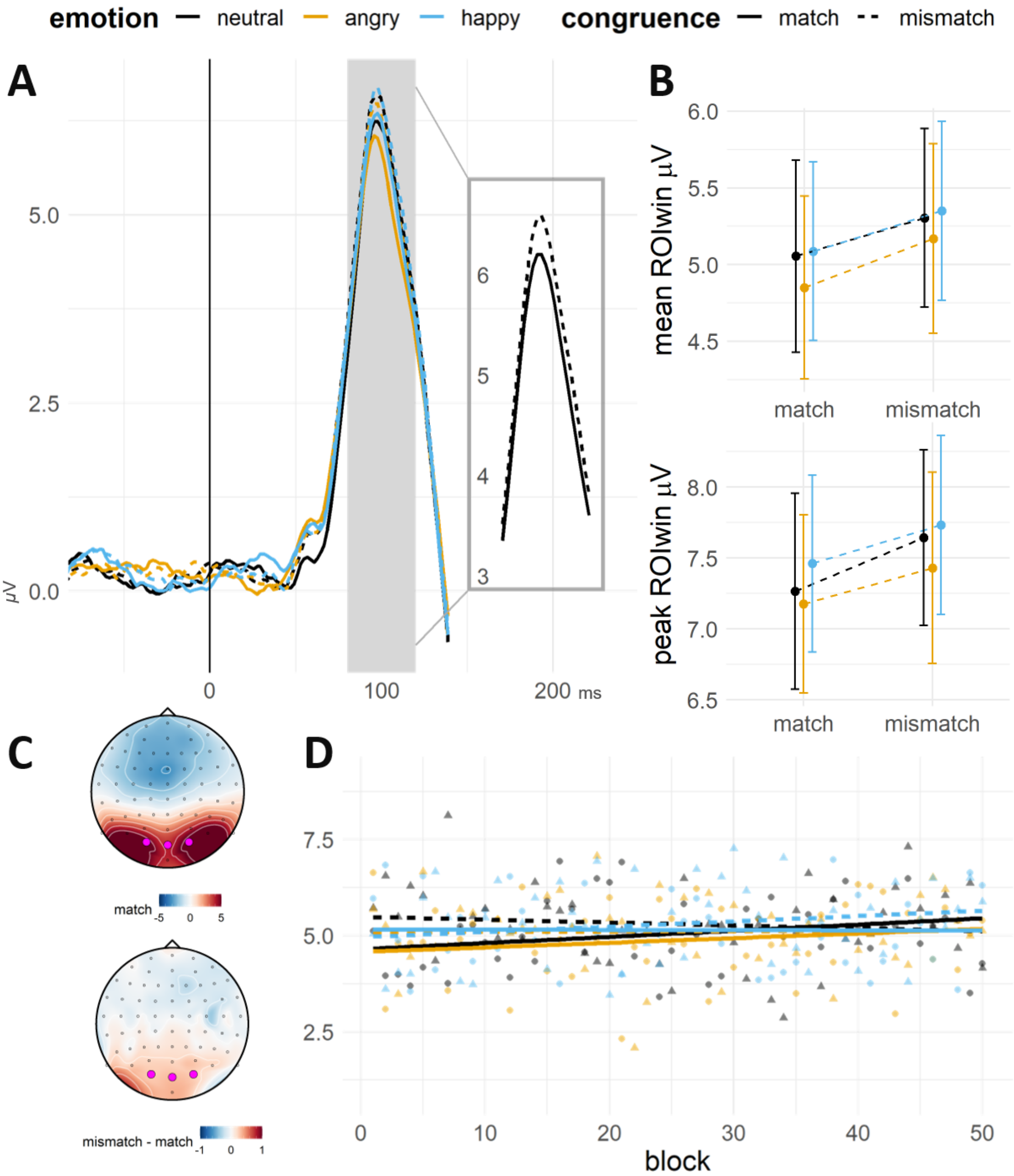
N170: N170 mean amplitudes were not significantly modulated by emotion ($\chi^2(2)$ = 1.7, p = .428) or congruence ($\chi^2(1)$ = 2.66, p = .103), but there was a significant interaction between emotion and congruence ($\chi^2(2)$ = 6.16, p = .046), with the difference between previously matching and mismatching trials was larger for neutrally associated faces (0.38 μV, t = 2.68, p = .008) compared to faces previously associated with affective sounds (β_{happy} = 0.13 μV, t = 0.92, p = .361; β_{angry} = -0.12 μV, t = -0.81, p = .421). For neutrally associated faces in previously mismatching trials (-5.47 μV), the N170 showed a less negative deflection compared to matching trials (-5.86 μV). In contrast to N170 mean amplitudes, peak amplitudes were neither significantly modulated by emotion ($\chi^2(2)$ = 0.18, p = .914), nor congruence ($\chi^2(1)$ = 0.98, p = .323) nor their interaction ($\chi^2(2)$ = 3.25, p = .197). **N170 during extinction:** For the N170 over time of the test session, a model with emotion, congruence, block, and all two-way and three-way interactions had the lowest GAIC.

EPN: For EPN mean amplitudes, no significant modulation by emotion was present ($\chi^2(2)$ = 1.59, p = .451). There was a main effect for congruence ($\chi^2(1)$ = 8.91, p = .003), with smaller amplitudes for faces associated with gender-mismatching compared to matching sounds, similar to the learning phase. The interaction between congruence and emotion ($\chi^2(2)$ = 2.8, p = .247) was not significant. **EPN during extinction:** Looking at the EPN during extinction over time, the data-driven selected model included the main effects of emotion, congruence, and block, as well as the interaction between emotion and congruence. No interaction with the factor block was in the final model.

LPC: LPC mean amplitudes were neither significantly modulated by emotion ($\chi^2(2)$ = 0.8, p = .669), nor by congruence ($\chi^2(1)$ = 0.04, p = .843), nor their interaction ($\chi^2(2)$ = 3.33, p = .189).

doi: 10.53846/goediss-9841

Figure 2.6
Face-locked P1 in the test session.

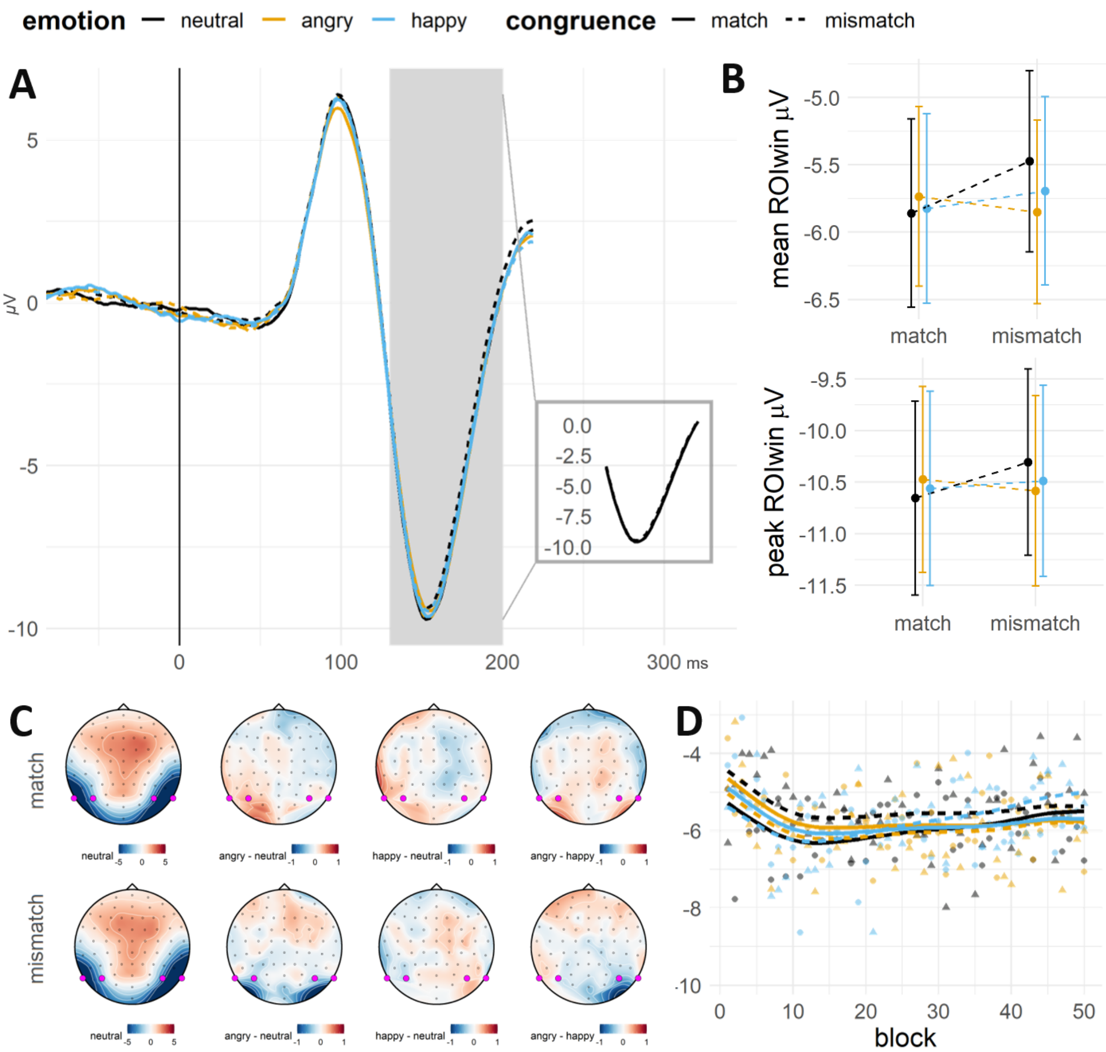


Notes: **A** Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender-congruence, averaged over all emotion conditions. **B** Grand-averages of the ROI mean amplitudes (upper panel) and peak amplitudes (lower panel), contrasted for all conditions. Errorbars indicate +/- 1 *SE* of the mean. **C** Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink. **D** Mean P1 amplitudes over the course of the test session. Dots represent the grand averages per block and condition. The lines represent the fitted values of the GAMLSS model.

doi: 10.53846/goediss-9841

Figure 2.7

Face-locked N170 in the test session.

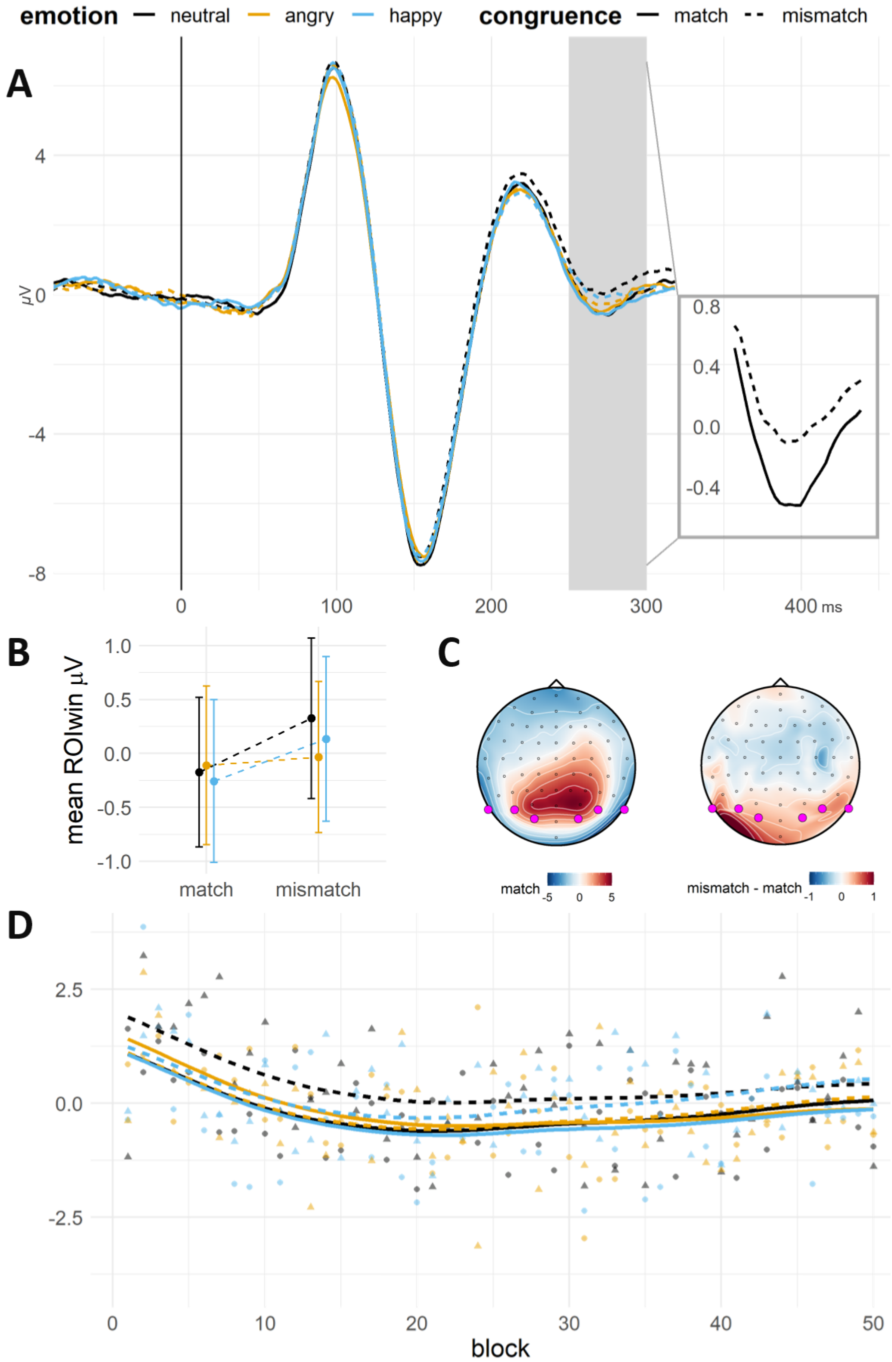


Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender-congruence, averaged over all emotion conditions. **B** Grand-averages of the ROI mean amplitudes (upper panel) and peak amplitudes (lower panel), contrasted for all conditions. Errorbars indicate +/- 1 *SE* of the mean. **C** Topographies of the ERP distribution for faces associated with neutral compared to emotional bursts, separately for gender-congruent and -incongruent faces. ROI channels are highlighted in pink. **D** Mean N170 amplitudes over the course of the test session. Dots represent the grand averages per block and condition. The lines represent the fitted values of the GAMLSS model.

doi: 10.53846/goediss-9841

Figure 2.8

Face-locked EPN in the test session.



Notes: **A** Grand average ERP of the averaged ROI channels. The highlighted area displays the ROI time window. The zoomed-in window shows the main difference of gender-congruence, averaged over all emotion conditions. **B** Grand-averages of the ROI amplitudes, contrasted for all conditions. Errorbars indicate +/- 1 *SE* of the mean. **C** Topographies of the ERP distribution for gender-congruent faces and the difference between gender-incongruent and congruent faces. ROI channels are highlighted in pink. **D** Mean EPN amplitudes over the course of the test session. Dots represent the grand averages per block and condition. The lines represent the fitted values of the GAMLSS model.

doi: 10.53846/goediss-9841

Pupil size modulations. For the pupil early time window (0 - 1000 ms) there were no significant effects of emotion ($\chi^2(2) = 1.05$, $p = .591$) nor congruence ($\chi^2(1) = 0.01$, $p = .911$) nor their interaction ($\chi^2(2) = 1.23$, $p = .540$). Similarly, pupil size was not modulated in a later time window (1000 - 2000 ms) by emotion ($\chi^2(2) = 0.37$, $p = .832$) or congruence ($\chi^2(1) = 0.07$, $p = .791$) nor their interaction ($\chi^2(2) = 3.37$, $p = .185$).

Behavioral measures. **Accuracy:** Prior to any re-coding or rejection, the overall accuracy for the gender decision task was 97% (N = 32), thereby higher than for the learning task. Accuracy was similar between match (98%) and mismatch trials (97%). The accuracies for each emotion category were 98% for happy, 97% for angry, and 97% for neutral trials. None of the subjects fell into the exclusion criteria (> 25% incorrect trials). After the re-coding of certain systematic error patterns, the overall accuracy was 97% for both faces, independent of whether they were presented in match (97%) or mismatch trials (97%), happy (98%), angry (96%) or neutral (98%) trials during learning.

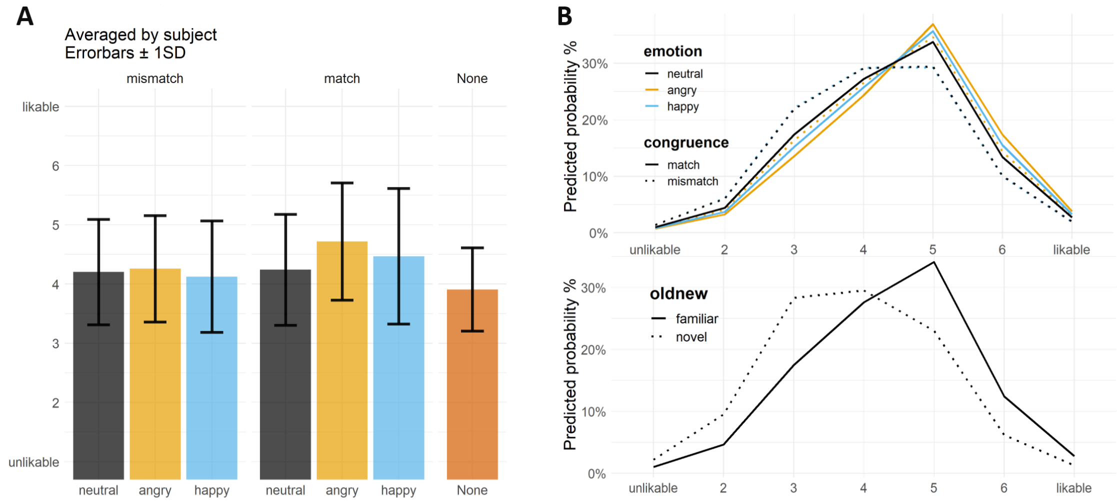
Response Time: The overall mean (non-aggregated) RT for the test session was 598 ms (*SD* = 119 ms). The RT-model estimation on aggregated data was conducted analogously to the learning session. RTs showed a modulation only by congruence ($\chi^2(1) = 5.06$, $p = .024$) but not by emotion ($\chi^2(2) = 2.7$, $p = .259$), nor by the interaction between both factors ($\chi^2(2) = 0.73$, $p = .694$). Gender decisions to faces previously presented with gender matching bursts (596 ms) were answered faster than those presented with mismatching sounds (601 ms; $\beta_{\text{mismatch}} = 2.51$, $SE = 1.12$, $t = 2.23$).

Likeability rating: We ran two cumulative linked mixed models to account for the ordinal scale of the likability ratings; one model only included the associated faces, comparing gender-congruence and emotion levels, and a second for the previously associated compared to novel faces. In both models, random intercepts for participant ID and face stimulus were included. First, when comparing the full with reduced models via likelihood ratio tests, a model including only congruence was significant ($\chi^2(1) = 4.371$, $p = .037$). However, when allowing for the congruence $\times$ emotion interaction, there were no statistically significant differences in likability ratings (all *CIs* of the *OR* included 1). Second, there was a significant difference between associated and novel faces ($\chi^2(1) = 140.32$, $p = <.001$), with novel being rated as less likeable (*OR* = 0.45 [0.31;0.65]). The predicted probabilities of both models for the likability ratings are shown in Figure 2.9. Both models' odds ratios and 95% CI can be found in Table A22 of Appendix A.

doi: 10.53846/goediss-9841

Figure 2.9

Likability rating of associated and novel faces.



Notes: **A** Barplots represent the likability ratings per condition, averaged within and across subjects. **B** Fitted values as predicted probabilities of the ordinal models. The upper panel shows the model including emotion and congruence. Please note that within the gender-mismatching condition, the dotted line for happy (blue) is mostly overlapping with the neutral (black) and hence difficult to see. The lower panel shows the collapsed familiar faces versus the novel faces.

2.4 Discussion

Prior work has documented that faces can gain additional relevance when associated with affective context information. Not only highly aversive stimuli such as loud noise bursts or electric shocks modulated face processing (for a review, see: Miskovic & Keil, 2012), but also verbal descriptions of behavior (Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017; Luo et al., 2016; Schindler, Bruchmann, Krasowski, et al., 2021; Suess et al., 2014; Xu et al., 2016), social and affective signals (Aguado et al., 2012; Bruchmann et al., 2021; M. J. Wieser, Gerdes, et al., 2014), and abstract forms of context like monetary reward and loss (Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017). In the present ERP and pupillometry study, we applied an emotion-implicit cross-modal association paradigm, with separate learning and test sessions, to investigate whether and how robustly neutral faces might acquire additional relevance, even if the emotional quality of the US, i.e., the affect burst, is not task-relevant.

Although we observed differences in the neural and behavioral responses towards emotional compared to neutral face-voice pairs, faces do not seem to have acquired additional relevance by their associations with emotional voices. During learning, we found emotion effects ranging from auditory processing (auditory P2) to pupil changes and behavioral responses, presumably triggered by the affect bursts. Hence, emotional sounds were automatically processed and elicited

doi: 10.53846/goediss-9841

typical responses at the neuro-physiological level. However, these emotion-based effects were not transferred to the conditioned faces (CS^{+}_{pos} and CS^{+}_{neg}), and emotion did not modulate face-locked ERPs as expected. Remarkably, faces acquired a different quality according to the task-relevant congruent or incongruent gender information, which was present in the behavior, auditory ERPs, and mid- and long-latency visual ERPs during learning. Most astonishingly, congruence effects were also present in the behavior, early and mid-latency visual ERP components during the delayed test session, indicating a substantial role of task-relevant stimulus features for successful associative learning. Thus, the absence of the associated relevance of the emotional voices cannot be attributed to a general disregard of the (emotional) voice stimuli since gender congruence was transferred from learning to testing. These findings suggest that sufficient associations of emotional valence to faces require explicated attention, as previously indicated by Hammerschmidt, Kagan, et al. (2018). In the following, we will discuss the emotion-related effects occurring during the learning phase and the conditioned effects visible both in the learning and test session. We will use the term “face-voice pairs” instead of CS^{+} and CS^{-} faces for pupil size, auditory ERPs, and RTs of the learning session in which it is impossible to disentangle immediate processing (e.g., reacting towards the affective voices) and effects that might have occurred because the faces gained predictive value.

Emotion and congruence effects of face-voice pairs

As hypothesized, during acquisition, gender-matching decisions were slower in emotional compared to neutral face-voice pairs. This finding corroborates studies demonstrating that task-irrelevant emotional stimuli are more difficult to disengage with and withdraw attentional resources from the actual task (Carretié, 2014; Dresler et al., 2008; Hur, Iordan, Dolcos, et al., 2016; Kotz & Paulmann, 2007; Schimmack, 2005; Vuilleumier, 2005). In addition, the conflicting emotional information of neutral faces and positive/negative affect bursts might have counteracted integrated person percepts (Föcker et al., 2011; Gelder & Vroomen, 2000). Presumably, this type of conflicting information is more detrimental for naturally co-occurring stimuli like faces and voices than for abstract stimuli. Independent of emotion, responses for mismatching compared to matching face-voice pairs were slower, replicating previous findings (Huestegge et al., 2019; Latinus et al., 2010). Gender-mismatching[8] information might be more difficult to integrate because of the long-term and repeated strengthening of the associations of gender-congruent faces and voices. To investigate at which point the emotion and congruence differences were reflected at the neural

[8] We are only referring to trials in which the participant’s response was correct; hence, subjective gender-mismatching corresponded to objective gender-mismatching according to our stimulus database.

doi: 10.53846/goediss-9841

level and additionally to the preregistered visual ERPs, we explored auditory ERPs (N1 and P2), time-locked to the voice-onset. We did not find a modulation of the N1 by emotion or congruence. There was an enhancement of the auditory P2 component for happy and angry compared to neutral face-voice pairs, similar to T. Liu, Pinheiro, Deng, et al. (2012). This difference might reflect the prediction effect of the emotion-congruence, i.e., pairs of a neutral face and a neutral voice, as shown in several studies: Typically, auditory suppression effects have been found when a preceding visual stimulus predicted the occurrence of a sound (Vroomen & Stekelenburg, 2010). These suppression effects were also reported in the context of dynamic faces and spoken utterances (Ho et al., 2015; Wassenhove et al., 2005) or emotionally congruent vocalizations (Jessen & Kotz, 2013; Kokinous et al., 2014). We also found a gender-congruence effect on the auditory P2 component with larger amplitudes for gender-matching compared to mismatching voices. These increased amplitudes might reflect "attention allocation costs", as described by the predictive coding framework (Feldman & Friston, 2010). With the specific task demands of *gender-congruence* decisions, feature-based attention towards *gender-congruent* voices might have led to prioritized processing over the course of learning, also reflected later in shorter RTs. The dissociation between emotional congruence and gender congruence at the level of the auditory P2 amplitudes provides novel evidence for its sensitivity to attention and predictive processing. This is in line with studies demonstrating a sensitivity of the P2 component to various types of incongruence (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010) and to interactive processes of prediction and attention (Schröger et al., 2015).

In our study, we expected that face-voice pairs containing affective compared to neutral bursts would elicit higher arousal and hence, larger pupil dilation, and that this response would be shifted towards the predicting face stimulus. In the time window of 0-1000 ms after face onset, overall, gender-mismatching trials elicited a stronger pupil constriction compared to matching trials, being most pronounced for face-voice pairs containing affective bursts. However, in the time window from 1000 to 2000 ms after face-onset and hence, covering the dilation of the pupil as well as the presentation of the voices, pupil size was impacted by emotion. Trials with neutral bursts elicited an overall larger dilation compared to affective bursts, which was most pronounced in gender-mismatching trials. This contradicts previous findings (e.g., Burley et al., 2017; Finke et al., 2021; Hammerschmidt, Kagan, et al., 2018; Schindler et al., 2022), with motivationally relevant and affective stimuli eliciting larger pupil dilation. However, pupil size also enlarges with increases in cognitive load (Oliva & Anikin, 2018; for a review on pupil size correlations see Zekveld et al., 2018). While our data indicates that both the predictability of emotional and gender congruence were interacting, it remains unclear which functional interaction between emotion and congruence

doi: 10.53846/goediss-9841

took place. Remarkably, there was no direct correspondence between pupil size and RTs during learning. To better understand these results, we looked at the habituation effects of the pupil during the learning phase across conditions. We found typical decreasing pupil responsiveness (especially reduced dilation magnitudes) from the beginning of the experiment towards the end. However, no systematic pattern was observable regarding emotion (e.g., no differences in the speed of habituation between emotion conditions).

Associated effects

Neither RTs, pupil size measures nor auditory ERPs can be treated independently from learning. Our conditioning paradigm allowed us to distinctively map the face-locked ERP modulations to different learning processes according to our experimental conditions both for the learning and the test session. Our main hypothesis was that positive and negative CS^+ faces would be processed differently than CS_{neu}. However, virtually none of the pre-registered ERP components indicated a modulation by emotion. Instead, several ERP components, pupil size, and RTs were modulated by (the previously conditioned) gender congruence of the face-voice pairs.

Learning session. Already in the learning session, we found a difference between the gender-matching and mismatching conditions on the EPN, with an enhanced negative amplitude for matching trials. The EPN is usually associated with the attentive encoding of emotional or motivational relevance (Schupp et al., 2006). Evidence from associative learning studies is mixed, with some reports of enhanced ERP negativities for CS^+ compared to CS^- (Bruchmann et al., 2021; Sá et al., 2018; Schellhaas et al., 2020) and others demonstrating differences between EPN effects of acquired and inherent relevance (Aguado et al., 2012; Hammerschmidt et al., 2017). Taking the functional link of the EPN to other attention-related ERP components (Schupp et al., 2007; Schupp et al., 2006) into account, it remains open whether the gender congruence manipulation in our study induced a kind of emotional relevance or whether the EPN effects reflect rather general (emotion-independent) attention processes.

Test session. In the test session, in line with our predictions, pupil size did not show modulations by emotion or congruence. Physiological measures have been shown to extinguish fast even when differentiation in neural measures is still persistent (e.g., Hammerschmidt, Kagan, et al., 2018; Pastor et al., 2015). Similar to the learning session but differently from what we expected, virtually all ERP modulations were related to gender congruence. These findings indicate a generalization of some processing differences to a different task. The most unexpected finding was the modulation at early processing stages, reflected in enhanced P1 amplitudes for the faces of the

doi: 10.53846/goediss-9841

previously gender-incongruent condition. The P1 has mostly been associated with the processing of lower-level stimulus properties, as well as with the rapid detection of conditioned motivated or emotional salience (e.g., Aguado et al., 2012; Hammerschmidt et al., 2017; Müller-Bardorff et al., 2018; but cf. Bruchmann et al., 2020). In general, early visual processing is influenced by task demands and perceptual load (Handy et al., 2001; Pratt et al., 2011). However, the CS-US pairing for each participant was fully randomized. Hence, even though some voices or faces might have been more difficult to extract gender information from, physical stimulus characteristics cannot explain this effect. Since participants could not anticipate whether a congruent or incongruent face would be presented due to the randomized order of presentation, P1 differences are necessarily related to the processing of the stimuli. Unlike other fear-conditioning studies with faces, which reported an enhancement of the CS^+, Sperl et al. (2021) found an increased response for CS^- faces in occipito-temporal channels at the typical P1 time window. The authors suggested that smaller P1 amplitudes for CS^+ might be caused by prolonged attention during learning and hence a smaller prediction error for CS^+. This illustrates an interesting dissociation between early (P1) and the subsequent processing stages. Y. Liu et al. (2011) analyzed P1 amplitudes as a function of stimulus repetition and found larger P1 amplitudes for CS^+ at the beginning with a switch towards the end of the experiment in favor of CS^- stimuli. In our study, gender-incongruent face-voice pairs seemed to be processed less elaborately, indicated by a larger prediction error with increased amplitudes of the P1 and decreased mid-latency and late ERPs, as well as slower RTs during learning. This prediction error apparently even remained present during extinction. The only preregistered visual ERP component that was modulated by an interaction of congruence and emotion was the *N170* component during the test session, which was characterized by a smaller mean (but not peak) amplitude for the conditioned incongruent CS_{neu} faces compared to all others. However, we expected that, if at all sensitive to our experimental condition, it would show an enhancement for matching and emotionally, especially negatively conditioned faces, as found in other studies (e.g., Bruchmann et al., 2021). We cannot exclude that (neutral) faces with neutral gender-mismatching voices elicited a stronger interference for configural processing of the faces, again implying functional dissociation of the N170 to the other components, although this assumption would need to be tested by future studies. Noticeably, the interaction effect was only significant for the N170 mean but not peak amplitudes. As the preregistered N170 component was measured at partly overlapping electrodes as the EPN component, the mean amplitude effect during the N170 time window might already represent a mixture of configural face processing and relevance encoding (for a discussion about distinct N170 and EPN emotion effects, see Rellecke et al., 2012b). However, for the *EPN*, the interaction between conditioned emotion and congruence failed significance, whereas the

doi: 10.53846/goediss-9841

difference between conditioned gender-matching and mismatching conditions became more pronounced. The latter finding is, albeit unexpected, interesting in several aspects: First, the pattern of the EPN differences between all conditions was very similar across both sessions and seemed to be robust enough over time and across different task demands. Second, the scalp topography of the congruence effect showed high conformity to typical emotion-based effects known from studies with inherent emotional salience like emotional expressions of faces, words, or complex scenes (e.g., Bayer & Schacht, 2014). The EPN modulation by conditioned gender congruence is corroborated by indicators of overt behavior, i.e., faster gender decisions during the test session. Eventually, gender-matching faces acquired positive valence because they were easier to process during learning. In comparison, gender-mismatching faces did not seem to acquire strong negative valence but were probably more treated as an artificial by-product of the task. Alternatively, gender-incongruent voices might have introduced uncertainty about the gender of the face due to a reflexive integration of the face and the voice towards a whole-person concept, which was not overcome by the repetitions of the learning session. Even when the voice was no longer present in the test session, this learned uncertainty made it more difficult to decide on the gender of the face. Finally, the LPC component was neither modulated by emotion nor by congruence in the test session, probably due to task settings (e.g., Hammerschmidt et al., 2017; Rossi et al., 2017)

Neural dynamics of learning and extinction

We investigated the dynamics of learning and extinction of the emotion- and congruence-based cross-modal associations. Y. Liu et al. (2011) described the learning-related temporal dynamics of the P1 component as non-linear, consisting of three phases: an initial decrease, a subsequent increase, and habituation. Similarly, we allowed for potential non-linear relationships between repetitions and ERP amplitudes as a function of learning. We averaged stimuli per block and condition to avoid arbitrary bin sizes, despite a lower signal-to-noise ratio. However, since our model does not capture rapid changes, the overall patterns of effects are similar up to a certain minimum number of bins. In Appendix A, we provide visualizations of the ERP amplitudes over the course of the session with different bin sizes. Sperl et al. (2021), who used electric shocks, reported very fast differentiation between CS^+ and CS^- stimuli within the first five repetitions in P1 and N170 time windows. We aimed to observe gradual associative strengthening between the CS^+ and US using medium-intense US. In our study, only the EPN appeared to change non-linearly across learning. Analogously, during extinction, only mean N170 and EPN amplitudes appeared to follow a non-linear function, with an initial decrease phase and then slowly returning to baseline. The exploratory results are data-driven and should be understood as an impetus for further research

doi: 10.53846/goediss-9841

and replication with an independent sample. For example, in our data amplitudes of the auditory P2 differentiated between both gender congruence and emotion levels quite consistently over the entire learning session. Similarly, LPC amplitudes during learning were best explained by a model with the factors congruence and block, like in the averages across the whole session. In contrast, for the EPN during learning and test and the P1 at test, models including the factor emotion could explain the overall data significantly better, even though averaged across the session, emotion was not significant. Unexpectedly, the reduced negative amplitude for the mismatching CS_{neu} in the N170 time window was apparent virtually throughout the test session. Moreover, when comparing ERP dynamics of the P1, N170, and EPN in the test session, we found more positive amplitudes for the mismatching CS_{neu} relative to the other conditions. This changed only towards the end of the session when it converged with the mismatching ${CS^+}_{pos}$. However, although these models show emotion-based effects in some of the ERP components, the effects of gender congruence were overall larger and more robust.

Emotional implicitness of the task

We expected emotional vocalizations to capture increased attention due to their social and biological relevance (Johnstone & Scherer, 2000) and to transfer relevance via associative learning to the faces. Looking at the acquisition period, when faces were conditioned with inherent emotional vocalizations, we found that the RTs were affected not only by the gender (in-)congruence but also by the emotion of the vocalization. Thus, emotional relevance indeed affected behavior as expected, but it did not transfer to learning to the same degree as the task-relevant gender congruence. Possible reasons will be discussed in the following:

Can faces be conditioned with auditory emotional expressions? Emotional expressions of the face and voice naturally show variations within people and among situations. One could argue that faces naturally show some “resistance” to be conditioned with information that is naturally very variable and not stable. However, Aguado et al. (2012) successfully associated faces with positive and negative emotional expressions of the identically portrayed individual (same modality). Cross-modal associations have also been demonstrated in fear-conditioning studies (Miskovic & Keil, 2012), with some of them using aversive screams (Bruchmann et al., 2021; Glenn et al., 2012; Schindler et al., 2022).

How much attention towards the US is needed to form stable associations? Crucially, in many fear-conditioning paradigms, the unconditioned stimuli were not task-relevant but consisted of highly aversive stimuli. If stimuli of lower intensity were used, most task instructions aimed

doi: 10.53846/goediss-9841

at making CS-US contingencies explicit. In contrast, we did not instruct participants about the CS-US contingency but implemented a task in which specific features, i.e., gender information, of the stimuli were relevant. In fact, the majority of participants were not aware of the emotional CS-US contingency, as indicated by our memory checks. Still, our task ensured that they attended to both the CS and US.

Attention and executive load can modulate conditioning effects also with stronger aversive stimuli like electric shocks (Hur, Iordan, Berenbaum, et al., 2016). Despite the high accuracy in the gender-matching task, one possible limitation of our study is that the gender-matching task might have prevented emotional conditioning, as indicated by the long-lasting and robust gender congruence effects on ERP measures. Therefore, it seems plausible that drawing attention to other (non-emotional) features of the same stimulus might have suppressed associative learning of emotion, although emotional vocalizations were processed differently from neutral sounds. Associative learning thus seems to depend on how much feature-based attention can be devoted to the US.

Future directions

We only tested the explicit CS-US contingency related to emotion. To rule out that the missing effect was caused by the missing CS-US contingency awareness, future studies should assess whether the gender congruence is explicitly retrievable. More generally, we used a paradigm in which emotion was neither task-relevant during learning nor test. It should therefore be investigated in a systematic cross-over setting what specific role feature-based attention during conditioning and retrieval plays. Investigating feature-based attention in the context of fear-conditioning and extinction might ultimately contribute to improving therapeutic interventions (e.g., exposure-therapy) in the context of clinical research.

Conclusion

We implemented an associative learning paradigm to investigate whether neutral faces automatically acquire emotional relevance when associated with cross-modal emotion from the voice, while emotion was not task-relevant. Emotion effects were limited to auditory ERPs, pupil size, and RTs in the acquisition period, possibly being promptly elicited by the emotional burst. In contrast, the task-relevant gender-congruence of the face-voice pairs impacted virtually all measures during acquisition. However, more strikingly, it modulated neural (P1, N170, and EPN) and behavioral responses towards previously conditioned faces during test on the following day and in a

doi: 10.53846/goediss-9841

different task. Our results indicate that associative learning of emotional relevance prerequisites global attention to the emotional stimulus and specific (feature-based) attention toward the emotional quality of the stimulus to be integrated.

Funding: This work was supported by Deutsche Forschungsgemeinschaft, Grant/Award Number: 254142454 / GRK 2070. **Acknowledgments**: We thank Jantje Becker for her help with recruitment and data collection. **Author contributions**: Annika Ziereis: Conceptualization, Data curation, Formal analysis, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing; Anne Schacht: Conceptualization, Writing - Review & Editing, Supervision.

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Chapter 3

Modelling response times in an associative learning task with delayed responses: A GAMLSS and drift diffusion approach

Abstract

From an evolutionary perspective, stimuli of higher biological relevance, such as social and emotional stimuli, should be preferentially processed and learned, and show slower extinction. Learning and extinction of faces (CS) conditioned with affective and neutral bursts (US) were investigated in two independent experiments ($N_{\text{Exp.1}}$ = 40, $N_{\text{Exp.2}}$ = 41) using emotion-implicit Pavlovian conditioning paradigms in which only the gender information of the faces and voices were task-relevant. We used a distributional (GAMLSS) approach to analyze response times (RTs) to capture the potential non-linearity of learning and extinction. In addition, we applied a mechanistic (drift-diffusion) approach to identify underlying processes accounting for differences between emotion conditions in RTs and choices beyond practice effects. In both experiments, learning was characterized by a general acceleration of responses, particularly at the beginning of the session, and by facilitation of gender-congruent compared to gender-incongruent face voice pairs. In contrast to our expectation, in both experiments, overall, participants responded fastest to neutral face-voice pairs, which was mainly reflected in the starting point parameter of the drift-diffusion model, indicating that the decision process started before the US. However, learning curves differed between emotion levels, with steeper learning curves for anger-related face-voice pairs. Responses in the test sessions were faster and showed fewer dynamics than learning sessions. Conditioned effects of emotion were only apparent in Exp.1. In contrast, conditioned gender-congruence effects were present throughout the test session in Exp.2. The applied methods revealed dynamics between valence conditions which would largely remain hidden when averaging RTs over the sessions.

doi: 10.53846/goediss-9841

3.1 Introduction

Early on, we learn about the contingencies of stimuli in our environment (e.g., the color of a cloud can be a good predictor of rain), about consequences of our own behavior (e.g., dropping a watermelon from a high building will probably make a big mess), about the meaning of words and the tastes of different types of food. Associative learning might also be one of the most important features of navigating the social environment. We learn about other people, their preferences and beliefs, which might change how we interact with and perceive them. To anticipate another person's behavior is beneficial in several ways, e.g., to react promptly to someone who is a physical or social threat and to respond differently to a known-to-be-friendly person.

The aim of this study was to investigate associative learning in the context of the perception and processing of faces associated with auditory laughter, screams, or neutral tone. Situational or learned contextual cues can alter the way we perceive and respond to faces (for reviews, see Miskovic & Keil, 2012; M. J. Wieser, Gerdes, et al., 2014). In an associative learning study of Aguado et al. (2012), non-expressive faces, which were repeatedly followed by an image of the same person's face with a different facial expression, elicited differential neural responses depending on whether the second facial expression was an emotional or non-emotional expression, suggesting that also emotional information can be associated to faces. As their study conditioned faces within the same (i.e., visual) modality, it had yet to be tested whether associative learning of emotion can occur similarly across modalities, and, in particular when emotion was not relevant for the current task goals during learning.

The underlying theoretical assumption of affective learning is that emotional stimuli are more salient (A. K. Anderson, 2005; Frischen et al., 2008; Hansen & Hansen, 1988) and increase the associative strength of the predicting stimulus and, in consequence, the learning rate, especially if the stimuli share (biological) relevance (Öhman, Erixon, et al., 1975; Öhman, Eriksson, et al., 1975). From this perspective, relevant stimuli should lead to faster learning and facilitated information uptake, possibly through attentional prioritization (e.g., Öhman & Mineka, 2001). We hypothesized that faces associated with affect bursts would gain additional relevance and thus become behaviorally more effective than faces conditioned with neutral bursts, although emotion was not task-relevant. With two different methods, we analyzed data from two independent experiments in which faces with neutral expressions (conditioned stimuli; CS) were cross-modally associated with affect bursts (unconditioned stimuli; US). We were particularly interested in two aspects: first, how emotional relevance influences responses to visually presented faces when the emotional context is not task-relevant, and second, how responses change as a function of associative learning and

doi: 10.53846/goediss-9841

extinction over time.

As it is difficult (if not impossible) to measure learning processes directly, their investigation requires the reduction to measurable behavior. Compared to explicit ratings (e.g., perceived differences before and after learning), the use of performance measures, such as response times (RTs) and (correctness of) choice reduces response biases by social desirability and the attempt to answer consistently (Nederhof, 1985; Nichols & Maner, 2008). Performance measures have been frequently used in psychological research, and the understanding of what affects RTs and their link to cognitive processes has constantly been growing, although it is far from being complete (for a review, see Schall, 2019). Studies on Pavlovian conditioning often have focused on behavior which is in part automatic, e.g., neural or physiological responses (for an overview, see Ojala & Bach, 2020). Whereas several studies have included RT measures as responses to inherently emotional stimuli in attentional tasks (e.g., Puls & Rothermund, 2017; Rooijen et al., 2017; Theodoridou et al., 2013) or in emotional priming studies (e.g., LeMoult et al., 2012; Purcell & Stewart, 2016; Sim et al., 2020; Voyer & Myles, 2017; Wagenbreth et al., 2016), in the context of associative learning of social and emotional CS stimuli, studies often reported response times only in addition to neural or physiological responses (e.g., Beckes et al., 2013; Björkstrand et al., 2022; Critchley et al., 2002; Dirikx et al., 2004; Pischek-Simpson et al., 2009; Watters et al., 2018). One reason might be that the relationship between task-relevant behavior and associative learning is not always clear. Different from studies on instrumental learning, in which the relation between acquired associations and behavioral responding is more direct, in Pavlovian conditioning, certain assumptions have to be made to relate intentional and task-dependent responses (e.g., a key press) to typical conditioned responses toward a stimulus. One of the assumptions in our study was that performance measures could indirectly capture the associative strength between stimuli, such that RTs would change as a function of learning and depending on the stimulus relevance. More precisely, if faces could gain additional relevance by being conditioned with affective bursts, we expected RT differences between the emotional and neutral face-voice pairs over the course of learning.

We were explicitly interested in the capacity of emotional US to be conditioned to faces when emotion was not task-relevant. To obtain overt and measurable responses, we implemented a gender-matching task of the face and the burst, ensuring that both stimuli were processed and attended to, but for which the emotional valence of the voices was irrelevant. We originally predicted faster responses in both the learning and test/extinction phase for faces associated with emotional compared to neutral voices. However, we did not specify predictions about gender-congruence effects as they were not the main focus of the study but included this factor in the model to control for potential differences in emotion effects between gender-matching and mismatching face-voice

doi: 10.53846/goediss-9841

pairs. Moreover, different from instrumental learning tasks, in our design, all information to make the correct choice was present in every trial. Thus, we expected that the accuracy of the gender-matching task and the gender-decision task would not change drastically with learning, but rather that learning was reflected in the speed of responding. Due to controversial findings in the literature about the attentional effects of conditioned stimuli and how they translate to response times (see, e.g., Ojala & Bach, 2020), we modeled learning over time to detect potential differences in learning curves between emotion categories which might not be detected when averaging over the whole experimental session. Different ways to analyze RT data have been discussed in the literature (e.g., Baayen & Milin, 2010; Green, 1971; Lo & Andrews, 2015; Marmolejo-Ramos, Barrera-Causil, et al., 2022; Ranger & Kuhn, 2013) to account for their typical properties: RTs can only be positive and are usually positively skewed (i.e., mode, mean and median differ), and participants additionally show inter-individual motor limits as well as intra-individual variation (e.g., due to differences in fatigue levels). Furthermore, not only the mean but also the spread of the response time distribution can be affected by task properties, e.g., by increases in task difficulty through less intense or ambiguous stimuli (e.g., Wagenmakers & Brown, 2007). To account for that, we followed two complementary approaches in the present study, of which the first aimed at monitoring changes over time for correct responses allowing for a higher temporal resolution, whereas the second targeted both accuracy and response times and interpretable parameters on a larger time scale.

First, we modeled the effects of emotion and gender congruence on RT changes over time to get better insight into the dynamics between emotion and congruence conditions. To do so, we fitted a generalized additive models of location scale and shape (GAMLSS, Rigby & Stasinopoulos, 2005), separately for the learning and test session of two independent experiments. GAMLSS models belong to the class of semi-parametric distributional models (see Kneib, 2013) and allow to model not only the mean but further parameters of the response distribution as a function of the predictor variables. As effects of learning and practice typically co-occur with fast improvements at the beginning of a new task (e.g., Dutilh et al., 2011, 2009; Heathcote et al., 2000), we used a penalized spline function to capture this potential non-linear effects in a data-driven way.

Second, we applied drift-diffusion (i.e., mechanistic) models on both experiments and sessions' RT and choice data. Mathematical models of RTs and accuracy have become increasingly popular (Voss et al., 2013), as some were specifically designed to provide interpretable coefficients to disentangle cognitive processes in binary choice tasks (and extensions thereof), e.g., to estimate speed-accuracy trade-offs. Moreover, they have been shown to have a very good fit for RT data (Wagenmakers, 2009), and some of them enable to model RTs and accuracy data concurrently.

doi: 10.53846/goediss-9841

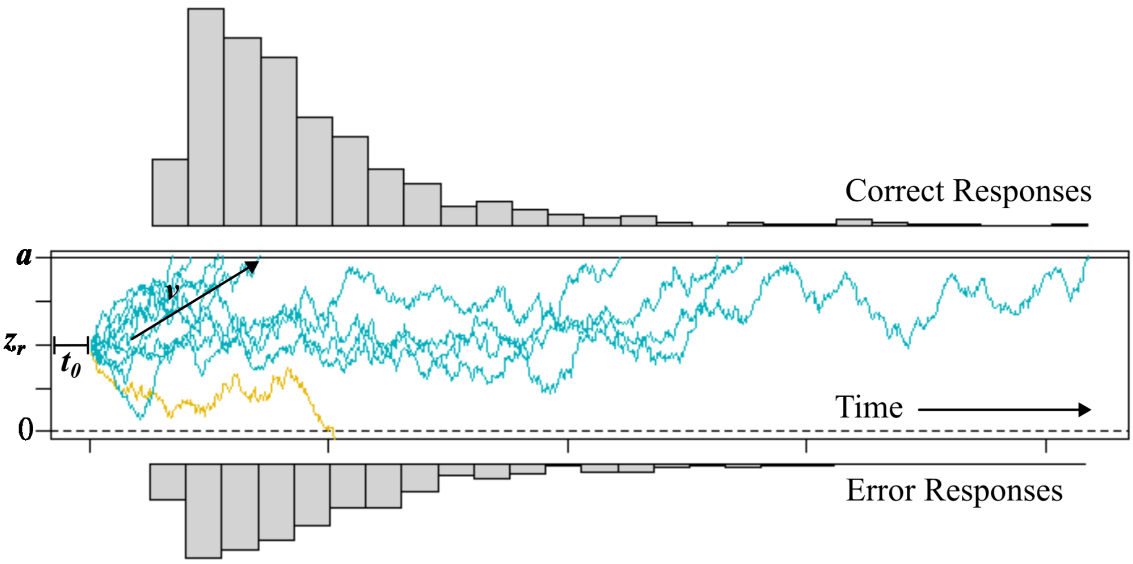
One of these models is the diffusion model, sometimes also called the Ratcliff diffusion model (Ratcliff & McKoon, 2008), which belongs to the class of sequential sampling models (Ratcliff et al., 2016; Ratcliff & Smith, 2004). The main assumption of the diffusion model is that in a speeded binary-choice task, the decision can be described as the accumulation of information over time until reaching one of two thresholds. The accumulation process can be described by a Wiener diffusion process characterized by a systematic drift v and Gaussian noise (Voss et al., 2015) and capturing a decisional process's systematic influence and random noise. The drift-diffusion model compromises several parameters for flexibility. The four main parameters are the drift rate v, which is a measure of the speed of information uptake, the starting point z_r reflecting a pre-decisional bias towards one of the boundaries, the boundary separation a informing about a certain decision criterion, and the non-decision time t_0 accounting for processes outside of the decisional process, e.g., the execution of a motor response or early stimulus encoding. With the drift-diffusion approach, we aimed to explore which changes in response times over learning would be accounted for by which parameters to possibly disentangle, e.g., effects of encoding and decision. An illustration of the diffusion process is shown in Figure 3.1.

Reports are mixed about what can be modeled with the drift-diffusion approach and about the specific requirements of the task. The basic drift-diffusion model was developed to handle rather fast decisional processes, e.g., excluding tasks with RT averages above 1.5 seconds (Ratcliff & McKoon, 2008) and multi-stage processes. However, simulation studies showed some robustness against changing parameters over time (e.g., Ratcliff & Tuerlinckx, 2002). Nevertheless, a minimum of trial numbers has been shown to be necessary for fitting. To account (partly) for this, we implemented a moving window approach to obtain a rough estimation for parameter changes across the learning and test sessions. Importantly, for uncontaminated data (i.e., no unsystematic slow or very fast responses), trial numbers of up to 200 to 500 trials per condition are beneficial for unbiased parameter estimation (Lerche et al., 2016). Not seldom, between 1000 to 10000 trials per participant have been reported (e.g., Dutilh et al., 2009). However, studies with fewer trials per condition also exist (C. J. Mueller & Kuchinke, 2016, 2015). Due to our learning paradigm, there was a natural limit of trial numbers per participant, condition, and stimulus repetition. Thus, we will report and discuss the findings of this exploratory approach while being aware that model estimates might lack some robustness.

We expected a general acceleration of RTs for all faces as a function of learning and not only practice with the task: At the beginning of the learning session, the processing of the face *and* the voice would be necessary to make the correct decision about whether they match their gender. However, with successful learning, the face itself would become more and more informative for

doi: 10.53846/goediss-9841

Figure 3.1
Illustration of ten trials of a simulated decision process.



Notes: The basic assumption of the drift-diffusion model is that information is accumulated over time, resembled by the drift rate v. The decision process ends with hitting one of the two thresholds (lower threshold 0 result in wrong answers depicted in yellow, upper threshold a results in correct answers depicted in turquise). The starting point z_r in this example is half the distance from 0 to a, but can theoretically be closer to one of the boundaries, which would result in a response bias (thus, the starting point is sometimes also called bias). The non-decision time t_0 subsumes all non-decisional processes, e.g., stimulus encoding and the motor response and corresponds to a shift in the RT distribution but does not affect accuracy. The histograms on top and below show the response times for the correct answers (top) and wrong answers (bottom). The RTs for the histograms were obtained by simulating 1000 trials of the depicted decision process with the same parameter values.

doi: 10.53846/goediss-9841

the correct decision, as the voice (and also the response key) were fixed for each face (and counterbalanced between participants). A further aspect, which might not be obvious at the first sight, is that participants might learn not only which voice would be presented but, in general, when a voice would be presented and can plan the motor response accordingly. Taken together, both associations should result in faster responses. Our main predictions for the drift-diffusion parameters were as follows: With learning, the informativeness of the face would lead to a time-shift of the decision process, i.e., the decision process would start earlier. However, the face's informativeness would only be indirectly captured due to the delayed response design (responses were only recorded *after* the onset of the voice, and whether a response had been recorded was indicated by the colour of a fixation cross). Thus, learning about the face would be reflected in a change of the starting point (z_r), i.e., a bias toward one of the boundaries. In addition to learning about the face, familiarity with the vocal stimulus might enhance its information uptake (Dutilh et al., 2009) and lead to faster recognition of the voice's gender, which should be reflected in positive changes in the drift rate v. However, v would resemble a mixture of the decision process about the face and the voice and its integration, especially at the beginning of learning when the decisional process would not be finished during face presentation. Moreover, in one of the experiments, we implemented a go/no-go condition, in which the participant had to await the onset of the sound stimulus to execute or inhibit responses. We suspected that with more familiar vocal stimuli, also this execution/inhibition process would be faster, leading to overall faster responses.

3.2 Method

The data in experiment one (Exp.1) were prepared and collected by Brückner (2019) as part of her Master's thesis. The data from experiment two (Exp.2) was part of the event-related potential (ERP) study described in Chapter 2 of this thesis. As many aspects were the same in both studies, we will highlight differences, which are, from our view, important for the interpretation of the results.

Participants

We analyzed data from 81 participants (Exp.1: 22 female, 18 male, 0 diverse, 19-31 years, M_{age}= 24, *SD* = 3.19; Exp.2: 30 female, 11 male, 0 diverse, 19 - 34 years, M_{age} = 23, *SD* = 3.59). Participants were reimbursed for their participation with either money (12 EUR in Exp.1, 8 EUR/h in Exp.2) or equivalent course credit.

doi: 10.53846/goediss-9841

Stimuli

Face stimuli were colored frontal portrait photographs with a neutral expression from the Göttingen Faces Database (Kulke et al., 2017). They were edited and combined with a transparency mask covering the hairline, ears, and neck and saved with a 200 $\times$ 300 pixels resolution. They were presented at visual degrees of approximately 8.09 $\times$ 12.08 (Exp.1) and 3.16 $\times$ 5.14 (Exp.2). Vocal stimuli were selected from the Montreal Affective Voices database (Belin et al., 2008) and consisted of laughter (happy), yells (angry), and sustained neutral tone (neutral). The selected stimuli for both experiments are listed in Table B1 of Appendix B.

Experimental Procedure

Both experiments were conducted in accordance with the Declaration of Helsinki and approved by the Ethics committee of the Institute of Psychology at the University of Göttingen. All participants gave written informed consent to participate in the study at the beginning of each session. The learning and test sessions took place on consecutive days.

Apparatus. *Experiment 1*: A maximum of six participants were tested simultaneously in a group laboratory with visual shielding between the participants. The computer-based study was presented with the Python module PsychoPy (Peirce, 2009) on Dell Notebooks (E5530) at a distance of approx. 50 cm from the participant. Auditory stimuli were presented at a constant volume for all participants over headphones (Beyerdynamic, DT 770 PRO). *Experiment 2*: Participants were tested individually in a sound-attenuated and electrically shielded testing room (in Exp.2, we simultaneously recorded the participant's EEG). The face stimuli were presented on a 24 inch monitor (BenQ, Zowie XL2411) with a distance of approx. 78 cm. Auditory stimuli were presented via speakers located at the monitor's height.

Learning task. In a Pavlovian procedure, we conditioned non-expressive faces (CS) with unconditioned stimuli (US) in the form of either happy (US_{hap}), angry(US_{ang}), or neutral affect bursts (US_{neu}), which could be either of the same gender as the face or of a different gender. The 12 face-voice pairs (50% with gender matching, 50% mismatching) were fixed within participants. Thus, every face predicted a specific burst and, at the same time, the correct response (gender matching vs. mismatching). Only in Exp.2 was face-voice pairing pseudo-randomized between participants (keeping the gender conditions balanced), whereas, in Exp.1, face-voice pairs were identical for all participants. The experimental task during learning was to decide whether the face-voice pairs

doi: 10.53846/goediss-9841

matched regarding gender. Each trial started with a central black fixation cross (500 ms), followed by the individual face stimulus (500 ms). A grey fixation cross and the auditory burst set in with the face offset. Participants were instructed to respond as fast and accurately as possible via key press as soon as the voice set in. In Exp.1, the two response keys were colored marked keys (“F” and “#”, German keyboard layout) which participants pressed with their index fingers; in Exp.2, participants used their dominant (right) hand only to press the answer keys (arrow keys). The key assignment was counterbalanced between participants. As soon as a response was recorded, the grey fixation cross was disabled. The next trial started earliest after the sound offset and a fixed 2000 ms inter-trial interval in Exp.1 or a variable inter-trial interval (*M* = 1800 ms; *SD* = 200 ms) in Exp.2. In addition, only Exp.1 included a time limit of 5 seconds, after which the next trial started irrespective of the recording of a response. Short breaks to rest after every fifth block were included in both experiments. One block was defined as the shuffled set of the 12 face-voice pairs, i.e., one repetition of a stimulus(-pair). Importantly, Exp.1 had an additional no-go condition with randomly interspersed no-go trials (90 trials) in which a beep was presented instead of the burst, and participants were instructed not to press any key. Furthermore, it entailed one-back tasks (30 trials) at random positions in which participants had to decide via key press (top and down keys) whether the face was the same as in the previous trial.

Test task. Only the previously conditioned faces without bursts or sounds were presented during the test session. Participants performed a gender decision of the face. Trials started with a black fixation cross (500 ms) which was replaced by the face stimulus (1500 ms in Exp.1, 1000 ms in Exp.2). Participants could respond via key press as soon as the face set in. Similar to the learning session, Exp.1 had a time limit to respond of 5 seconds, and Exp.2 contained interspersed one-back tasks (60 trials).

Data preprocessing. The lengths of Exp.1 and Exp.2 differed, with Exp.1 having 30 blocks during learning and 20 during test and Exp.2 having 50 blocks each. First, we trimmed RTs using a maximum cut-off of 5000 ms, resulting in a loss of 1% of trials for each experiment and session. It has been frequently proposed to trim data to a fixed minimum cut-off (e.g., 100 - 200 ms). However, due to the design of our study, such trimming could likely introduce a systematic bias in our data. In our experiment, for highly (en-)trained participants, it should have been feasible to time their responses close to the point of the onset of the voice. Thus, instead of setting a global cut-off value, we defined extremely fast responses based on the individual participant’s range, which has been shown to result in a higher power (Ratcliff 1995). To do so, we applied

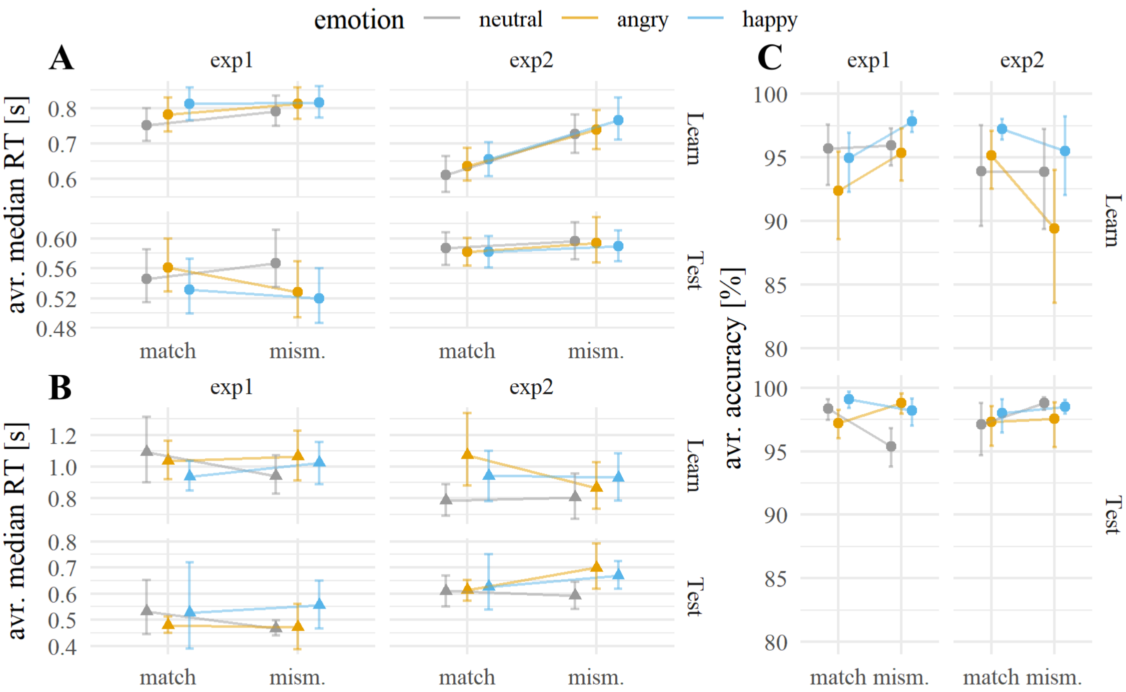
doi: 10.53846/goediss-9841

a skewness-adjusted boxplot method to exclude extreme values separately for every subject and session, using the function “adjbox” of the package “robustbase” (Maechler et al., 2021; based on: Hubert & Vandervieren, 2008). With this method, 501 (3.5%) observations for learning and 616 (6.4%) observations for test of Exp.1, 569 (2.3%) observations for learning and 1258 (5.1%) observations for test of Exp.2 were excluded.

3.3 Descriptive results

Figure 3.2

Median response times and accuracy per condition, session, and experiment.



Notes: Values are averaged over participants. Error bars are 95% non-parametric bootstraped Confidence Intervals around the mean. **A** shows response times in seconds for correct answers, **B** for wrong answers. **C** shows the accuracy as averaged percentages over participants.

Overall, behavioral responses differed between learning and test session and, albeit to a lesser degree, between the two experiments. As can be seen in Figure 3.2, correct responses in the learning sessions were slower in Exp. 1 with less pronounced differences between gender-matching and mismatching face-voice pairs. However, during learning, both experiments showed the fastest correct responses for neutral face-voice pairs. During the test session of Exp.1, faces conditioned with happy and angry gender-mismatching voices were responded to fastest, but there was an overall larger variation between participants’ answers. In contrast, RTs in the test session of Exp.2 were overall comparable between conditions, with a small difference between conditioned gender-

doi: 10.53846/goediss-9841

matching and -mismatching faces. A commonality between experiments was that during learning, wrong responses were slower than correct responses, whereas, during the test sessions, correct and incorrect responses were more similar. Moreover, the overall accuracy was comparable between experiments, although descriptively, during learning, Exp.1 showed higher accuracy for gender-mismatching face-voice pairs, whereas, for Exp.2, the gender-matching face-voice pairs' accuracy was higher. Accuracies in the test sessions were higher in both experiments. However, for Exp.2, accuracies were comparable between conditions, whereas, for Exp.1, faces conditioned with neutral bursts had lower accuracy within the conditioned gender-mismatching trials. Thus, in Exp.1, participants responded to faces conditioned with neutral bursts not only more slowly, but they appeared to be also more unsure about the faces' gender, indicated by the lower accuracy.

3.4 GAMLSS approach

We included only RT data of correctly answered trials in the GAMLSS model and analyzed each experiment and session independently. In the first step, we tested potentially suitable conditional distributions for the response time variable. Some theoretical distributions that account for typical RT properties (e.g., skewness, Luce et al., 1986) are the Lognormal (Zandt, 2002), the exponentially-modified Gaussian (ex-Gaussian) (e.g., Brewer, 2011; Hohle, 1965; Lacouture & Cousineau, 2008; Navarro-Pardo et al., 2013; Ratcliff & Murdock, 1976), the Gamma (e.g., J. R. Anderson, 1974; McGill & Gibbon, 1965), and the Weibull distribution (e.g., Palmer et al., 2007; for a comparison of these and others, see Zandt, 2000). For both experiments, we first fitted intercept models (without predictors) and compared the residuals. Moreover, we compared the density of the model distribution (by random sampling from the distributions) with the approximated density of the experimental data (see Figure B1 of Appendix B). The ex-Gaussian distribution matched visually the data best and in terms of the generalized Akaike Criterion (GAIC, Akaike, 1983), likely for the flexibility of a third parameter. The GAMLSS package includes a variety of different classes of distributions, of which also distributions with more than three parameters are available. A few distributions showed a similar or even a slightly better fit when more parameters were included (see Table B4 in Appendix B). However, to make results more comparable with previous RT studies, we restricted ourselves to a maximum of three parameters, of which the ex-Gaussian had the best fit for both experiments and sessions, although it did not capture the very left and right tail of the response distribution appropriately.

The ex-Gaussian is the convolution of two additive processes, a Gaussian (normal) function and an exponential function (Luce et al., 1986). The probability density function of the ex-Gaussian

doi: 10.53846/goediss-9841

distribution is defined as

$$f_Y(y|\mu,\sigma,\nu) = \frac{1}{\nu}\text{exp}\left(\frac{\mu - y}{\nu} + \frac{\sigma^2}{2\nu^2}\right)\Phi\left(\frac{y-\mu}{\sigma} - \frac{\sigma}{\nu}\right)$$

with $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ and $\nu > 0$, where Φ is the cumulative distribution function of the standard normal distribution. The mean of the ex-Gaussian is described as $\mu + \nu$ and the variance as $\sigma^2 + \nu^2$. The parameter ν is the mean of the exponential component and informs about the skewness of the distribution (Lacouture & Cousineau, 2008).

Model choice

The candidate model was set by our research question, i.e., we included the factor emotion (happy, angry, neutral), the factor congruence (match, mismatch), and block (z-transformed) and all higher-order interactions between the three variables in our model. To account for dependency in the data, we started with a maximum random structure (Barr et al., 2013; Marletta & Sciandra, 2020) and omitted random effects until convergence was reached, leaving intercepts for participants and a random slope for blocks for the estimation of the μ parameter. Thus, we estimated population effects for emotion, congruence, and block and their interactions on the expectation and variance of response times, while taking into account that participants might differ in how fast they respond in general and over time. By omitting random effects of emotion and congruence we could not account for individual differences in learning curves for each emotion and congruence level over time, which, from a theoretical point of view, would have seem plausible. However, this would have required different data or estimation methods due to the sparseness of the present data (Increasing the trial number per participant and condition without increasing the number of stimulus repetitions would have only been possible by including more distinct face-voice pairs).

Following the notation of Marmolejo-Ramos, Tejo, et al. (2022), the GAMLSS model can be described as follows,

$$g_k\left(\theta_k\right) = \mathbf{X}_k\beta_k + \sum_{j=1}^{J_k}\mathbf{Z}_{jk}\gamma_{jk},$$

where $g_k(\cdot)$ is the specific link function for each parameter θ_k (e.g., the distributional parameters μ, σ and ν in the ex-Gaussian case), the term $\mathbf{X}_k\beta_k$ is the product of the covariate design matrix and the vector of fixed effects. The second term $\sum_{j=1}^{J_k}\mathbf{Z}_{jk}\gamma_{jk}$ includes the the covariate design matrix Z_k and the vector of random effects γ_{jk}.

doi: 10.53846/goediss-9841

The model formula of the full model in common R syntax would be

```
responsetime    ~ emotion * congruence * pb(block.z) + re(random = ~block.z|sub_id),
sigma.formula = ~ emotion * congruence * pb(block.z),
nu.formula =    ~ emotion * congruence * pb(block.z)
```

Here, `*` abbreviates both the additive and the interaction term, e.g., `emotion * congruence` can also be rewritten as `emotion + congruence + emotion:congruence`. `pb(block.z)` is the penalized spline function for the z-transformed vector of blocks. We limited the number of knots for the splines relative to the number of blocks included in the data, which was different between experiments and sessions. For readability, we will in the following use the term "block" while referring to `pb(block.z)` if not explicitly stated otherwise. Note that the term for the random slope and random intercept for participant was only included in the formula for $\hat{\mu}$, which is represented by the first line. The models were "treatment"-contrast coded, i.e., the intercept refers to the estimated value of the reference levels, here: "neutral" (for emotion) and "match" (for congruence), and the estimates $\hat{\beta}$ to the respective slopes (or differences for factors). Block was z-standardized for each experiment and session, i.e., the obtained estimates refer to the midpoint of the respective experimental session.

We reported likelihood ratio tests (LRT) to compare the full model with a subset of reduced models (excluding one predictor at a time) in explaining the importance of a predictor for the overall model. Theoretically, it would be possible for predictors to impact, e.g., only μ but not σ or any other combination. However, since we did not have specific predictions about effects on individual parameters (μ, σ and ν), we only looked at the effects of the predictors on the *whole* distribution, thus dropping one term each from the full model in all parameters simultaneously. As a consequence, results inform only about the overall contribution of a predictor but not whether it affected exclusively the mean, the variance, or both. Since reporting model coefficients, including a non-linear function of time and interactions with the predictors for emotion and congruence, is not fully informative, we visualized the model-predicted averaged RTs in Figure 3.3 and Figure 3.5 for the learning and the test session, respectively, to give a better overview over the differences between conditions. The tables with the model coefficients for the learning and the test session can be found in Table B2 and Table B3 of Appendix B.

Results of the learning sessions

Exp.1 included 30 blocks and was shorter than Exp.2, which included 50 blocks of stimulus repetitions. Averaged over the learning session and irrespective of the condition, responses were

doi: 10.53846/goediss-9841

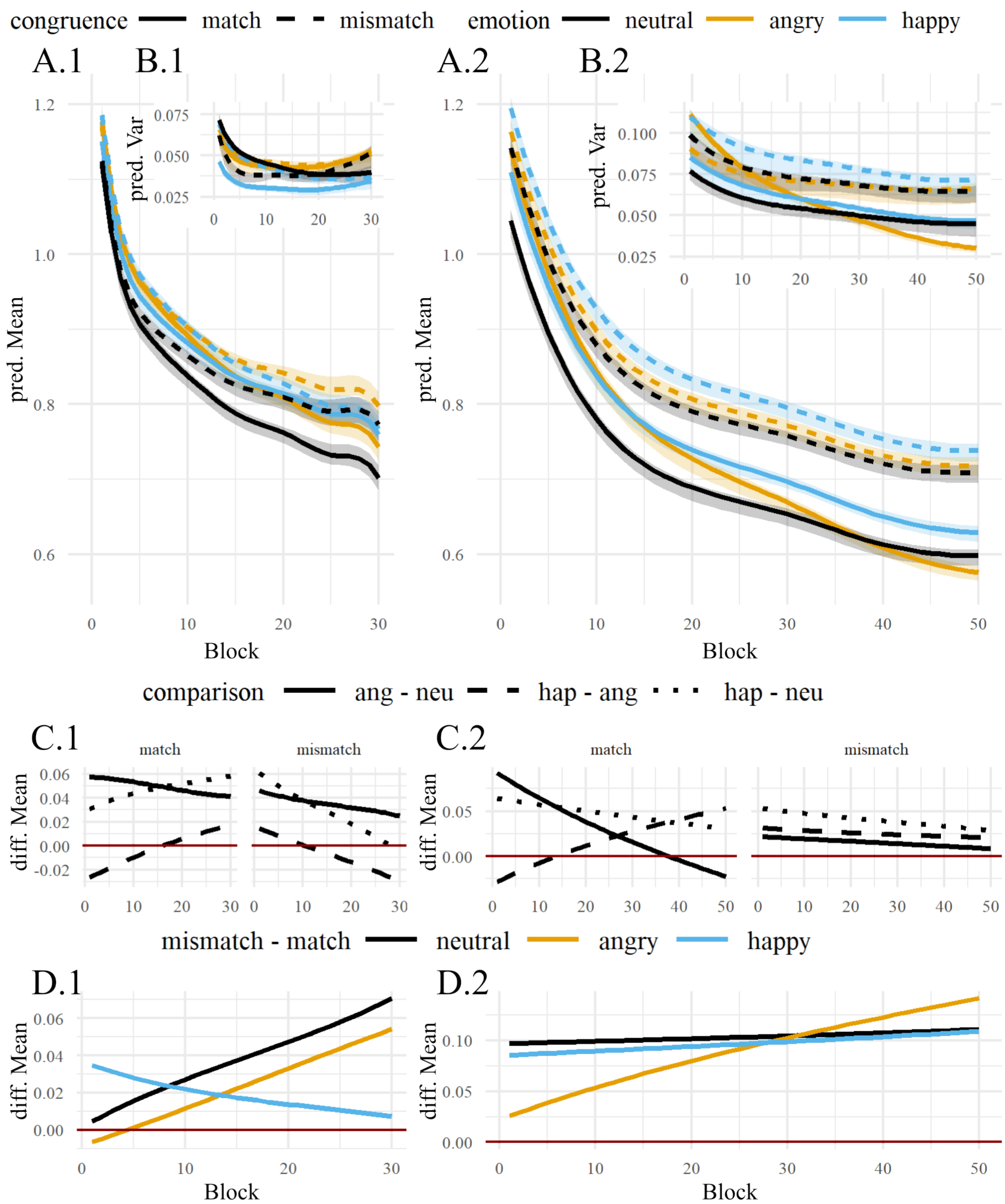
slower in Exp.1 (M = 848 ms, SD = 324) compared to Exp.2 (M = 774 ms, SD = 414). However, note that Exp.1 had fewer blocks, and responses became faster over time in both experiments. For both Exp.1 and Exp.2, according to LRTs, including a smoothing function for block resulted in a significantly better fit compared to including a linear function of block (Exp.1: $\chi^2(292.32) = 11.57$, $p < .001$; Exp.2: $\chi^2(679.72) = 10.81$, $p < .001$). Moreover, in both experiments, results of the LRTs of nested models suggested that the full model, including the three-way interaction of emotion $\times$ congruence $\times$ block, overall contributed significantly to explaining the variance in the data (Exp.1: $\chi^2(27.51) = 15.15$, $p = .026$; Exp.2: $\chi^2(46.75) = 15.16$, $p < .001$). This was similarly reflected by the comparison of the relative importance via GAIC for different penalties, although in Exp.1, not the full model but a model including block and interaction between emotion and congruence ranked highest (AIC = -4508.72, df = 109.2, k = 2). Increasing the penalty for the number of parameters, an additive model with emotion, congruence, and block but no interaction showed the best relative fit (GAIC = -3658.0, df = 103.2, k = 10). In contrast, in Exp.2, the full model (AIC = -3766.5, df = 125.1, k = 2) ranked highest. Increasing the penalty, the model with emotion, congruence, and block without interactions ranked highest(GAIC= -2920.3, df = 103.8, k =10). The results of the LRT models for learning are shown in Table 3.1.

Estimated moments of the RTs in learning session. The model-predicted global RT means and variances over blocks for both experiments are presented in Figure 3.3. Shaded areas depict the range of the predicted RTs based on all leave-one-out models as a model-stability measure. The predicted distributions and differences between condition levels for the learning sessions are shown in Figure 3.4. For learning, in both experiments, a non-linear change in RTs over time was observable, with the fastest decrease of RT at the beginning of the session. Moreover, there was some indication of different changes between emotion and congruence conditions over time. Moreover, not only the mean but also the variance of RTs decreased over time, as visible in the shape of the distributions. Overall, mismatching trials were slower than matching trials in both experiments, whereas the difference was more pronounced in Exp.2. Remarkably, the RT differences between gender congruent and incongruent trials remained relatively stable across emotion categories or increased even slightly, with the exception of the happy condition in Exp.1 for which the differences decreased. Another similarity between experiments was that overall, neutral face-voice pairs were responded to faster than the other emotional categories across congruence conditions. Within matching pairs, the interaction between angry trials and block indicated that, despite starting slower, angry trials showed a stronger RT decrease over time than neutral and happy pairs, particularly in Exp.2. There were also differences between experiments: Emotion effects were

doi: 10.53846/goediss-9841

Figure 3.3

Model-predicted RT mean and variance of the learning sessions.



Notes: All estimates are based on the full model and included only correct answers. **A.1** displays the predicted RT mean in seconds over blocks of Exp 1., **B.1** the predicted variance. Shaded areas show the range of predicted values for the leave-one-out models, i.e. predictions when excluding one participant. **C.1** represents the predicted emotion differences of mean RTs separately for matching and mismatching trials. **D.1** represents the predicted congruence difference of mean RTs separately for each emotion level. **A.2**, **B.2**, **C.2**, and **D.2** show the corresponding values for Exp 2.

doi: 10.53846/goediss-9841

Table 3.1

LRT of the GAMLSS models of the learning sessions

	Exp1			Exp2		
LRT:Model	χ^2	df	p	χ^2	df	p
pb(block) vs. block in the full model (FM)	11.57	292.32	<.001	10.81	679.72	<.001
emo × cong + pb(block) vs. FM	15.15	27.51	.026	15.16	46.75	<.001
cong × pb(block) + emo vs. FM	18.06	55.01	<.001	18.22	43.44	<.001
emo × pb(block) + cong vs. FM	15.17	56.92	<.001	15.11	34.58	.003
emo + cong + pb(block) vs. emo × cong + pb(block)	6.00	36.25	<.001	6.07	10.84	.097
emo + cong + pb(block) vs. emo × pb(block) + cong	5.97	6.84	.333	6.13	23.01	<.001
emo + cong + pb(block) vs. cong × pb(block) + emo	3.08	8.75	.035	3.02	14.15	.003
emo × pb(block) vs. emo × pb(block) + cong	2.90	43.04	<.001	2.91	686.94	<.001
cong × pb(block) vs. cong × pb(block) + emo	6.25	188.85	<.001	6.28	97.99	<.001
emo × cong vs. emo × cong + pb(block)	13.43	368.78	<.001	12.23	1025.21	<.001

Notes: FM = full model including all interactions in μ, σ and ν; pb(block) = penalized spline smoothing function for the z-transformed block variable; emo = emotion, cong = congruence; the model formulas for μ, σ and ν was kept parallel, i.e. dropping one factor of a model means dropping this factor for μ, σ and ν.All models included the random intercept of participant and slope of block for μ.

overall more pronounced in Exp.1 than in Exp.2, and dynamic differences between emotion levels within the mismatch condition were larger for Exp.1, whereas they were largely absent in Exp.2.

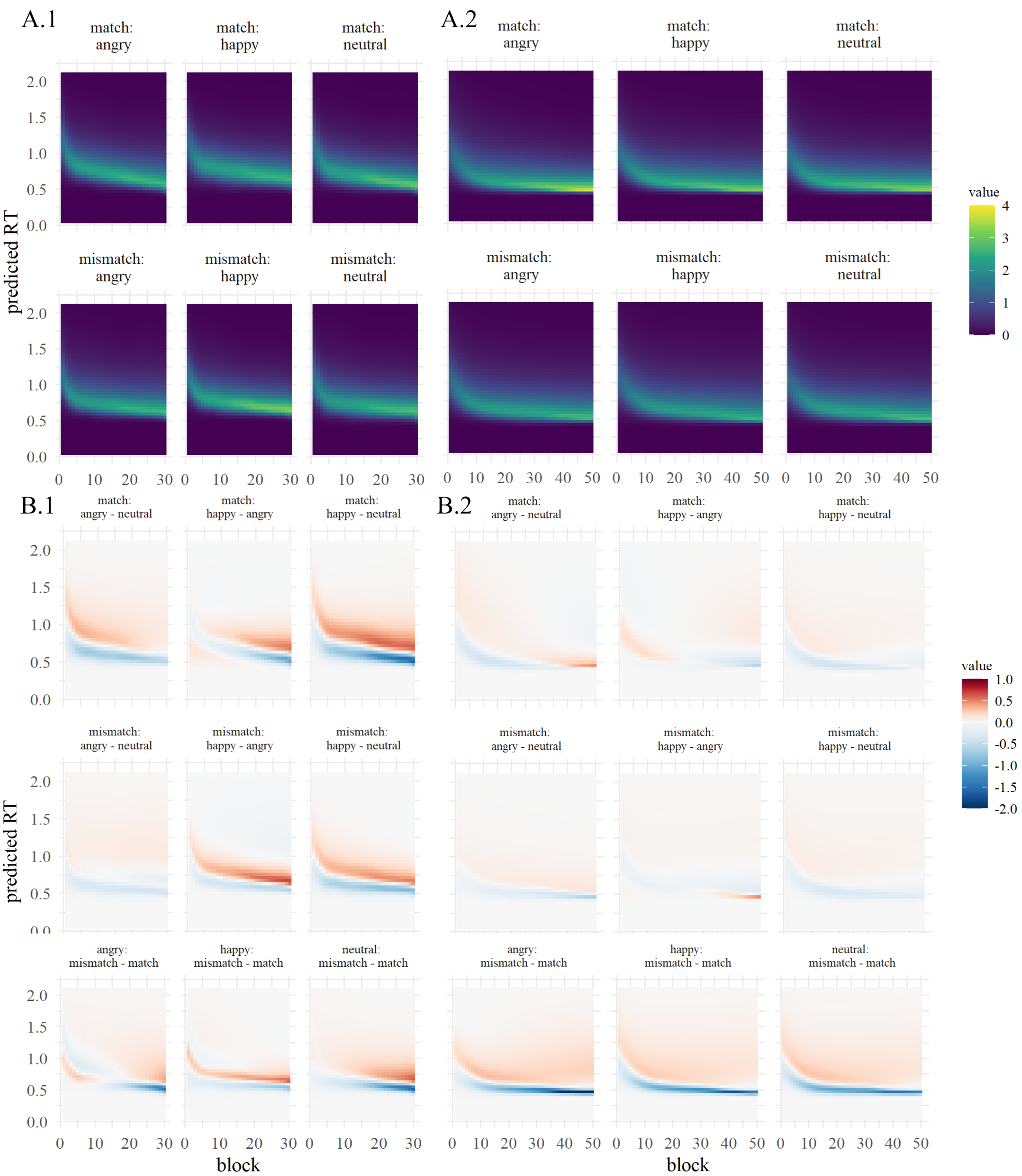
Results of the test sessions

Responses of the test session were, on average, faster than RTs of the learning session. Moreover, responses were faster in Exp.1 (*M* = 563 ms, *SD* = 181) than in Exp.2 (*M* = 617 ms, *SD* = 206), but, similar to the learning session, the test session of Exp.1 contained data of only 20 blocks, whereas Exp.2 included 50 blocks. LRTs suggested that in the test session, including a smoothing function for block instead of a linear function resulted only in a better fit in Exp.1 (χ^2(92.20) = 8.19, p <.001), but not in Exp.2 (χ^2(6.55) = 3.79, p = .144). Moreover, in neither of the experiments, the full model, including the three-way interaction, significantly improved the data fit compared to a model with block as an additive factor (Exp.1: χ^2(18.35) = 14.67, p = .226; Exp.2: χ^2(15.05) = 15.04, p = .451). In Exp.1 and Exp.2, the two-way interaction of emotion × congruence was significant (Exp.1: χ^2(116.86) = 6.05, p <.001; Exp.2: χ^2(16.73) = 6.20, p = .012). In Exp.1, there was also a trend for the interaction between emotion × block (χ^2(11.98) = 5.85, p = .058). In contrast, in Exp.2, none of the models which included the interaction with block improved the data fit significantly. According to ranked GAICs, in Exp.1, the model involving the interaction between emotion and congruence and an additive term for block was the best (GAIC = -10734.88, *df* = 100.1, k = 2), also when increasing the penalty (GAIC = -9934.44, *df* = 100.1, k = 10). In Exp.2, the highest ranking model included the main effects and interaction between congruence and emotion, but not block (k = 2; GAIC = -38265.49, *df* = 92.95). Increasing the penalty (k = 10) for Exp.2, the null model had the lowest GAIC (-37559.28, *df* = 77.83). The results of the LRT

doi: 10.53846/goediss-9841

Figure 3.4

Predicted density plots over block for the learning sessions.



Notes: **A.1** shows the predicted RT densities of Exp.1 over blocks. **B.1** shows pairwise comparisons between emotion levels within congruence levels and differences between congruence levels within emotion levels for Exp.1. **A.2**, **B.2** show the predicted distributions and differences for Exp 2.

doi: 10.53846/goediss-9841

Table 3.2
LRT of the GAMLSS models of the test sessions

	Exp1			Exp2		
LRT:Model	χ^2	df	p	χ^2	df	p
pb(block) vs. block in FM	8.19	92.20	<.001	3.79	6.55	.144
emo × cong + pb(block) vs. FM	14.67	18.35	.226	15.04	15.05	.451
cong × pb(block) + emo vs. FM	17.70	133.64	<.001	18.24	31.12	.030
emo × pb(block) + cong vs. FM	14.88	123.24	<.001	15.00	21.17	.132
emo + cong + pb(block) vs. emo × cong + pb(block)	6.05	116.86	<.001	6.20	16.73	.012
emo + cong + pb(block) vs. emo × pb(block) + cong	5.85	11.98	.058	6.24	10.62	.113
emo + cong + pb(block) vs. cong × pb(block) + emo	3.02	1.58	.669	2.99	0.66	.882
emo × pb(block) vs. emo × pb(block) + cong	3.04	21.93	<.001	3.02	76.65	<.001
cong × pb(block) vs. cong × pb(block) + emo	6.81	193.87	<.001	6.05	19.38	.004
emo × cong vs. emo × cong + pb(block)	9.68	110.96	<.001	6.80	8.83	.248

Notes: FM = full model including all interactions in μ, σ and ν; pb(block) = penalized spline smoothing function for the z-transformed block variable; emo = emotion, cong = congruence; the model formulas for μ, σ and ν was kept parallel, i.e. dropping one factor of a model means dropping this factor for μ, σ and ν. All models included the random intercept of participant and slope of block for μ.

models for the test session are shown in Table 3.2.

Estimated moments of the RTs in the test session. As can be seen in Figure 3.5, leave-one-out estimates were more variable in Exp.1 compared to Exp.2, suggesting stronger influences on the model by individual participants and thus lower stability of estimates. Remarkably, in Exp.1 but not in Exp.2, the happy and angry mismatching trials were, relative to neutral trials, overall slower in the learning session but then faster in the test session. Furthermore, in Exp.2, differences between emotion categories were smaller compared to Exp.1, and emotional and neutral matching trials were largely overlapping, particularly toward the end of the test session. In Exp.2, differences between (previously) matching and mismatching trials, albeit smaller compared to learning, were still observable during test. The differences between emotion effects in Exp.1 and Exp.2, as well as the larger consistency of congruence effects in Exp.2, are well observable when comparing the differences of the RT distributions (see Panels B of Figure 3.6).

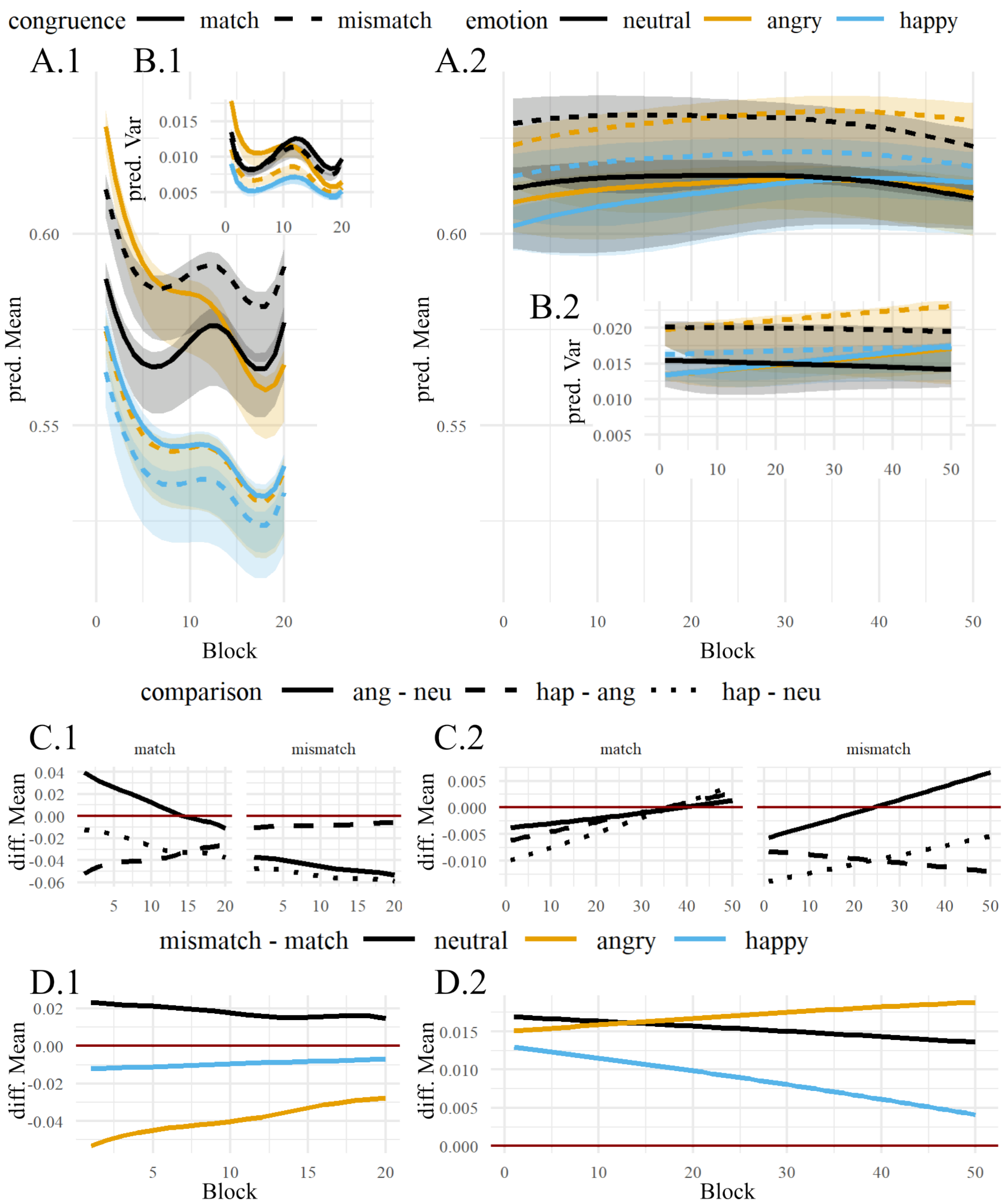
3.5 Drift diffusion approach

For the diffusion models, we included both correct and incorrect answers in the data. It has been suggested to exclude especially fast RTs (Voss et al., 2015), as those can bias estimates particularly. However, in the case of our learning paradigm, very fast responses do not necessarily correspond to unintentional responses but could be meaningfully interpreted as the result of training. The anticipation of the exact point in time when answering was allowed (i.e., with face offset) could be an important consequence of learning, i.e., learning both about the “when” and not only

doi: 10.53846/goediss-9841

Figure 3.5

Model-predicted RT mean and variance of the test sessions.

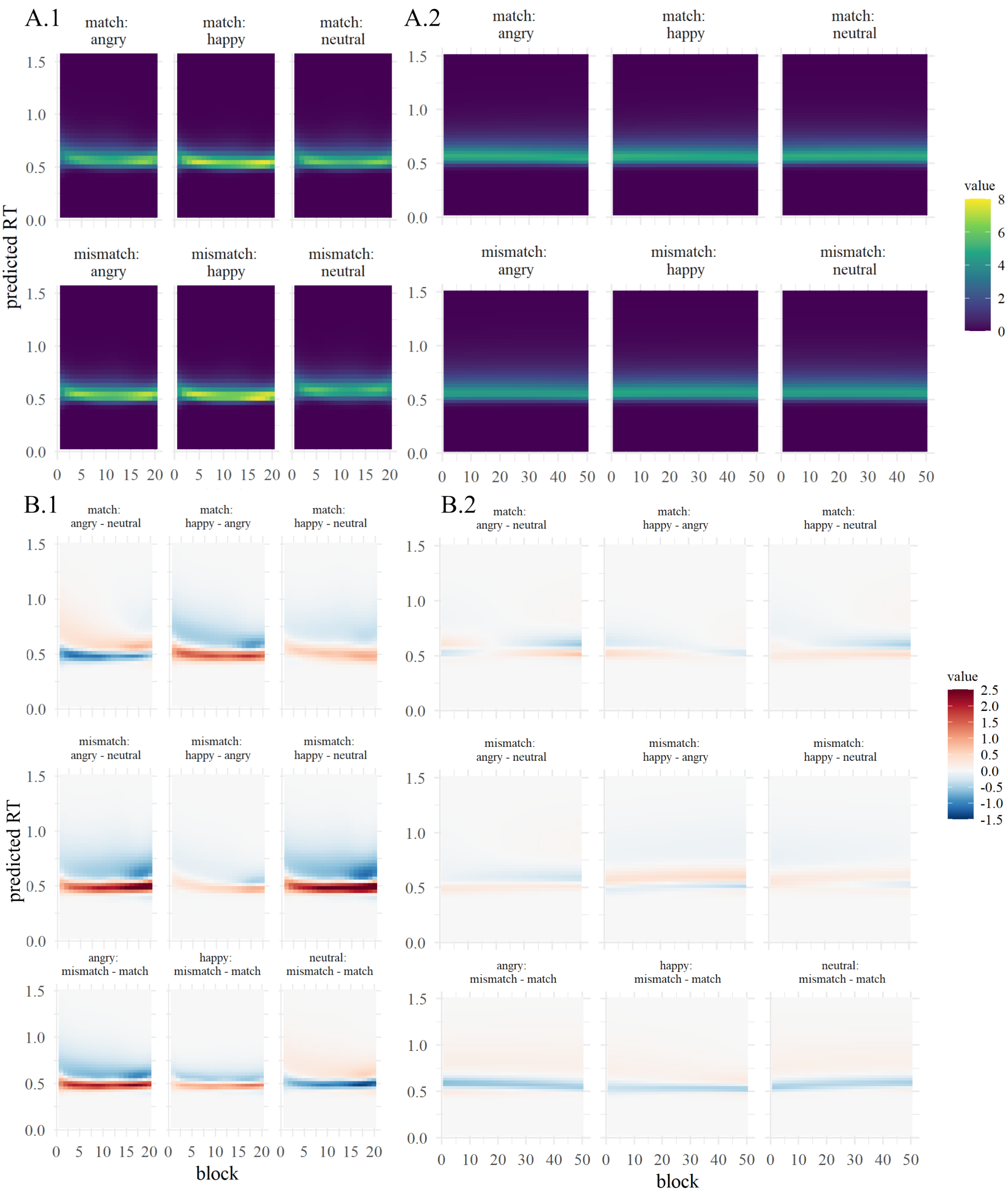


Notes: All estimates are based on the full model and included only correct answers. **A.1** displays the predicted RT mean in seconds over blocks of Exp 1., **B.1** the predicted variance. Shaded areas show the range of predicted values for the leave-one-out models, i.e. predictions when excluding one participant. **C.1** represents the predicted emotion differences of mean RTs separately for matching and mismatching trials. **D.1** represents the predicted congruence difference of mean RTs separately for each emotion level. **A.2**, **B.2**, **C.2**,and **D.2** show the corresponding values for Exp 2.

doi: 10.53846/goediss-9841

Figure 3.6

Predicted density plots over block for the test sessions.



Notes: **A.1** shows the predicted RT densities of Exp.1 over blocks. **B.1** shows pairwise comparisons between emotion levels within congruence levels and differences between congruence levels within emotion levels for Exp.1. **A.2**, **B.2** show the predicted distributions and differences for Exp 2.

doi: 10.53846/goediss-9841

about the “what” (Balsam et al., 2010). Thus, we did not set a fixed lower threshold for responses but used the same adjusted-boxplot method as for the correct RTs to detect extreme values separately for every participant and condition. Nevertheless, two participants (ID25 and ID32) did not provide enough trials for the parameter estimation of all conditions of interest and hence were excluded from all drift-diffusion models.

Model estimation was carried out using the software fast-dm-30.2 (Voss et al., 2015). The data for the models were accuracy coded, such that the upper boundary resembles correct responses and the lower boundary incorrect responses. The models were fitted for each participant and session separately, including a global model over the whole experimental session (without a variable of experiment time or stimulus repetition). As we were interested in the change of parameters over time, but the estimation of the drift-diffusion parameters requires a minimum number of data points, we additionally fitted sub-models, including a subset of blocks in a sliding window fashion, for which we created dummy variables with ten blocks each in steps of five (i.e., blocks 1-10, 6-15, 11-20, an so on) for each participant, condition, and parameter set. We fitted models with different parameter sets on the data and tested whether data fit improved when the drift or/and starting point were allowed to vary with the conditions of interest. The full model included both z_r and v depending on emotion and congruence. The boundary separation a and the non-decision time t_0 were always estimated globally, and the remaining available parameters were fixed (for more details, see Voss et al., 2015) to reduce the complexity of the model for estimation of sparse data (Lerche et al., 2018). In addition, to account for the small trial numbers per condition, we used the Kolmogorov-Smirnov (KS) method, which has been reported to have the highest robustness in the case of outliers (Lerche et al., 2016). When comparing fits, the full model, allowing both z_r and v to depend on emotion and congruence, showed overall the best fit. The estimated parameters of the full model for both sessions and experiments were averaged over participants, and non-parametric bootstrapped 95% CIs were obtained. The results for the drift-diffusion parameters are shown in Figure 3.7. We inspected model fits by comparing empirical median RTs with theoretical median RTs based on the model parameters for each participant, block window, and condition. The models showed a good fit for correct answers, but there were substantially larger misfits for wrong answers, such that the largest differences between predicted and observed median RTs included mainly overestimations of RTs for incorrect answers, likely due to the low amount of incorrect answers overall. To check whether the stability of findings was affected by these specific model misfits, we compared the drift-diffusion parameters, including all participants, with those obtained when excluding participants for which models showed strong misfits. As a cut-off for being classified as a misfit, we chose all predicted RTs exceeding 5000 ms, which was also the overall upper

doi: 10.53846/goediss-9841

cut-off of our study. This eliminated not all but the majority of strongly deviating predictions. The comparison of estimated diffusion parameters for both the full and the reduced participant set suggested overall similar results. Observed vs. predicted median RTs before and after exclusion for Exp.1 can be found in Figure B2 and for Exp.2 in Figure B3 of Appendix B. The drift-diffusion parameter estimates of the subset (excl. misfits) are shown in Figure B4 of Appendix B.

Results of the learning sessions

Independent of the model (global vs. time), in both experiments, there was a difference in the starting point z_r for match and mismatch trials, with larger values for the matching condition. Moreover, starting points of the matching condition shifted towards the upper bound as a function of block in both experiments, especially within the first 30 blocks. However, for the mismatching condition, there was no such shift. With regard to emotion, in both experiments, the starting point was overall highest for the gender-matching neutral condition. A difference between matching happy and angry pairs appeared only in Exp.1, with higher z_r for angry pairs, whereas levels were comparable in Exp.2. Moreover, within the mismatching condition, differences in emotion levels were more pronounced in Exp.1 compared to Exp.2. In contrast, drift rates v were slightly increasing in both congruence conditions and experiments over time. While for Exp.1, the drift rate increased more strongly for the mismatching trials, for Exp.2, it was for the matching trials. The largest emotion difference in the drift rate was in Exp.1 between happy and the other emotion levels of the mismatching condition. Both a and t_0 decreased over time in both experiments, i.e., correct and incorrect answers became faster over time (due to t_0), and correct responses became even faster while maintaining high accuracy (due to lower a and higher v). Notably, when comparing average parameters of the global model and the models over time, in both experiments, the non-decision time parameter t_0 of the global model was closer to the minimum of t_0 over time, opposite to the boundary separation a, which was globally closer to the maximum over time. z_r in the global model tended to be smaller compared to z_r over time. For v, there was no such systematic bias.

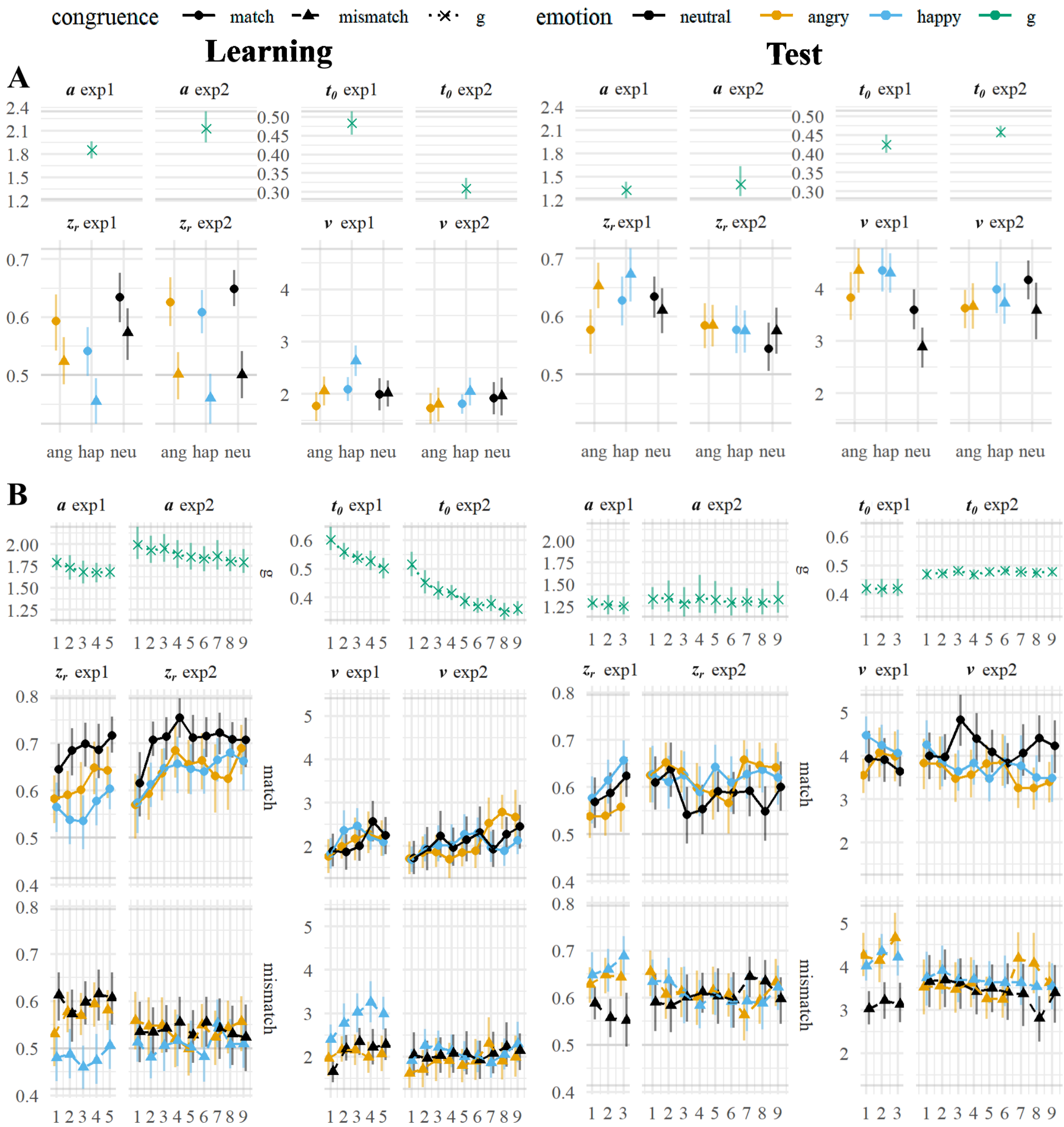
Results of the test sessions

Compared to learning in both experiments, the global model of the test session showed overall lower values for the boundary separation a, i.e., a lower response caution. Non-decision times t_0 were more similar between Exp.1 and Exp.2 compared to the learning session. Both starting point z_r and drift rates v tended to be more homogeneous between emotion levels in Exp.2 than in Exp.1. Most notably, drift rates in the test session were, irrespective of the condition, overall

doi: 10.53846/goediss-9841

Figure 3.7

Averaged drift diffusion parameters for the learning and test session.



Notes: The left panel shows the parameters for the learning sessions, the right for the test sessions. Error bars indicate 95% non-parametric bootstrapped confidence intervals around the sample mean. The boundary separation a and the non-decision time t_0 were estimated independent of the stimulus conditions. The starting point z_r and drift rate v were allowed to vary between conditions, i.e., we obtained separate estimates for each emotion and congruence level. **A** shows the global model across the sessions. **B** shows parameters changes as a function of stimulus repetition. Here, the x-axis refers to the sliding block windows of the experiment, of which one corresponds to blocks 1-10, two to 6-15, etc.

doi: 10.53846/goediss-9841

higher compared to learning. Parameter estimates for the boundary separation and non-decision times were largely constant over time in both experiments. Regarding the starting point z_r and the drift rate v of Exp.1, it is noteworthy that within mismatching trials, there was a difference between emotional and neutral conditions with lower values for neutral. In contrast, in Exp.2, the mismatching condition showed more similar estimates for the emotion categories compared to the matching condition. Descriptively, there was a slight tendency for neutral matching trials in Exp.2 to have a higher drift rate and a lower starting point, which was different from the learning session.

3.6 Discussion

We used a distributional and a mechanistic approach to gain insight into the dynamical change of response times (RTs) as a function of learning and extinction in cross-modal associative learning studies. In the two studies from which the data were analyzed, non-expressive faces were associated with either happy, angry, or neutral affect bursts of the same or different gender as the face. The participant's task was emotion-unrelated, and they had to decide about the gender-congruence of the face-voice pairs (learning) or the gender of the face (test). Despite similar stimuli and associative paradigms, the experimental designs of the two studies differed in several aspects, including total duration, randomization, and, most importantly, the paradigm of Exp.2 included a go/no-go task. Although the task of both studies obtained delayed answers during learning (answers could be given with the onset of the voice), Exp.2 included an additional decision process to execute or inhibit responses and thus allowed to compare performance results for the two experiments with different stages of decisional processes. To be able to compare RTs over time and to capture the dynamics for the individual emotion and congruence levels, we used GAMLSS models and included a non-linear function for stimulus repetition. Using drift-diffusion models, we aimed to obtain interpretable parameters that help disentangling the effects of practice with the task and learning about the conditioned stimuli. In addition, we estimated and compared standard drift-diffusion parameters for the learning and test sessions of the two associative learning studies to explore which parameters change as a function of learning. Moreover, we explored whether the standard drift diffusion parameters would be responsive at all for our experimental task which included delayed responses and entailed multiple-stage decision processes.

Our overall hypothesis was that faces associated with emotional voices would gain more relevance than those with neutral voices. Thus, emotional relevance would facilitate decisional processes, which would translate to faster, correct answers. Across experiments, participants showed clear signs of practice and became faster throughout the learning session, especially at the begin-

doi: 10.53846/goediss-9841

ning (similar to Dutilh et al., 2011, 2009; Heathcote et al., 2000). Because accuracy was overall high, we focussed on the correct answers in the GAMLSS approach and included wrong responses only in the drift-diffusion models. In contrast to the emotional-relevance hypothesis, for correct answers during learning, not the emotional but the neutral face-voice pairs resulted in the fastest RTs across experiments, whereas faces associated with happy voices were responded to slowest. Moreover, gender-mismatching face-voice pairs differed from matching face-voice pairs, particularly in Exp.2, which showed faster responses for congruent pairs (see also Huestegge et al., 2019; Latinus & Taylor, 2006). Both results suggest that congruent information, i.e., emotional (neutral-neutral) and gender congruence, facilitated information uptake and decision in the gender matching task. Possibly, emotional voices were more difficult to disengage with (e.g., Carretié, 2014; Dresler et al., 2008; Hur, Iordan, Dolcos, et al., 2016; Schimmack, 2005) and impaired focussing on the actual, gender-matching task, whereas (perceived) gender-mismatching face-voice pairs might have interfered the integrated person perception (Föcker et al., 2011; Gelder & Vroomen, 2000).

In addition to the general acceleration of RTs during learning, both experiments showed similar dynamics between emotion levels. Whereas on average, matching neutral trials were the fastest, matching angry trials showed steeper slopes, i.e., became relatively faster over time in both experiments. In Exp.2, which had a longer session duration and more repetitions, matching angry pairs even surpassed neutral pairs toward the end. A further common result between the experiments was the difference between gender-matching and mismatching pairs, which became more pronounced over the course of the session, i.e., increased with learning.

As hypothesized, drift-diffusion parameters of the learning sessions changed over time, i.e., with stimulus repetitions. We expected mainly changes in the starting point z_r as a function of learning, as the decision process might have already been at an advanced stage before the voice onset (i.e., before the actual response was recorded). Indeed, such a shift was apparent in both experiments and was most likely a result of the delayed response task (cf. Dutilh et al., 2011). Remarkably, z_r increased more in the gender-matching trials, whereas for the mismatching trials, z_r remained largely constant and at a lower level. Moreover, matching neutral trials overall had the highest z_r, corroborating that neutral matching trials were fast *and* accurate. A small difference between matching and mismatching z_r was already observable at the beginning of the session in both experiments, possibly due to the overall difficulty of integrating subjective mismatching gender information of face-voice pairs. Unfamiliarity with stimulus categories, e.g., words vs. non-words, has been shown to bias responses toward the more familiar answer, especially without previous practice (Dutilh et al., 2011, 2009). The differences between matching and mismatching and emotion levels of z_r suggest that the changes in starting point indeed reflect learning processes

doi: 10.53846/goediss-9841

about the stimulus and not merely the effects of becoming familiar with the task. However, also non-decision times t_0 became shorter, and boundary separations a became smaller throughout the learning session. To truly disentangle whether those changes reflect practice with the task or stimulus-specific effects, it would be necessary to include novel stimuli intermixed with the learned ones, similar to Dutilh et al. (2011). The effect of practice on non-decision times has been related to the facilitation of memory access and stimulus encoding. Perceptual learning of a specific stimulus and changes in early sensory areas constitute part of the total learning effect (Gilbert et al., 2001; Petrov et al., 2011). However, these effects are not always observable, e.g., C. C. Liu & Watanabe (2012) related all practice effects to the decision time, whereas early encoding or motor responses were unaffected by training. In both of our experiments, drift rates v increased over time. In Exp.1, both matching and mismatching trials benefited from stimulus repetition. This was slightly different for Exp.2, in which matching face-voice pairs showed a slightly stronger increase in v. We suspect that either sensory encoding (t_0) and the starting point z_r only partly captured the acceleration of response times, and thus, the remaining variation needed to be accounted for by the drift rate, or the increase in v described the learning about the *vocal* stimulus. In Exp.1, there was an overall higher drift rate for mismatching happy pairs but also a lower starting point for this condition, which needs to be considered together. We observed the effects of learning on several parameters, which could indicate functional dissociable mechanisms (e.g., Voss et al., 2004; J. Zhang & Rowe, 2014).

The experimental design of our studies allowed not only for learning about the face-voice pairs but also about the timing when stimuli will be presented. We included fixed durations of the fixation cross and the face stimulus. Thus, participants could learn not only about *which* voice would follow a face but also *when* faces and voices would become available. How temporal synchronization might help to time the decision process was tested in Petrov et al. (2011). Using a beep that systematically preceded the stimulus onset, they tested whether participants would use the temporal structure to synchronize the decision process to the stimulus onset. The authors reported a decrease in non-decision times and their variability with practice, suggesting that the temporal synchronization contributed to the learning process. In addition, increased drift rates were found to be partially stimulus-specific, indicating that learning improved the quality of the sensory input. Unfortunately, we can not disentangle stimulus encoding from motor planning processes in our design, as both the onset of the facial stimulus and the onset of the voice could have been learned. Therefore, future studies could systematically vary the predictability of encoding the stimulus vs. planning the motor response to give insight into the relative contributions on acceleration that are comprised in non-decision times.

doi: 10.53846/goediss-9841

The test sessions differed from the learning sessions in that way that only a single decision process and no delayed task were present. Only previously conditioned faces were presented (without voices) that had to be categorized into male and female faces. Thus, when referring to emotion or congruence levels, we exclusively refer to what was associated during learning.

Overall, RTs of the test session were faster compared to the learning session, suggesting that the decision on the gender on the face was easier than the gender-matching decision on face-voice pairs. Although tasks of the test session did not differ between the experiments, RTs in Exp.1 were faster than RTs in Exp.2 and showed more variation over the course of the session and between conditioned emotion levels. Whereas general variability might be traced back to a shorter learning session in Exp.1, the greater emotion effects were in contrast to our predictions. Remarkably, in Exp.2, there was still a gender-congruence effect present with faster decisions for faces previously associated with a gender-matching voice. Since conditioned gender-congruence was irrelevant for the task in the test session, and all faces were presented equally often during learning, this effect has to be caused by the learned association between faces and voices. Gender-mismatching face-voice pairs could have introduced some uncertainty about the gender of the face, which, in consequence affected the gender-decision on the face also in, and importantly, throughout the test session. Alternatively, the gender-congruence effect of the test session could have been caused by a more general impaired processing and integration of mismatching stimulus pairs in learning (see, e.g., Huestegge et al., 2019; Latinus et al., 2010). However, since congruence was not relevant for the task in the test session, we would have expected to see this difference rather at the beginning of the session and then to decrease over time.

Participants of Exp.1 showed an overall acceleration at the beginning of the test session, which was not apparent in Exp.2. Notably, there was no observable change in t_0 at the beginning of the test sessions. Dutilh et al. (2011) reported decreased t_0 for repeated compared to novel stimuli and interpreted this as a facilitation of the perceptual process, which is included in the non-decision time. Possibly, due to the high number of repetitions during learning, familiarity with the stimulus was already established such that more repetitions at test did not improve encoding significantly. Moreover, lower boundary separation a and high starting point values z_r indicate that the task during test (gender decision of the face) was overall easier compared to learning (gender-congruence decision of the face-voice pair). That correct responses were faster compared to wrong responses was indicated by z_r being closer to the upper boundary.

Parameter differences between emotion levels were only observable in Exp.1, with faster responses for mismatching emotional compared to mismatching neutral faces, reflected in pronounced differences for z_r and v. Other influences of emotion on drift-diffusion parameters were

doi: 10.53846/goediss-9841

inconclusive, e.g., the higher drift rate in Exp.2 for matching neutral faces corresponded to smaller z_r in some of the time windows, probably reflecting model instability.

Limitations

The application of the drift-diffusion account to our specific experimental design had, aside from interesting findings, also some limitations. First, the small number of trials per condition might have caused issues with parameter estimation despite using a rather robust method (KS). Moreover, although there exist applications of diffusion models in learning, e.g., reinforcement learning (e.g., Fontanesi et al., 2019; Pedersen & Frank, 2020), only a few studies reported changes of diffusion parameters as a function of other types of learning, e.g., practice (e.g., Dutilh et al., 2011, 2009) and perceptual learning (e.g., J. Zhang & Rowe, 2014). A commonality of these studies is that they usually still include a larger number of trials per condition and block (usually at least > 200, e.g., Dutilh et al., 2009), whereas we included 20 per condition and window. Although it was reported that reliable results could even be obtained for small trial numbers (e.g., Lerche et al., 2018; Metin et al., 2013), future studies should replicate the findings with an increased number of face-voice pairs.

Some results of our study were difficult to interpret. For example, the non-decision times of learning in Exp.1 were overall longer compared to Exp.2, which could not be explained by the different settings or hardware equipment. If that had been the case, we would have expected a similar pattern in the test session, which we did not find. Moreover, it seemed not very plausible that due to the individual testing situation in the EEG study (Exp.2), participants might have tried not to make mistakes (higher a) or that less distraction might have facilitated sensory encoding (smaller t_0). On the contrary, it would have been more intuitive if t_0 had been longer in Exp.2 during learning because of the no-go condition. In the end, it was for this specific task that participants had to wait for the auditory stimulus to decide on the execution or inhibition of their response. Moreover, the change of direction of the effects in Exp.1 from learning to test was contrary to our intuition. Although it could be argued that the facilitation of emotion-based associations on RT might rather be present during test, i.e., when emotional sounds were not distracting, it is difficult to explain why this should be stronger for previously gender-mismatching faces. In line with this, particularly problematic is the missing randomization or counterbalancing in Exp.1 of the face-voice pairs. Hence, potential emotion and congruence effects in this study might at least partly be due to specific effects of the face or voice stimulus (e.g., some were more distinctive compared to others and thus easier to learn). Lastly, we would like to acknowledge that the Cognitive Sciences have produced a variety of theoretical and computational models of associative learning (e.g., Cochrane

doi: 10.53846/goediss-9841

et al., 2022; Gershman & Niv, 2012; Luzardo et al., 2017; Melinscak & Bach, 2020; Pearce & Bouton, 2001; Rescorla, 1988), which make precise predictions about learning curves and in the end, help to map behavioral and neural measures to one another (Schall, 2019). However, since the design of our study involved a multi-stage learning and decision process, including practice, familiarizing with the stimuli, learning about the contingency of the face-voice pairs, and learning about the contiguity of stimulus presentations, we decided, as a first step, in favor of an approximate, data-driven approach.

3.7 Conclusion and outlook

We applied two complementary methods, one distributional and one mechanistic approach, to model RTs as a function of cross-modal associative learning. This allowed us to reveal dynamics between our experimental conditions that, simply averaged, would not have been apparent. Across experiments, we observed an overall high performance for neutral and gender-matching face-voice pairs during learning and a steeper learning curve for angry face-voice pairs. Moreover, gender-congruence effects increased slightly during learning, suggesting that the potential "surprise" of hearing a gender-mismatching voice was not overcome with learning trials. Astonishingly, in Exp.2, this gender-congruence effect was still to some degree present. Overall, performance for emotional face-voice pairs was lower than neutral, matching face-voice pairs. Replications of these effects with a different, e.g., emotion-related task, could reveal whether it is emotional congruence or task intention (focus) that caused this effect.

There exist implementations of (hierarchical) Bayesian drift-diffusion models that handle smaller trial numbers and covariates such as block or time more robustly (e.g., Bürkner, 2019; Wiecki et al., 2013) and estimate the effects on a population level, instead of averaging individual estimates over participants. However, the number of parameters per condition is (to date) not trivial to implement, requires informative priors, and is computationally expensive when including a relatively large number of participants (see, e.g., Weindel et al., 2021). Although it is not in the scope of the present thesis, investigating associative learning over time with (hierarchical) drift-diffusion models would be very valuable for the generalizability of effects. In general, future studies might implement specifically designed paradigms that include more trials per condition, e.g., by including more different face-voice pairs to increase trial number and difficulty with the advantage of slowing down learning and making changes better observable. Moreover, the delayed responding task and the no-go task led to multi-stage decisional processes, which might not be fully captured by the standard drift-diffusion model but require customized models.

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Chapter 4

Motivated attention and task relevance in the processing of crossmodally associated faces: Behavioral and electrophysiological evidence

Abstract

It has repeatedly been shown that visually presented stimuli can gain additional relevance by their association with affective stimuli. Studies have shown effects of associated affect in event-related potentials (ERP) like the Early Posterior Negativity (EPN), Late Positive Complex (LPC), and even earlier components as the P1 or N170. However, findings are mixed as to the extent associated affect requires directed attention to the emotional quality of a stimulus and which ERP components are sensitive to task instructions during retrieval. In this preregistered study (https://osf.io/ts4pb), we aimed to test cross-modal associations of vocal affect-bursts (positive, negative, neutral) to faces displaying neutral expressions in a flash-card-like learning task, in which participants studied face-voice pairs and learned to correctly assign them to each other. In the subsequent EEG test session, we applied both an implicit ('old-new') and explicit ('valence-classification') task to investigate whether the behavior at retrieval and neurophysiological activation of the affect-based associations depended on the type of motivated attention. We collected behavioral and neurophysiological data from N = 40 participants who reached the preregistered learning criterium. Results showed EPN effects of associated negative valence after learning and independent of the task. In contrast, modulations of later stages (LPC) by positive and negative associated valence were restricted to the explicit, i.e., valence-classification, task. These findings highlight the importance of the task at different processing stages and show that cross-modal affect can successfully be conditioned to faces.[1]

[1]This manuscript has been submitted for publication and is currently in revision.

doi: 10.53846/goediss-9841

4.1 Introduction

The human brain navigates the complexities of our everyday social lives very efficiently, e.g., by quickly extracting various information from other people's faces (Haxby et al., 2000). Research has repeatedly shown that what we know about a person and what is relevant for us affects how we perceive that person (e.g., Bublatzky et al., 2014; Davis et al., 2009; Heisz & Shedden, 2009; for a review, see M. J. Wieser & Brosch, 2012). This includes, but is not limited to, biographical information and relevant experiences with that person. In the laboratory, relevance is often manipulated through associations with valence-laden stimuli and actions, ranging from receiving monetary (Hammerschmidt, Kagan, et al., 2018; Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017) or social reward and punishment (Aguado et al., 2012; M. J. Wieser, Gerdes, et al., 2014) to highly aversive stimuli like loud noise bursts (Watters et al., 2018) or electric shock (Rehbein et al., 2014). It has been repeatedly shown that various types of affective stimuli impact face processing promptly (for a review, see M. J. Wieser & Brosch, 2012) and through learned associations (for a review, see Miskovic & Keil, 2012).

Although the term *attention* is not clearly defined in the literature, there is consensus that certain stimuli are preferentially processed to others, because they are physically salient, they resemble targets matching our current goals, or because we have learned their relevance through past experience. Especially experience-driven attention (B. A. Anderson et al., 2021) aims to explain phenomena like impaired performance in the presence of learned aversive distractors (e.g., Öhman et al., 2001; Vuilleumier, 2005) or self-referential cues as described in the cocktail party effect (e.g., Röer & Cowan, 2021). To date, there is more evidence for conditioning effects with threat-related stimuli. However, also appetitive cues have been shown to be associated to different types of stimuli, e.g., faces, objects, or abstract stimuli like meaningless words (e.g., Aguado et al., 2012; Blechert et al., 2016; Davis et al., 2009; Hammerschmidt, Kagan, et al., 2018; Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017; Rossi et al., 2017; Steinberg, Bröckelmann, Rehbein, et al., 2013; Ventura-Bort et al., 2016). Although recognizing and reacting to both, appetitive and aversive environmental cues appears adaptive, the necessity to detect and respond fast is higher for a threatening environment (Öhman et al., 2001). Also, avoiding predictable and unpleasant situations may be preferred over detecting potentially pleasant ones (e.g., Gottfried et al., 2002).

Neurophysiological research allows to investigate processes beyond overt behavior and has demonstrated that some acquired associations with affective stimuli elicit differential neural responses, even if a conditioned behavioral or physiological response has extinguished (Antov et

doi: 10.53846/goediss-9841

al., 2020; Apergis-Schoute et al., 2014). However, in the case of absent effects in behavioral and neural measures, it remains open whether the information was learned at all, or whether it did not show under the specific test condition.

The overarching aim of this study was to investigate how directed, experience-driven attention through task requirements impacts face perception on different levels (i.e., early/automatic vs. later/elaborate processing). More specifically, we tested whether the retrieval of valence-implicit or valence-explicit features moderates the neurophysiological and behavioral response to valence-based associations in faces. Several affect-sensitive ERPs have been related to different stages of the processing of associated faces: The P1 usually peaks around 100 ms after face onset with an occipital, bilateral positivity. It is generated from the extrastriate cortex (Hillyard & Anllo-Vento, 1998; Russo, 2003), and has been reported to be enhanced for faces associated with affect-laden or valent stimuli, e.g., monetary reward (Hammerschmidt et al., 2017), emotional expressions of the associated face (Aguado et al., 2012), and threatening stimuli, although some fear-conditioning studies reported even earlier effects (e.g., Steinberg, Bröckelmann, Rehbein, et al., 2013). More reliably than for the P1, associated and conditioned effects were reported for the N170 and subsequent components. The N170 is a face-sensitive neural marker in the form of a negative deflection peaking around 170 ms over occipito-temporal region, generated to a large extent by the fusiform face area (Gao et al., 2019). N170 effects of conditioned faces have been reported for a number of fear-conditioning studies (e.g., Bruchmann et al., 2021; Camfield et al., 2016; Schellhaas et al., 2020; Sperl et al., 2021) and for studies on associated person knowledge (Luo et al., 2016; Schindler, Bruchmann, Krasowski, et al., 2021) and conditioned facial expressions (Aguado et al., 2012). Modulations of the early posterior negativity (EPN), a relative negativity over occipito-temporal regions related to the early detection of emotional relevance and most pronounced around 200-300 ms for face stimuli, have also been reported for conditioned faces with different kinds of unconditioned stimuli (US), e.g., in conditioned fear (e.g., Bruchmann et al., 2021; Sá et al., 2018; Schellhaas et al., 2020), and verbal descriptions about a person (e.g., Luo et al., 2016; Suess et al., 2014; Xu et al., 2016) and produced by a person (M. J. Wieser, Gerdes, et al., 2014). Sustained motivated attention has been related to the late positive complex (LPC). Effects on the LPC, a centro-parietal positivity, have been reported in the context of the perception of faces associated with different context for fear-conditioning (Bruchmann et al., 2021; Panitz et al., 2015; Rehbein et al., 2018; Sá et al., 2018; Sperl et al., 2021; Wiemer et al., 2021), reward (Hammerschmidt, Kulke, et al., 2018), and person knowledge studies (Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017; Xu et al., 2016). While there are several studies on cross-modal perception that include faces and affective voices (e.g., Gelder & Vroomen, 2000;

doi: 10.53846/goediss-9841

Pell et al., 2022), to our knowledge, the processing of faces that previously have been associated with both positive and negative affect bursts has not yet been tested.

The relevance of information for a specific situational task or context has been shown to play an important role both in learning and in retrieval (Shin et al., 2020). In learning, task-relevant (and hence context-congruent) information is supposed to be integrated more easily into a pre-activated schema (Kesteren et al., 2010). However for retrieval the relationship between task-relevance and context is not as straight-forward, considering reports of generalized effects across different tasks. Similar to the processing of faces with emotional expressions (Hudson et al., 2021; Rellecke et al., 2012a; Schindler et al., 2020; Valdés-Conroy et al., 2014) and affective stimuli in general (Olofsson & Polich, 2008), task requirements likely moderate conditioned effects especially at later stages of processing (Schupp et al., 2006). That early and late effects have been reported in (not necessarily the same) fear-conditioning studies may primarily be caused by the use of intense and highly arousing stimuli. Additionally, most fear-conditioning studies have implemented valence and arousal ratings of the conditioned stimuli (CS faces), before and after the conditioning phase (e.g., Panitz et al., 2015; Rehbein et al., 2018; Sá et al., 2018; Sperl et al., 2021), which might influence the attentional processes in other tasks, i.e. during learning and retrieval. Previously target-defining features of a stimulus have been reported to automatically draw processing resources away even when not anymore task-relevant (e.g., Kyllingsbæk et al., 2014). Nevertheless, conditioned effects have also been reported for valence-unrelated tasks, e.g. in old-new categorization of faces (e.g., early effects: Hammerschmidt et al., 2017; late effects: Abdel Rahman, 2011; Baum et al., 2020; Kissler & Strehlow, 2017), and passive-viewing tasks (e.g., Xu et al., 2016).

Only a few studies systematically investigated the role of attention on the perception of faces associated with context information. Three recent studies tested the effects of feature-/ and memory-based attention with different tasks, which included a) discrimination of lines that overlayed the faces, b) the faces' gender (Bruchmann et al., 2021; Schindler et al., 2022) or age (Schindler, Bruchmann, Krasowski, et al., 2021), and c) the associated CS category. In their threat-conditioning study, Schindler et al. (2022) reported interactions between task and conditioning for the P1 and the EPN, but not for the N170 and LPC components and hence show no clear distinction between task influences on early and later processing. In contrast, associated verbal descriptions of crime-related actions in Schindler, Bruchmann, Krasowski, et al. (2021) differentially moderated early and late processing, of which the N170 was enhanced in all tasks for negatively associated faces, whereas both associated effects on the EPN and LPC were only reported for the valence-focussing condition. In these studies, associated context information was presented also during the test, either interspersed (in 33% of trials in Bruchmann et al., 2021; Schindler et al., 2022) or

doi: 10.53846/goediss-9841

at the beginning of each task (Schindler, Bruchmann, Krasowski, et al., 2021). In experimental studies, researchers often have used intense and highly aversive stimuli to maximize differentiability between conditions. While this is valid and important to demonstrate a general potential of associated context to faces, it neglects the real and true diversity of affective context, especially positive associations. Moreover, it has not been particularly well researched whether associating affective stimuli of lower intensity and whose contextual relevance was not made explicit for learning also elicit robust effects in face perception. Expressions in faces and voices naturally co-occur and construct a perception of the whole person (Freeman & Ambady, 2011b). On the one hand, emotional expressions in both modalities are situation-dependent and naturally vary within individuals. On the other hand, expressions in the face and voice share inherent social and biological relevance (Straube et al., 2011). Both factors might impact the effectiveness of using these stimuli in associative learning and thus make them a compelling research topic.

Aim of the study

To address this gap, we investigated the role of attentional focus in the retrieval of faces associated with cross-modal affect. To do so, we applied a valence-implicit (old-new) and valence-explicit (valence-classification of the associated voices) task in a delayed test session to investigate associated valence in different attention conditions while recording face-sensitive ERPs. For learning, we paired faces displaying neutral expressions with short auditory affect bursts of positive, negative, and neutral valence because they unfold emotional information rapidly and do not have the segmental structure of speech. In our newly developed internet-based learning phase, unlike in classical (Pavlovian) or instrumental learning paradigms, our participants studied the face-voice pairs to correctly assign them to each other (similar to learning with flashcards). Moreover, we did not provide further information about the task requirements of the test session to not prompt participants to pay attention to specific stimulus features.

Hypotheses

Our global hypothesis was that task requirements during the test would activate (goal-directed) memory-based attention to the associated face-voice pairs, which in turn would modulate the processing of the faces. More precisely, we expected the differential effects of task on different processing stages of valence-based associations, with early processing being less impacted than later, more elaborate processing (according to Rellecke et al., 2012a). In this sense, goal-directed attention through the task and experience-based attention through the relevance

doi: 10.53846/goediss-9841

association would produce additive effects on visually-evoked potentials.

Learning. In our online learning hub, participants could study the face-voice pairs flexibly and according to their own schedules. As a result, we expected high variability in the individual learning styles and the time needed until reaching the predefined learning criterion (95% correct in 24 subsequent test trials) and analyzed learning data only in an exploratory way.

Test: Behavioral hypotheses. We expected an effect of task difficulty with slower responses and lower accuracy for the valence-classification task (3-choice-responses) compared to the old-new-task (2-choice-responses). Furthermore, we expected an interaction effect of task $\times$ valence with larger RT and accuracy differences between the affectively and neutrally associated faces for the valence-classification task compared to the old-new task, as the latter required only superficial recognition of the faces. Regarding the valence effect in the valence-classification task, we expected higher accuracy for affectively compared to neutrally associated faces. Furthermore, faces previously associated with voices expressing elation and amusement should be rated as more likable than faces associated with neutral bursts, and analogously, bursts of negative emotions should be rated as less likable compared to neutral bursts (similar to but not as pronounced as in Suess et al., 2014).

Test: ERP hypotheses. We expected that the visually evoked potentials related to face perception would be modulated by the emotional valence of the associated voices. Additionally, we expected modulatory effects of task, and interaction effects of task $\times$ valence, especially on the mid- (EPN) and long- (LPC) latency ERPs. However, for comparability with other studies and due to inconsistent findings on the influence of goal-directed attention (via task demands), we tested the interaction (task $\times$ valence) in all of our models. We predicted larger (mean and peak) amplitudes of the **P1** for affectively (i.e., positive or negative) than neutrally associated faces and similar effects in both tasks. Although we expected valence-based modulations of the **N170** and **EPN**, we did not specify the direction of effects, as effects of associated valence have only inconsistently been reported for these components. In contrast to the P1 and N170, we expected that EPN differences between affectively and neutrally associated faces would be more pronounced in the valence-classification task. The EPN is suggested to reflect enhanced sensory encoding of valence-laden stimuli (independent of the task). However, it is unknown whether recently conditioned faces would also produce an EPN component modulated by the associated valence in a superficial task like the old-new task (e.g., Rellecke et al., 2012a for reduced emotion effects on

doi: 10.53846/goediss-9841

the EPN for facial expressions in superficial tasks). Lastly, we did not expect effects of associated valence the **LPC** due to non-reported effects in similar studies. However, in the case of effects by associated valence, we expected it to be exclusive for the valence-classification task.

4.2 Method and Materials

This study was preregistered on https://osf.io/ts4pb.

Participants

Of the 61 participants, who signed up for the study, 54 started learning, and 43 completed the EEG test session, of which our target sample size, i.e., 40 participants, matched the required number of trials (min. 30 valid EEG trials per valence condition and task after artifact rejection). Our sample consisted mainly of students (38 out of 40; 30 female, 10 male, 0 diverse; Age: 18 - 32, M = 21.62 years), reporting normal or corrected-to-normal vision (max. +/- 1 diopter), normal hearing, and no neurological or psychiatric disorders. All participants were right-handed, according to Oldfield (1971), and proficient in German. We recruited via advertisements on campus and postings on social media, the university's job portal, the website of the Institute for Psychology, and the department's recruitment database. Participants were reimbursed with a fixed amount of money for completing the online learning phase and an additional hourly rate for the test session in the laboratory or an equivalent amount of course credits.

Stimuli

Twenty-four faces were selected from the Goettingen Faces Database (Kulke et al., 2017) and presented with their natural color on a light grey background. Face stimuli were edited and combined with a transparency mask that covers the hairline, ears, and neck. In the test session, they had a visual angle of approx. 3.2 $\times$ 4.8 degrees and a resolution of 200 x 300 pixels. The mean luminance (HSV; Dal Ben, 2019) of images ranged from 0.45 to 0.48 (M = 0.47; $\chi^2(528) = 550$, p =.246). Affect bursts (happy, elated, angry, and disgusted) were selected from a validated database (Cowen et al., 2019) and supplemented with neutral vocalizations (clearing throat, yawning) from the social media platform "youtube". All silent periods at the beginning and the end of the sound files were trimmed manually and normalized to -23 LUCS (Loudness Units Full Scale) with the open-source software Audacity(R) (v.2.4.2, Audacity Team, 2021). The perceived loudness of the audio files was normalized based on an algorithm following the EBU R 128 recommendation

doi: 10.53846/goediss-9841

(https://tech.ebu.ch/docs/r/r128.pdf) on limiting the loudness of audio signals. Compared to other normalization methods (peak- and RMS normalization), this method resulted in the smallest range of estimated loudness across stimuli using the R package Soundgen (Anikin, 2019). There were two separate stimulus sets for faces. One set included the 12 CS^+ faces used in the learning and test phase, and the other set contained 12 new faces, which were used for the old-new and rating task of the test session. The assignment of the CS^+ faces to the US voices was counterbalanced and matched for gender. Every emotion category (amusement, elation, yawning, throat-clearing, disgust, and anger) of the voices entailed two stimuli in our experiment (one female and one male). Hence, every valence category entailed four stimuli (four positive, four negative, and four neutral), resulting in a total of 12 face-voice pairs included in the learning phase. There were six different versions of the learning set for the face-voice pairs. Participants were pseudo-randomized to one of the six versions to ensure a balanced distribution of stimulus-set versions.

Procedure

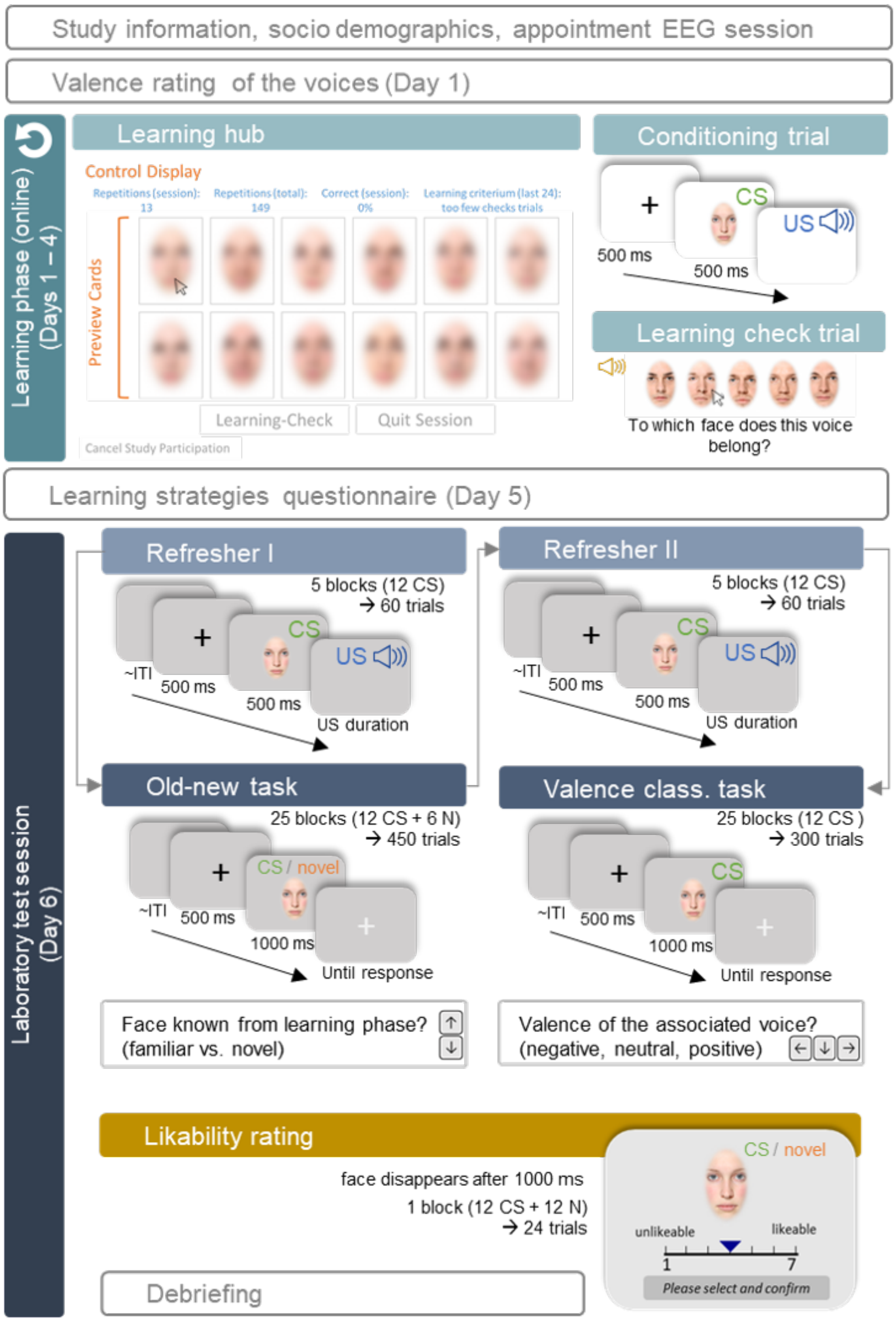
The study was approved by the local ethics committee and conducted according to the Declaration of Helsinki. Before participation, interested participants visited a website that informed them about the complete procedure, inclusion criteria, data policy, corona regulations, and remuneration of the online learning phase and EEG session. They were redirected to a form to indicate contact and socio-demographic information if they gave their consent. We set appointments for the EEG sessions with eligible participants and created personalized links and participant codes for the learning platform (learning hub). The link was activated six days before the scheduled EEG session. To participate in the EEG testing in the laboratory, participants must have achieved a learning criterion (95% correct out of 24 test trials) during one of the learning sessions and, independently of that, completed obligatory learning checks on the first four days. Participants were free to choose the length and number of learning sessions, repetitions, and learning checks within the learning phase. An overview over the learning and test procedure is shown in Figure 4.1.

Learning phase (online). The online experiment was programmed in JavaScript with self-written and existing functions from the open-source library jsPsych (v6.3.1, Leeuw, 2014). The experiment was integrated into JATOS (v3.5.8, Lange et al., 2015) on a local server installation at the University of Goettingen for data management. Participants could start the learning sessions with a personalized link and participation code. Instructions were given compulsorily at the beginning; for later sessions, they were optionally displayed. When participants logged in for the first time, they rated the valence of the auditory stimuli: On two sliders (without initial thumb),

doi: 10.53846/goediss-9841

Figure 4.1

Procedure of the online learning phase and test session.



Notes: After general information about the study procedure and the collection of socio-demographic and eligibility information, an appointment was made for the EEG session. Participants started the learning phase (online) six days before the EEG session. On the fifth day, ie the day before the EEG session, participants completed a questionnaire about their individual learning strategies. Participants were invited to the EEG session if they had reached the pre-defined learning criterion. The EEG session had the same fixed order for all participants: Refresher of the face-voice pairs, the old-new task, a second block of refresher trials, the valence classification task, and the likability rating of the faces.

doi: 10.53846/goediss-9841

they were asked to rate 1) how positive vs. negative the mood of the speaker expressing the vocalization was and 2) how pleasant vs. unpleasant they found the auditory stimulus. Following that and independent of their ratings, we individually presented all auditory stimuli with emotion labels to set an anchor for stimuli that may have been ambiguous for the participant. Subsequently, participants were redirected to the learning hub, where they could start their first learning session.

During the learning (association) phase, participants aimed to learn the pairing of the 12 face-voice stimuli (i.e., which face belonged to which voice) within four days by using the learning hub. The main page consisted of 12 preview cards showing blurred versions of the 12 faces. Clicking on one of the blurred cards started a conditioning trial of a central fixation cross, followed by the unblurred CS^+ face and the auditory US starting with face offset. The number of conditioning trials per CS-US pair was recorded for each session and in total. To assess whether they were able to allocate the face-voice pairs, participants needed to do at least one obligatory **learning check** (including 24 test trials, approx. two minutes) per day. A learning check trial consisted of a pseudo-randomly selected US voice, played while participants had to select the correct face out of five gender-matching faces. Immediate feedback on the correctness of the answer was provided. If the latest 24 learning check trials within any session were answered correctly in 95% of trials, the learning criterium had been reached. To prevent early and late learners from having different time delays between learning and test, we required daily learning checks independent of the learning criterium. On top of the learning deck, information about the number of repetitions (conditioning trials) per session and in total, as well as the accuracy in learning checks for the session and the last 24 learning checks, was displayed. The order of the preview cards was shuffled at the beginning of each new session. For learning checks, a list of all face-voice pairs was shuffled, and the number of test trials was sampled from this list without replacement. The order of faces to choose from in the learning checks was also random. Participants could cancel their participation in the online study at any time and request the deletion of their data. Once participation was canceled, it was impossible to resume or restore the data or participate in the EEG test session.

Questionnaire about learning strategies: One day before the test session, participants completed a questionnaire about their strategies to study the face-voice pairs. Participants who reached the learning criterium were informed in more detail about the procedure and the safety regulations and asked for their confirmation of the test session.

Test session. After giving written consent, participants were prepared for the EEG session and seated in a dimly lit, electrically shielded room in front of a computer screen at a distance of approx. 78 cm. Two speakers were located on the left and right at the monitor's height. Participants

doi: 10.53846/goediss-9841

positioned their chins in a height-adjustable chin rest. For the presentation of the laboratory experiment, along with standard Python (2.7) libraries like NumPy and SciPy we used functions of PsychoPy (Peirce, 2009) for the presentation of the faces, PyGame (Shinners, 2011) as the audio library, and PyGaze (Dalmaijer et al., 2013) for the communication with the eye-tracker. After a welcoming message, the eye tracker was calibrated with a 9-point calibration. For all participants, the test session had the same order: Refresher trials I (5 $\times$ 12 trials), Old-new task (25 $\times$ 18 trials), refresher trials II (5 $\times$ 12 trials), valence-classification task (25 $\times$ 12 trials), and a likability rating (24 trials). To not reveal that any emotion- or valence-relevant task would be part of the experiment, the specific task instructions were shown before each task. Refresher trials were passive-viewing trials in which participants did not need to respond. However, we instructed them to focus on the face-voice pairs to "refresh" what they had learned. For the other tasks, specific instructions were followed by four example trials using not real faces and the correct answer as a label on top to familiarize with the response keys. Only here participants got feedback on whether they were correct and had the possibility to clarify the remaining questions with the experimenter. Breaks to stretch and relax were scheduled between tasks, and additionally, there were four breaks within the old-new task and three within the valence-classification task. A drift correction of the eye-tracker (1-point-calibration) was implemented to resume or start the next task.

In all tasks, the order of faces or face-sound combinations was shuffled at the beginning of each block, with each block consisting of a single set of associated faces (or all 12 associated faces plus six randomly selected faces from the set with the novel faces). Assignments of response keys were alternated between participants. We instructed participants to answer as accurately and fast as possible and guess if unsure. **Refresher trials** started with a black fixation cross at the center of the screen for 500 ms, which was replaced by one of the CS^+ faces displayed for 500 ms. With the offset of the face, the US set in (duration varied between stimuli). After a jittered inter-trial interval (M = 2800 ms, SD = 200 ms), the subsequent trial started. In the **old/new-task**, the participant's task was to decide if a face was known from the online learning phase or of a novel set of faces. A black fixation cross was displayed for 500 ms at the center of the screen replaced by either a CS^+ face from the learning phase or a CS^- face (novel). All faces were presented individually for 1000 ms and participants could respond as soon as the face stimulus set in. With the offset of the face, a grey fixation cross was displayed if no answer had been registered yet and continued until an answer was given via keypress. After the face-offset and a registered response, the next trial started after an additional jittered inter-trial interval (M = 1800 ms, SD = 200 ms). In the **valence-classification task**, participants had to recall the valence category (negative, positive, neutral) of the associated voices, while seeing only the CS^+ faces. The presentation duration of trial elements

doi: 10.53846/goediss-9841

(fixation cross, face, response fixation cross, inter-trial interval) was identical to the old-new task. At the end of the test session, participants rated the **likability** of the CS^+ and the novel CS^- faces. Faces were presented individually for 1000 ms and rated on a Likert scale in the appearance of a 1-7 slider positioned below the faces. There was no value or choice shown by default. A value was selected by clicking with the mouse on the slider but had to be confirmed to start the next trial. At the very end of the session, participants were informed about the main aims and background of the study (presented on the computer screen) and could clarify questions with the experimenter.

Collected data

For every learning session, the number of conditioning trials for each individual face-voice pair and the accuracy of learning checks trials was recorded. In the test session, in addition to performance (RT and accuracy), we recorded EEG and pupil during the refresher, old-new, and valence-classification tasks. No neurophysiological measures were collected during the likability rating.

EEG recording and preprocessing

The continuous EEG was recorded with a sampling rate of 512 Hz (bandwidth: 102.4 Hz) at 64 active electrodes (AgAgCl) mounted in an electrode cap (Easy CapTM). The arrangement was based on the extended 10-20 system (Pivik et al., 1993). Additionally, two external electrodes were used, one each for the left and right mastoids. Reference electrodes were the common mode sense (CMS) active electrode and as ground electrode, the driven right leg (DLR) passive electrode. The scalp voltage signals were amplified by a BiosemiActiveTwo AD-Box and recorded with the software ActiView. The data was preprocessed offline in MATLAB (2018) with functions of the toolbox EEGLAB (2019.9, Delorme & Makeig, 2004). To account for a systematic delay that was measured with a photodiode, event markers were shifted by a constant of 26 ms. The continuous data was re-referenced to average reference (excl. external electrodes), filtered with a 0.01 Hz second-order Butterworth filter, and the remaining 50 Hz line-noise was corrected with a function of the plugin "CleanLine" (v1.04, Mullen, 2012). Prior to performing independent component analysis (ICA), data was epoched from -500 ms to 1000 ms around face-onset and the mean of the pre-stimulus baseline (-500 ms to 0 ms) was subtracted. Extended Infomax ICA was performed after a PCA reduction to 63 channels on a 1 Hz high-pass filtered copy of this dataset. The resulting ICA weights were transferred to the original 0.01 Hz filtered dataset. Independent components (ICs) were removed if labeled as muscle (>80%), eye (>90%), and channel-noise

doi: 10.53846/goediss-9841

(>90%) components using "IClabel" (v1.2.4, Pion-Tonachini et al., 2017). Remaining diverging channels (>3 *SD*) were spherically interpolated. Then, epochs were trimmed to -200 - 800 ms and baseline-corrected (-200 ms to 0 ms). Trial-wise artifact rejection was performed: amplitudes exceeding -100/100 μV (M = 42.15; 4.84%), steep amplitude changes (> 100 μV within an epoch; M = 3.80; 0.44%), improbable activation (>3 *SD* of the mean distribution for every time point; M = 108.33; 12.4%) were excluded. Overall the mean rejection rate was 15.29%. Eye blinks during baseline or face presentation were excluded in a separate step using pupil data. We extracted the following ERPs based on time windows and regions of interest (ROI) electrodes of a previous study (Ziereis & Schacht, submitted): Mean and peak amplitudes for the **P1** (80 - 120 ms) at an occipital electrode cluster (O1, O2, and Oz); mean (and peak)[2] amplitudes for the **N170** (130 - 200 ms) at an occipitotemporal electrode cluster (P10, P9, PO8, PO7); mean amplitudes for the **EPN** (250-300 ms) at an occipitotemporal cluster (O1, O2, P9, P10, PO7, and PO8); mean amplitudes for the **LPC** (400 - 600 ms) at an occipito-parietal electrode cluster (Pz, POz, PO3, and PO4). Moreover, we investigated ERP effects between familiar and novel identities in the old-new task, we included two further ERPs, which we included in the Appendix C, together with details on pupil-related results of this study.

Statistical analysis

Tables with statistical models (incl. estimates, confidence intervals, stability measures, and likelihood ratio tests) are in Appendix A. All statistical analysis was conducted in R (v 4.0, R Core Team, 2020). All statistical models but the beta inflated distribution model (see below) were sum-contrast-coded, reflecting main effects rather than marginal effects. Here, the intercept corresponds to the (unweighted) grand mean, and lower-level effects are estimated at the level of the grand mean. The significance of the predictors was tested with likelihood ratio tests (LRT) of models including the predictor against reduced models and a null model. Post-hoc contrasts were used to test the difference between factor levels using "emmeans" (Lenth, 2020). We used the conventional significance level α = 0.05 (two-sided) and for posthoc tests Šidák-correction to adjust for multiple comparisons. To estimate the parameters in the analyses, we used the maximum likelihood (ML) estimator. For the 95% confidence intervals we used non-parametric bootstrapping (nsim = 999) if not specified otherwise.

Ratings of the voices. Due to the nature of the slider response measure with lower and upper bounds, we used a beta inflated distribution model (GAMLSS family "BEINF," Stasinopoulos

[2]not preregistered

doi: 10.53846/goediss-9841

& Rigby, 2007) to estimate the ratings of the voice stimuli. The model allows zero and one as values for the response variable. The beta inflated distribution is given as $f(y) = p_0$, if (y = 0), $f(y) = p_1$, if (y = 1), and $f(y|\alpha, \beta) = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}$ for y $\in$ (0,1).

The full model included the predictors emotion of the voice stimulus, type of the rating (valence of the voice vs. personal reaction to the voice), their interaction, and a random intercept for participant ID.

Learning phase (learning speed). We modelled accuracy of the learning check trials until the learning criterium was reached (for the first time) with a binomial mixed model (GLMM). Predictors of the binomial mixed model were valence, number of learning checks (per valence), and their interaction. We included random slopes of valence and check number and the random intercept participant ID.

Test session. For response time data, only correct trials were selected. Separately for every participant, task and condition, data was trimmed to a maximum cutoff of 5000 ms post face onset and a skewness adjusted boxplot method to exclude extreme values (function “adjbox” of the package “robustbase,” Maechler et al., 2021; based on Hubert & Vandervieren, 2008). After averaging across participant and conditions, response time data still resulted in skewed residuals. By taking the natural log of the averaged response times, the distribution of residuals became less skewed. We report all model parameters on the log scale. Our model included valence, task, the valence $\times$ task interaction as fixed effects, valence and task as random slopes and participant ID as random intercept. The model allowed random slopes and the random intercept to be correlated. Additionally, for the old-new task, we tested response time differences between familiar and novel faces in a separate model by adding the level (“novel”) to the predictor variable valence.

We ran mixed logistic regression models (binomial GLMMs) on accuracy data. The predictor variables (UVs) were task (old-new and valence-classification), valence of the associated sounds (negative, neutral, positive) and their interaction. Because we detected overdispersion in the preregistered model (which included only participant ID as a random intercept), we maximized the random effects structure for the model including valence with a random intercept of participant ID, a random slope for valence and a random slope for task. Including a slope for the interaction between valence and task resulted in singularity issues and was dropped from the model.

For the likability rating, we ran ordinal-mixed-models with valence (incl. novel) or emotion as fixed effect and random intercepts for face and participant ID. Model estimates and their 95% confidence intervals are reported as Odds ratios.

doi: 10.53846/goediss-9841

For the ERP analysis only correctly answered trials were included. The study had a 2(task: old-new / valence-classification) × 3 (valence: negative/positive/neutral) within-subject design. For all outcomes (P1, N170, EPN and LPC amplitudes), mixed models with the fixed effects valence (positive, negative, neutral), task (old-new and valence-classification), their interaction, and the random effect (intercept) participant ID were analyzed. In addition to these ERPs, for the old-new task, we analyzed the FN400 and LPON in a separate models and added the level ("novel") to the predictor variable valence (see Appendix C). Furthermore, we explored auditory processing of the refresher trials in voice-locked N1-P2 ERP complex with N1 (90 - 145 ms) and P2 (165 - 300 ms), both with the identical fronto-central electrode cluster: F3, F1, Fz, F2, F4, FC1, FC3, FC2, FC4, C3, C1, Cz, C2, C4, CP1, CP3, CPz, CP2, and CP4. Although we expected the associated effects to reflect valence rather than the individual emotion categories, we analyzed all ERPs with the fixed effects task and emotion (6 levels) and their interaction.

4.3 Results

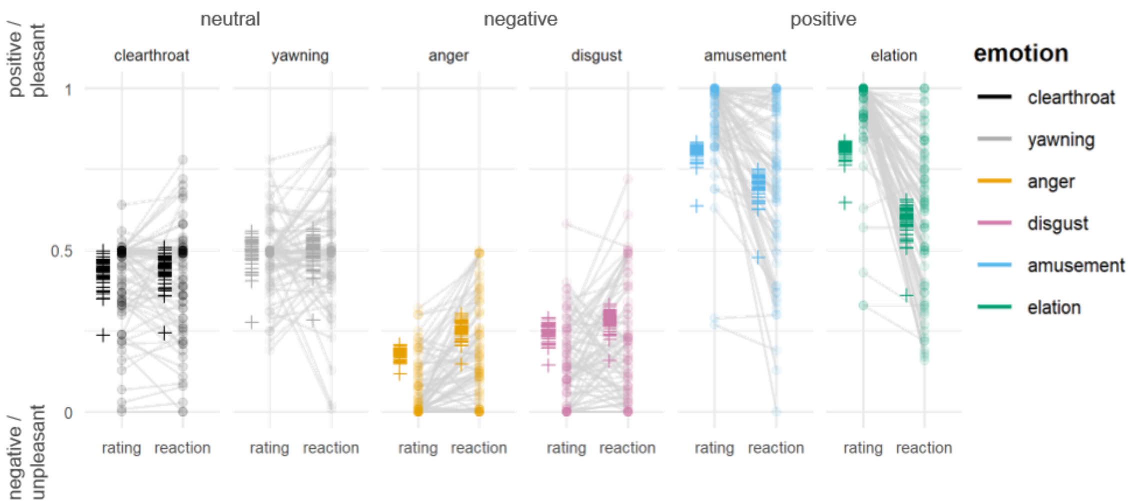
Learning session

Valence rating of the voices. Before the first learning session, participants evaluated the individual voices along the dimensions "valence of the speaker's expression" and "reaction to the burst". The zero-one-inflated model showed a main effect of emotion ($\chi^2(0.7) = 6, p = .008$), rating type (valence rating vs. reaction: $\chi^2(23.1) = 828.21, p <.001$) and a significant emotion × rating type interaction ($\chi^2(6.34) = 137.78, p <.001$). In line with the pre-specified valence categories, participants rated the bursts as negative, neutral and positive vocal expressions. However, their overall personal reaction towards the stimuli was more homogeneous across emotion categories, with less positive reactions to elation ($\beta_{\text{elation\_reaction}} = -1.48$, $CI = [-1.85; -1.12]$) and amusement ($\beta_{\text{amusement\_reaction}} = -0.84$, $CI = [-1.2; -0.48]$) and less negative reactions to anger ($\beta_{\text{anger\_reaction}} = 0.67$, $CI = [0.3; 1.04]$), see Figure 4.2.

Repetitions of face-voice pairs. The number of repetitions of face-voice pairs varied across participants, with a total number of repetitions ranging from 18 to 428 and, for a given valence group, from 4 to 7 to 125 to 157. When proportions were considered, positive face-voice pairs were repeated least frequently (median = 32%) but had the largest range (25% - 47%), followed by neutral (33%; 22% - 41%) and negative (35%; 26% - 45%) face-voice pairs.

doi: 10.53846/goediss-9841

Figure 4.2
Rating of the vocal bursts (pre-learning).



Notes: Stimuli were rated regarding the speakers' emotional valence (positive/negative) and the personal reaction towards the stimuli (unpleasant/pleasant) on separate sliders with no initial thumb. Dots represent the raw rating data per stimulus and participant. Crosses represent the predicted values per participant based on the zero-one-inflated GAMLSS model.

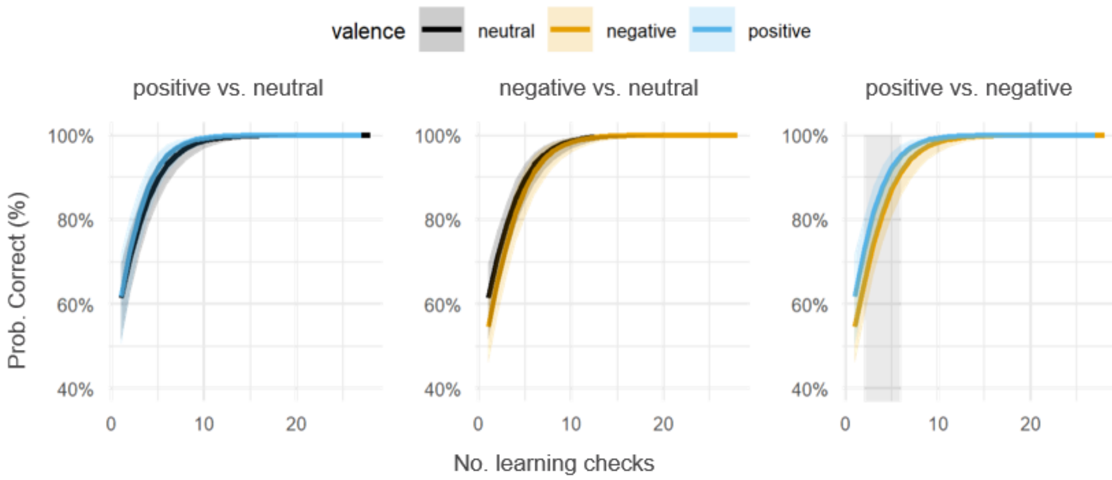
Learning speed (accuracy) by valence. There was a main effect of number of learning checks ($\chi^2(1) = 51.09, p < .001$), no effect of valence ($\chi^2(2) = 1.9, p = .387$), but a valence × check number interaction ($\chi^2(2) = 7.95, p = .019$). Until the learning criterium was met, there were differences in learning speed between valence categories. Positive face-voice pairs were significantly learned faster compared to negative face-voice pairs at early check trials (predicted accuracies were outside 95% point-wise CI of the other valence category between the second and sixth learning checks per valence category). Differences between other valence categories over time were not significant (see Figure 4.3).

Learning strategies. Except for the mandatory learning checks, the learning phase could be organized flexibly by participants. To gain more knowledge about how they experienced learning (i.e., perceived difficulty and subjective learning styles) we asked all participants to complete an online questionnaire the day before the lab session. Overall, participants varied in how difficult they rated the learning task. On a Likert scale from 1 (very hard) to 5 (very easy), studying the face-voice pairs and reaching the learning criterium were rated on average as rather easy (M = 3.76, SD = 0.85). All participants indicated certain face-voice pairs to be harder to memorize. However, participants differed in what they specified as difficult: high similarity between faces (n = 23), lack of distinctive facial features (n = 16), gender of the face (female faces easier (n = 3), male faces easier (n = 5)), emotion (neutral more difficult (n = 4), anger and disgust within

doi: 10.53846/goediss-9841

Figure 4.3

Predicted accuracy for learning checks.



Notes: The x-axis refers to the number of learning checks separately per valence category, the y-axis to the predicted probability of a correct answer in a learning check. The light-grey highlighted area reflects significant differences between valence categories (predicted accuracy of one curve outside of 95% *CI* of the other).

male faces more difficult (n = 1)), and subjective mismatch between faces and voices (n = 4). The majority of participants (n = 33) indicated that they used at least one specific strategy to study the face voice pairs, of which mnemonic device (n = 28) was mentioned most often, followed by focus on specific distinct facial features (n = 26) in order to be able to distinguish faces. Less frequently, they reported to form sub-groups of stimuli (e.g., female pairs first) and learned them separately (n = 5). Most participants began by using the card deck (n =25). However, after a while some (n = 6) preferred to use mainly the learning checks to look at the faces for a longer duration and to get feedback on which faces still needed practice. Only two participants used spatial information of the preview cards at the beginning but stopped because the positions in each session were shuffled.

Participants rated their everyday ability to memorize faces on a 5-point Likert scale from 1 (very hard) to 5 (very easy) as rather high (M = 3.92, SD = 1.1). This self-reported ability did not significantly correlate with the number of learning checks needed until the learning criterium was met ($r(36) = -0.24$, p = .814).

doi: 10.53846/goediss-9841

Table 4.1

Means and standard errors for accuracy, RTs, and ERP amplitudes of the test session for valence and emotion by task

		neutral			positive			negative			novel
Task	Measure	avr.neu	yawning	clearthroat	avr.pos	elation	amusement	avr.neg	disgust	anger	novel
old-new	Acc.Test	98.6 (4.04)	99.1 (1)	98.1 (7.69)	98.83 (3.23)	99.45 (1)	98.2 (6.19)	98.35 (3.17)	98.25 (4.53)	98.45 (3.72)	98.45 (1.76)
	RT	725 (136)	726 (149)	723 (121)	722 (139)	721 (137)	723 (146)	733 (144)	743 (164)	724 (133)	756 (116)
	P1	4.05 (4.22)	4.02 (4.32)	4.04 (4.25)	4.08 (4.08)	3.98 (4.06)	4.17 (4.25)	4.08 (4.05)	4.16 (3.96)	3.93 (4.17)	
	P1.peak	7.04 (4.15)	7.2 (4.36)	7.51 (4.16)	6.93 (3.8)	7.06 (3.9)	7.4 (3.94)	6.88 (3.89)	7.32 (3.8)	7.1 (4.05)	
	N170	-6.73 (3.78)	-6.87 (4.01)	-6.55 (3.74)	-6.61 (3.9)	-6.75 (3.82)	-6.49 (4.04)	-6.74 (3.85)	-6.51 (3.91)	-7.05 (3.86)	
	N170.peak	-10.73 (4.57)	-11.14 (4.69)	-10.97 (4.55)	-10.7 (4.68)	-10.97 (4.62)	-10.96 (4.8)	-10.88 (4.56)	-10.93 (4.6)	-11 (4.54)	
	EPN	-2.08 (3.3)	-2.21 (3.48)	-1.91 (3.31)	-2.19 (3.25)	-2.38 (3.11)	-2.03 (3.5)	-2.51 (2.97)	-2.33 (3.23)	-2.73 (2.89)	
	LPC	5.32 (3.07)	5.25 (3.04)	5.36 (3.25)	5.25 (3.45)	5.21 (3.52)	5.28 (3.51)	5.24 (3.19)	5.39 (3.6)	5.1 (2.91)	
valence-class	Acc.Test	95.38 (9.5)	96.2 (9.11)	94.55 (11.84)	97.08 (4.58)	97.05 (6.23)	97.1 (5.51)	96.7 (6.09)	97 (8.66)	96.4 (8.46)	
	RT	957 (148)	943 (152)	974 (167)	930 (143)	919 (122)	944 (187)	934 (178)	928 (182)	945 (202)	
	P1	4.08 (4.41)	3.97 (4.43)	4.14 (4.46)	3.96 (4.18)	3.98 (3.92)	4 (4.6)	3.98 (4.01)	4.02 (3.96)	3.91 (4.2)	
	P1.peak	6.81 (4.26)	7.07 (4.24)	7.34 (4.26)	6.91 (3.99)	7.2 (3.67)	7.2 (4.46)	6.86 (4.08)	7.12 (4.02)	7.05 (4.35)	
	N170	-7.04 (3.7)	-7.14 (3.76)	-6.99 (3.75)	-7.15 (3.76)	-7.2 (3.79)	-7.12 (3.82)	-7.28 (3.8)	-7.21 (3.78)	-7.38 (3.84)	
	N170.peak	-10.71 (4.4)	-11.2 (4.51)	-10.9 (4.41)	-10.86 (4.46)	-11.18 (4.51)	-11.15 (4.51)	-10.86 (4.47)	-11.08 (4.48)	-11 (4.53)	
	EPN	-2.06 (3.21)	-2.1 (3.36)	-2.03 (3.3)	-2.16 (3.45)	-2.2 (3.75)	-2.12 (3.36)	-2.58 (3.1)	-2.52 (3.34)	-2.74 (3.09)	
	LPC	5.06 (3.42)	5.06 (3.38)	5.16 (3.57)	5.61 (3.8)	5.7 (3.96)	5.55 (3.78)	5.55 (3.46)	5.63 (3.56)	5.51 (3.54)	

Notes: Accuracy is in %, Response times are in ms, and ERP amplitudes in μV

doi: 10.53846/goediss-9841

Test session

Response times. *Associated valence:* In line with our hypothesis, responses were slower in the valence-classification task than the old-new task ($diff_{valclass\text{-}oldnew}$ = 212 ms; $\chi^2(1) = 56.42$, $p < .001$). There was no main effect of valence ($\chi^2(2) = 2.14$, $p = .343$), but an interaction between valence and task ($\chi^2(2) = 6.55$, $p = .038$). In the valence-classification task, neutral trials were descriptively slower than positive (and to a lesser extend also negative) trials but post-hoc differences were not significant (all adjusted p >.05). *Associated emotion:* Due to singularity issues we reduced the model structure to a random intercept model with participant ID. Also in this model there was a main effect of task ($diff_{valclass\text{-}oldnew}$ = 213 ms; $\chi^2(1) = 406.07$, $p < .001$). However, neither emotion ($\chi^2(5) = 2.68$, $p = .749$), nor the interaction between emotion and task was significant ($\chi^2(5) = 4.64$, $p = .461$). *Old/new task comparison:* When comparing the valence categories and novel stimuli in the old new task, participants responded more slowly to novel faces compared to faces known from the learning phase, irrespective of their valence ($\chi^2(3) = 19.67$, $p < .001$). The largest difference was between positive and novel faces ($diff_{pos\text{-}nov}$ = -37 ms, $p < .001$).

Accuracy. *Associated valence:* A hypothesized, the valence-classification task had significantly more errors compared to the old-new task ($OR_{valclass/oldnew} = 0.32$, $\chi^2(1) = 13.85$, $p < .001$). Valence was not significant ($\chi^2(2) = 0.13$, $p = .938$) and there was also no significant interaction between task and valence ($\chi^2(2) = 0.56$, $p = .756$). *Associated emotion:* A model with single emotion levels resulted again in a main effect of task ($\chi^2(1) = 14.16$, $p < .001$). There was no main effect of emotion ($\chi^2(5) = 2.74$, $p = .740$) but a significant interaction between task and emotion ($\chi^2(5) = 12.72$, $p = .026$). Post-hoc tests showed that for all emotion categories but anger ($OR = 0.61$ $p = .198$) and elation ($OR = 0.48$ $p = .051$) the valence-classification task had a significant lower accuracy compared to the old-new task (all $p <= .05$).

ERP results.

P1. *Associated valence:* P1 mean amplitudes were neither modulated by task ($\chi^2(1) = 0.46$, $p = .497$) nor valence ($\chi^2(2) = 0.14$, $p = .931$) nor their interaction ($\chi^2(2) = 0.54$, $p = .764$) Similarly, P1 peak amplitudes were neither modulated by task ($\chi^2(1) = 0.67$, $p = .413$) nor valence ($\chi^2(2) = 0.22$, $p = .896$) nor their interaction ($\chi^2(2) = 0.91$, $p = .635$). *Associated emotion:* Replacing valence with the single emotion categories did not change results of P1 mean amplitudes (task: $\chi^2(1) = 0.26$, $p = .607$; emotion: $\chi^2(5) = 2.07$, $p = .839$; task $\times$ emotion: $\chi^2(5) = 0.96$, $p = .965$).

doi: 10.53846/goediss-9841

Similarly, a model including emotion did not explain P1 peak amplitudes significantly (task :$\chi^2(1) = 0.99$, $p = .319$; emotion: $\chi^2(5) = 5.33$, $p = .377$; task $\times$ emotion: $\chi^2(5) = 1.41$, $p = .923$).

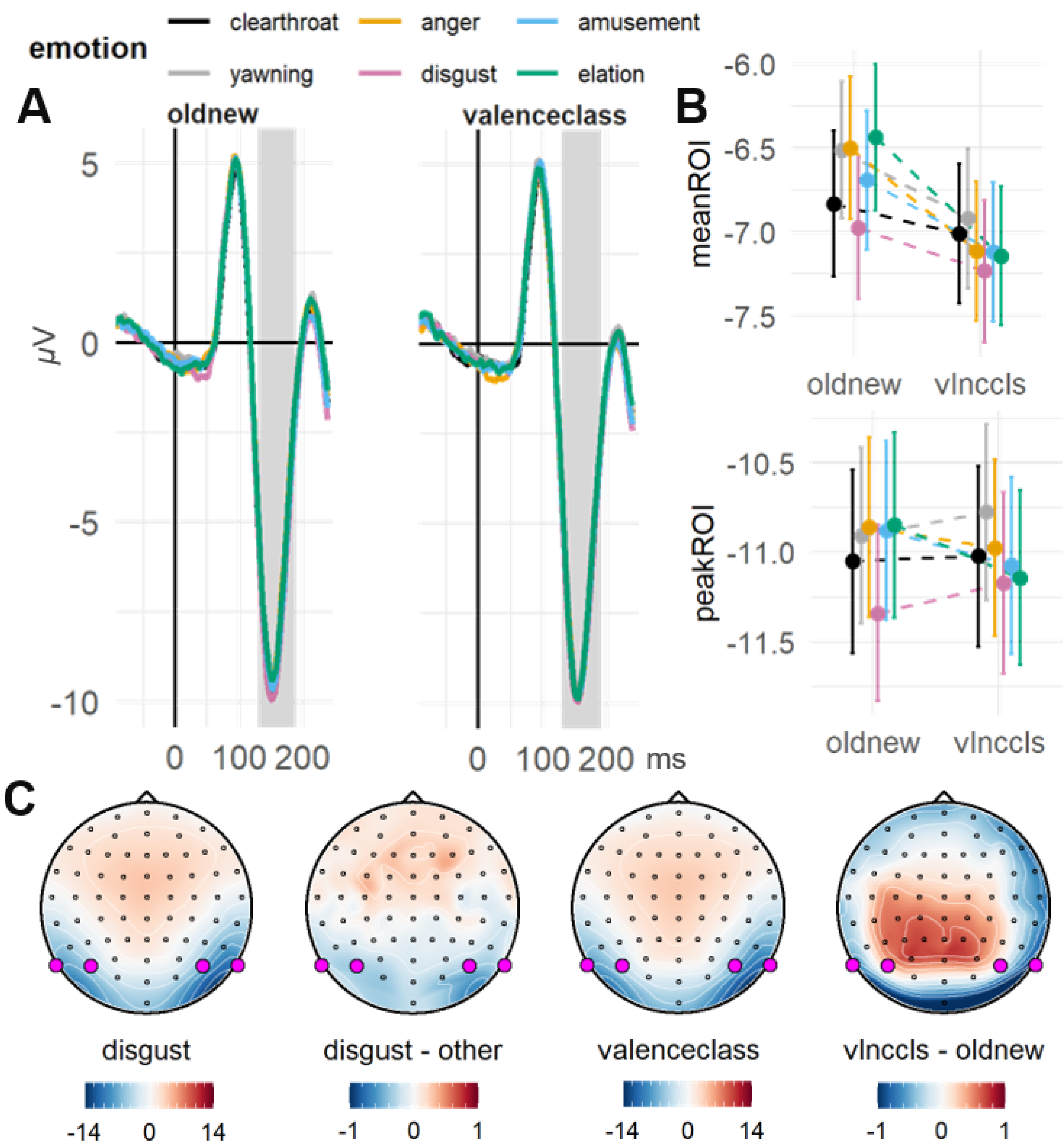
N170. *Associated valence:* N170 mean amplitudes were not modulated by valence ($\chi^2(2) = 2.1$, $p = .350$), but there was a main effect of task ($\chi^2(1) = 28.72$, $p <.001$). Mean amplitudes averaged across valence conditions were significantly more negative in the valence-classification task (-7.16 μV; $\beta_{\text{valclass}} = -0.23$, $SE = 0.04$, $t = -5.49$) compared to the old-new task (-6.69 μV). No interaction between valence and task ($\chi^2(2) = 1.69$, $p = .430$) was present. N170 peak amplitudes were not modulated by valence ($\chi^2(2) = 1.65$, $p = .437$), task ($\chi^2(1) = 0.21$, $p = .648$) or the valence $\times$ task interaction ($\chi^2(2) = 0.77$, $p = .681$). *Associated emotion:* Looking at emotion categories separately, N170 mean amplitudes were significantly modulated by emotion ($\chi^2(5) = 13.67$, $p = .018$), with disgust showing an enhanced negative mean amplitude (-7.21 μV; $\beta_{\text{dis}} = -0.28$, $SE = 0.09$, $t = -3.04$). Also, in this model, a main effect of task was present ($\chi^2(1) = 32.96$, $p <.001$) with more negative mean amplitudes for the valence-classification task (-7.17 μV; $\beta_{\text{valclass}} = -0.23$, $SE = 0.04$, $t = -5.78$). However, there was no interaction between task and emotion present ($\chi^2(5) = 3.58$, $p = .612$). Similar to mean amplitudes, emotion significantly modulated peak amplitudes ($\chi^2(5) = 11.58$, $p = .041$) with enhanced peak amplitudes for disgust (-11.42 μV; $\beta_{\text{dis}} = -0.31$, $SE = 0.1$, $t = -3.01$). There was no effect of task ($\chi^2(1) = 0.76$, $p = .383$) and no interaction between emotion and task ($\chi^2(5) = 1.66$, $p = .894$).

EPN. *Associated valence:* There was a main effect of valence on EPN amplitudes ($\chi^2(2) = 10.86$, $p = .004$). This was due to enhanced negative amplitudes for negatively (-2.54 μV; $\beta_{\text{neg}} = -0.28$, $SE = 0.09$, $t = -3.23$) compared to neutrally ($\text{diff}_{\text{neu-neg}} = 0.47$, $p = .006$) and positively ($\text{diff}_{\text{pos-neg}} = 0.37$, $p = .045$) associated faces. There was no no main effect of task on EPN amplitudes ($\chi^2(1) = 0.01$, $p = .925$) and no interaction between valence and task ($\chi^2(2) = 0.13$, $p = .936$). *Associated emotion:* Looking at emotion categories separately, there was a main effect on EPN amplitudes by emotion ($\chi^2(5) = 21.61$, $p = <.001$) due to enhanced negative amplitudes for disgust (-2.73 μV; $\beta_{\text{dis}} = -0.46$, $SE = 0.12$, $t = -3.79$) compared to the neutral categories throatclearing ($\text{diff}_{\text{dis-clt}} = -0.58$, $p = .031$) and yawning ($\text{diff}_{\text{dis-yaw}} = -0.76$, $p = <.001$) and compared to the positive category elation ($\text{diff}_{\text{dis-el}} = -0.66$, $p = .008$), collapsed across tasks. Also in this model, task did not modulate EPN amplitudes ($\chi^2(1) = 0.04$, $p = .838$) and the emotion $\times$ task interaction ($\chi^2(5) = 1.47$, $p = .917$) was not significant.

doi: 10.53846/goediss-9841

Figure 4.4

Face-locked N170 by emotion.

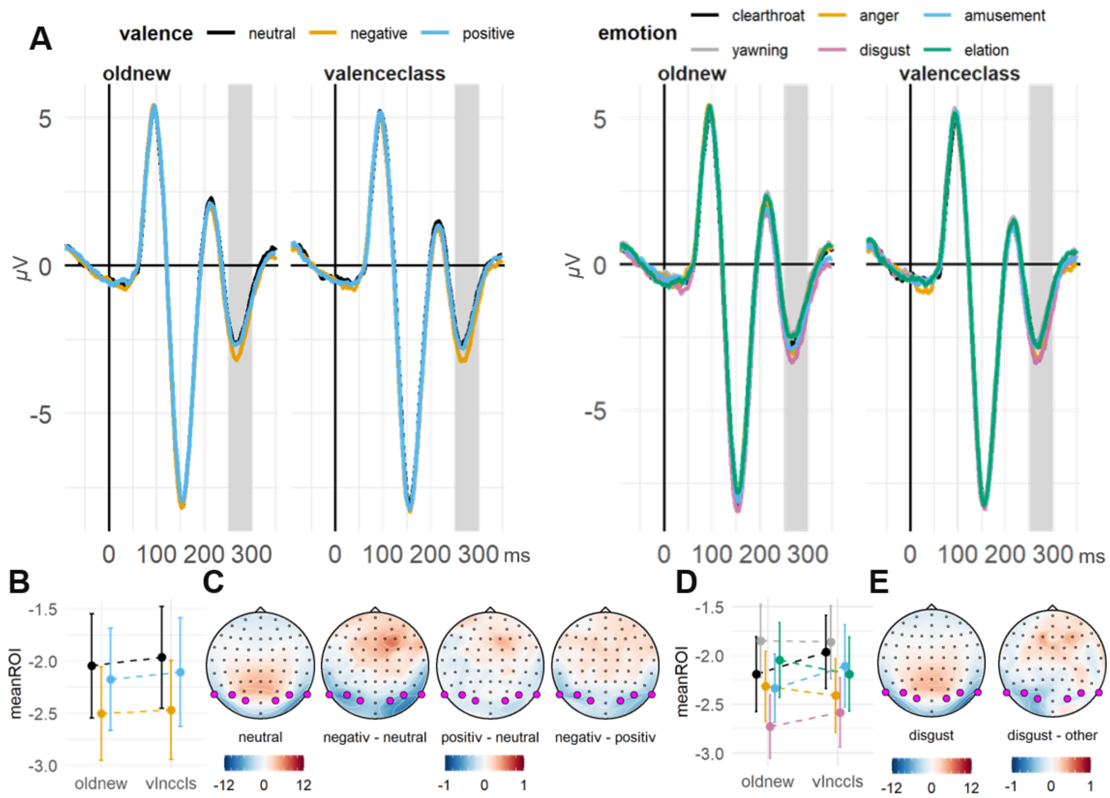


Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand-averages of the ROI mean amplitudes (left panel) and peak amplitudes (right panel), contrasted for the implicit and explicit task and all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **C** Topographies of the ERP distribution for faces associated with disgust bursts contrasted with all other emotion conditions, averaged across the implicit and explicit tasks. ROI channels are highlighted in pink.

doi: 10.53846/goediss-9841

Figure 4.5

Face-locked EPN by valence and emotion.



Notes: **A** Grand average ERP time series of the averaged ROI channels for valence (left panel) and emotion (right panel). The highlighted area displays the ROI time window. **B** (valence) and **D** (emotion): Grand-averages of the ROI mean amplitudes, contrasted for the implicit and explicit task and all valence/emotion conditions. Errorbars indicate +/- 1 SE of the mean. **C** Topographies of the ERP distribution for faces between valence conditions, averaged across the implicit and explicit tasks. **E** Topographies of the ERP distribution for faces associated with disgust bursts contrasted with all other emotion conditions, averaged across the implicit and explicit tasks. ROI channels are highlighted in pink.

doi: 10.53846/goediss-9841

LPC. *Associated valence:* LPC amplitudes did not show a modulation by valence ($\chi^2(2) = 2.64$, $p = .268$) or task ($\chi^2(1) = 1.08$, $p = .298$). Although, the valence × task interaction was not significant, ($\chi^2(2) = 4.46$, $p = .108$), looking at the time course of the component, affectively compared to neutrally associated faces appeared to show a different activation in the valence classification task. Post-hoc tests showed that there was a trending positive difference between positive and neutral ($\text{diff}_{\text{pos-neu}} = 0.55$, $p = .054$), and there was a trending positive difference between negative and neutral categories ($\text{diff}_{\text{neg-neu}} = 0.49$, $p = .098$) which was only present in the valence classification task. *Associated emotion:* LPC amplitudes were not significantly explained by single emotion levels ($\chi^2(5) = 5.01$, $p = .414$), or task ($\chi^2(1) = 2.46$, $p = .117$) or an emotion × task interaction ($\chi^2(5) = 6.2$, $p = .287$). Descriptively, in the valence classification task, the neutral categories resulted in lower amplitudes, but also here, none of the post-hoc contrasts was significant.

Likability rating. We ran two (one for valence and one for emotion levels) cumulative linked mixed models to account for the ordinal scale of the likability ratings. In both models, random intercepts for participant and face-stimulus were included. Likelihood ratio tests of both models and a model without a fixed effect showed that valence significantly explained the variance of the rating data ($\chi^2(3) = 96.19$, $p = <.001$). However, separating emotion categories did not explain the data better compared to the valence categories ($\chi^2(3) = 0.79$, $p = .851$). The odds ratios and 95% CI of both models are in Table C20 of Appendix C. Mean ratings and model predictions are shown in Figure 4.7. *Associated valence:* Associated valence modulated the likability ratings in line with our hypothesis, with positively associated faces being rated as most likable compared to negatively associated faces as the least likable. With the exception of novel and negatively associated faces, all pairwise differences between valence categories were significant (all adjusted $p <.01$). *Associated emotion:* The ordinal model including emotion categories showed a grouping of levels according to the pre-specified valence categories. There was no significant differences within valence categories (neutral: throat-clearing and yawning; negative: anger and disgust; positive: amusement and elation). Pairwise comparisons of emotion categories between valence levels were significant (all adjusted $p <.05$), with the exception of throat-clearing and all positive categories, yawning and amusement, and novel and all negative categories.

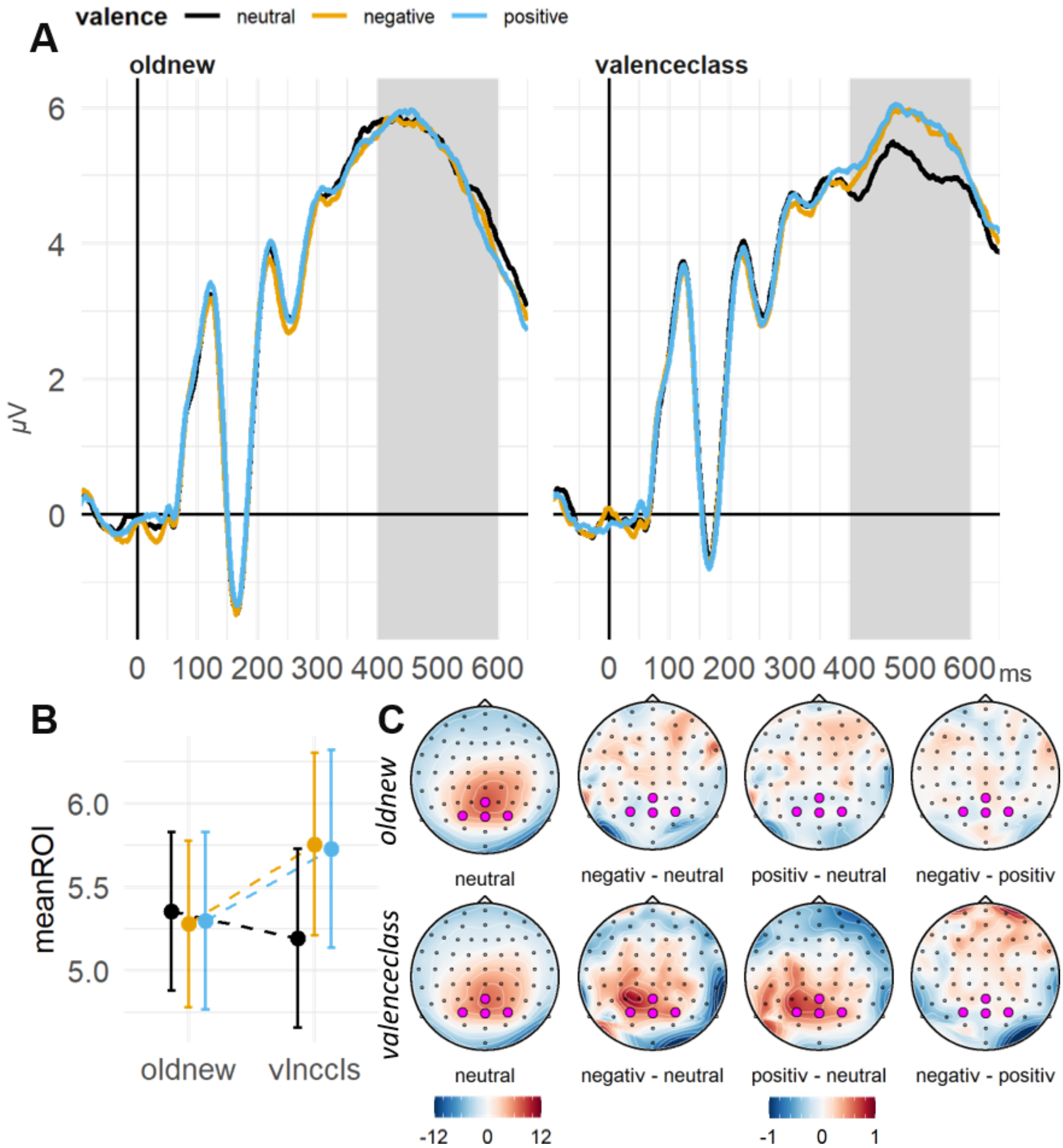
4.4 Discussion

The present study aimed to investigate memory-based attention effects on the retrieval of valence-based associations in face perception. After having faces associated with affective bursts

doi: 10.53846/goediss-9841

Figure 4.6

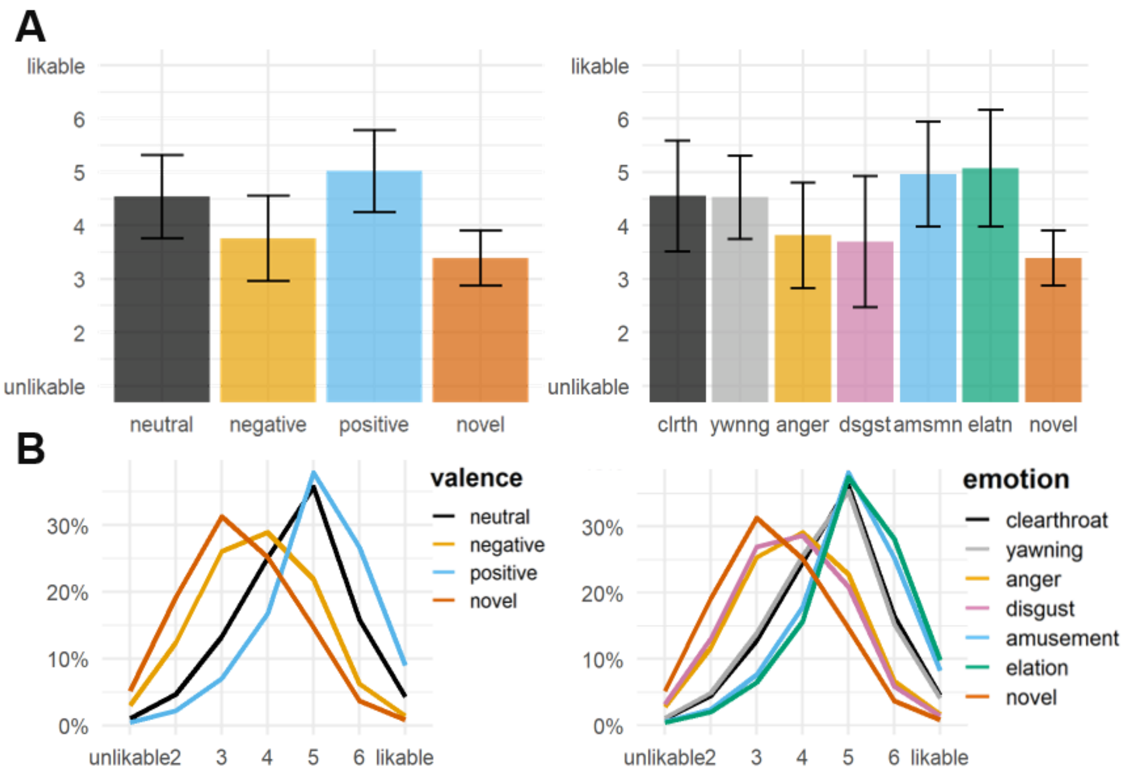
Face-locked LPC by valence.



Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand-averages of the ROI mean amplitudes, contrasted for the implicit and explicit task and all valence conditions. Errorbars indicate +/- 1 SE of the mean. **C** Topographies of the ERP distribution for faces between valence conditions, separately for the implicit (old-new) and explicit (valence class.) tasks. ROI channels are highlighted in pink.

doi: 10.53846/goediss-9841

Figure 4.7
Likability rating of faces.



Notes: Left panels show the likability by valence, right panels by emotion; **A** Barplots represent the likability ratings per condition, averaged within and across subjects. Errorbars show +/- 1 SD. **B** Fitted values are presented as predicted probabilities of the ordinal models.

doi: 10.53846/goediss-9841

in an online learning paradigm, we measured short-, mid-, and long-latency ERPs to faces associated with positive, neutral, or negative valence in a valence-implicit and a valence-explicit task. Consistent with our hypotheses and previous research, we found that faces previously associated with affective bursts were not only rated according to the valence of the context but also elicited differential neural responses from faces associated with a neutral context. Moreover, associated effects in late components were strongly affected by task requirements, suggesting that goal-directed attention on specific associated features affected especially later, more elaborate processing of the faces.

The first associated effect was present in the **N170**. Although the averaged associated valence did not moderate N170 amplitudes, there were differences between single emotion levels, with an enhanced negative amplitude for disgust-associated faces. A number of studies reported N170 effects for valence-associations (e.g., Aguado et al., 2012; Bruchmann et al., 2021; Camfield et al., 2016; Luo et al., 2016; Schellhaas et al., 2020; Schindler, Bruchmann, Krasowski, et al., 2021; Sperl et al., 2021). Due to its measured spatial overlap with the EPN, the N170 has been suggested to represent a mixture of configural face processing and relevance encoding (Rellecke et al., 2012b). In addition, there was an independent effect of task starting in the N170 time window and extending to a positive-going deflection over the lateral occipito-temporal areas peaking around 200 ms (similar to findings by Itier & Neath-Tavares, 2017; Schindler, Bruchmann, Krasowski, et al., 2021). The interpretation of this effect is not straightforward: the visually evoked P2 component has been linked to higher-order configural processing (Latinus & Taylor, 2006), differences in task difficulty (Philiastides, 2006), tasks requiring expertise to subgroups of faces (Stahl et al., 2008), and face typicality (Pell et al., 2022), all of which could be roughly related to deeper processing demands (Banko et al., 2011) of faces in the valence-classification task and differences in processing depth (Itier & Neath-Tavares, 2017). Remarkably, these early task differences did not extend to the EPN time window.

EPN amplitudes were modulated by associated valence, with enhanced amplitudes for the negative compared to the positive and neutral conditions. Several studies reported enhanced neural processing towards negatively but not positively associated faces (e.g., Luo et al., 2016; Suess et al., 2014; M. J. Wieser, Gerdes, et al., 2014). This negativity bias has also been shown for threatening facial expressions (e.g., Schupp et al., 2004; for a review, see Schindler & Bublatzky, 2020). That negatively-associated faces were preferentially processed in our study is remarkable in that affect bursts resemble rather low-intense stimuli. In addition, task neither modulated EPN amplitudes nor did it moderate valence effects, suggesting a fast and automatic allocation of attention toward negative information related to faces (similar to Baum & Rahman, 2021; Bruchmann et al., 2021;

doi: 10.53846/goediss-9841

cf. Schindler, Bruchmann, Krasowski, et al., 2021). Similar to the N170, EPN amplitudes were particularly pronounced for disgust-related faces. Usually, expressions of disgust serve to detect and reject things potentially offensive, toxic, or contaminating to keep oneself safe and healthy (e.g., spoiled food or open wounds). Expressions of disgust directed at us could also serve as a social-communicative signal and be interpreted as a risk of social exclusion (e.g., Amir et al., 2005; Gan et al., 2022; Judah et al., 2015). Although facial expressions of disapproval, disgust, and anger have been shown to trigger different neural processes (Burklund et al., 2007), auditory expressions of disgust could be perceived as more ambiguous and give more room for interpretation of social disapproval.

We hypothesized that the attentional focus of the task would especially affect later processing. Consistent with this hypothesis and previous research on the processing of faces with emotional expressions (for a review, see Schindler & Bublatzky, 2020) and conditioned faces (Schindler, Bruchmann, Krasowski, et al., 2021; cf. Bruchmann et al., 2021), associated valence modulations of the **LPC** were only descriptively present in the valence-explicit task. While early ERPs in the test session showed only effects of negative associations and specifically of disgust-related faces, later processing was modulated by both negatively and positively associated faces with no such strong differences between single emotion categories. Moreover, also positively associated valence was not extinguished but instead triggered by goal-directed memory retrieval, although it did not show in the valence-implicit task. In our study, LPC effects are strongly related to the task-relevant goals while at the same time differentiating between affective and neutral but not between positive and negative associations. Our results add to findings of previous research reporting LPC effects of positively associated faces Hammerschmidt, Kagan, et al. (2018) and other kinds of visual stimuli (e.g., Schacht et al., 2012) and show that also positive affect burst can be cross-modally associated to faces.

P1 amplitudes were not modulated by task and, differently from what we predicted, not modulated by associated valence. The P1 has been related to the processing of lower-level stimulus properties, and selective attention through sensory gain mechanisms (Hillyard & Anllo-Vento, 1998; Russo, 2003). Although several studies have reported a sensitivity for valence-based associations of the P1 (e.g., Aguado et al., 2012; Hammerschmidt et al., 2017; Muench et al., 2016; Schacht et al., 2012) and even earlier processing (e.g., Rehbein et al., 2014; Sperl et al., 2021; Steinberg et al., 2012), other studies have not investigated (e.g., Baum & Rahman, 2021) or reported P1 effects (e.g., Hammerschmidt, Kulke, et al., 2018; Schindler, Bruchmann, Krasowski, et al., 2021). It is possible that the association with affective vocal stimuli of lower intensity in our study was not sufficient to elicit differential activation of the P1. Although associated emotional

doi: 10.53846/goediss-9841

expressions of the face have been shown to modulate P1 amplitudes (Aguado et al., 2012), and comparable effect sizes of cross-modal and within-modal associations have been reported (Hofmann et al., 2010), the variability in learning might have played a more important role. As other studies reported stable associations after very few conditioning trials (e.g., Rehbein et al., 2014; Steinberg, Bröckelmann, Rehbein, et al., 2013; Ventura-Bort et al., 2016), it is unclear what drives neural changes at early processing of conditioned stimuli (e.g., the number of CS-US couplings, the (dis-) similarity between CS, the intensity of the US, the stimulus duration, or the consolidation period, etc.). By including a learning criterium in our study, we ensured associations between the faces and affective bursts. In addition, we included refresher trials between the valence-implicit and valence-explicit task to counteract extinction. However, the number of face-voice conditioning trials varied between participants and thus differed from typical conditioning studies. Some participants developed their own strategies and preferred studying the pairs by doing learning checks, which allowed them to see the faces longer and to get feedback on their answer. However, in these trials, not one face but five faces and the voice were presented simultaneously. Possibly, the association of the face and the voice occurred here on a more explicit level and was rather defined by attending to specific facial features then gradually tuning sensory discrimination through associative learning.

We included valence ratings of the voices prior to any association with faces. Overall, participants rated the vocal expressions according to our pre-specified valence categories. Interestingly, ratings on their reaction towards the bursts were less extreme than the expression ratings and showed a larger inter-individual variation. Behavioral performance between valence categories only differed in the learning phase of our study, in which faces with positive bursts were learned faster (similar to reward-associated faces in Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017; and reward-associated words or symbols Bayer et al., 2018; Kulke et al., 2019; Rossi et al., 2017). In contrast, during test, there was no clear evidence for accuracy and response times being affected by associated valence, although descriptively, in the valence-explicit task, responses for the neutral condition were slower than the positive and negative conditions. Nevertheless, as expected, RTs were shorter and accuracy higher in the old-new task than in the valence-classification task, probably due to the number of choices (two vs. three) and required depth of processing needed (recognition vs. explicit recall). The old-new task might have become more difficult over time due to the repetition of the novel faces. However, the behavioral results suggest that, overall, the valence-classification task was more difficult than the old-new task. If only cognitive load suppressed ERP effects of associated valence, we would have expected it to occur in the more difficult, i.e. in the valence-classification task.

doi: 10.53846/goediss-9841

The **likability rating** at the end of the test session was affected by associations with voices expressing positive and negative emotions during the learning (similar to Suess et al., 2014). More specifically and as we hypothesized, ratings were according to our pre-specified valence categories, while emotion within valence categories did not differ. It is possible that valence rather than the specific emotion category changed decisions on likability, although we cannot rule out the possibility that the preceding valence-classification task increased homogenization within valence conditions. Although likability ratings supported the ERP results, in our opinion, the ratings may resemble rather contingency awareness then true changes in likability and furthermore might be biased due to the focus on valence differences in the preceding task.

Our novel learning paradigm allowed participants to study face-voice pairs flexibly, which participants made use of, as documented by the learning strategy questionnaire. Despite some variation, participants followed similar self-chosen strategies to memorize the pairs, although we did not include any hints or recommendations on how to study the face-voice pairs. Participants actively searched for distinct facial features to combine them with what, in their view, would fit the emotional valence of the voice (e.g., the man with tired eyes yawned, the woman with warm brown eyes giggled). As the pairing of the faces and voices was randomized, participants reported taking whatever features would differentiate best between the faces and voices, and some participants even took notes to study. Hence, the type of learning was very different from classical Pavlovian conditioning or instrumental learning, in which associations might form more gradually. It is remarkable that faces associated with moderately negative bursts elicited distinct neural activation irrespective of the task requirements and despite this variability in learning. One limitation of the study might be that, although we chose this option deliberately, we fixed the order of the tasks in the test session (refresher I, old-new task, refresher II, and the valence-classification task). To ensure that only the valence-classification task would trigger explicit attention to the valence-based associations and to prevent spill-over effects to the valence-implicit task, we set the valence-explicit task at the last position of the experimental part, in which we recorded ERPs . Nevertheless, early effects were similar between tasks and the valence effects in the LPC occurred only in the valence-classification, i.e., second task, which should have been more prone to be affected by extinction of the associations or simply fatigue.

The study provides new evidence that faces cross-modally associated with affective stimuli of both positive and negative valence have the potential to elicit neurophysiological responses similar to inherent affective stimuli. During test, task demands affected later, more effortful processing, whereas earlier processing indicated an automatic discrimination of negative from other information across both tasks. We demonstrated that associations with even mildly negative stim-

doi: 10.53846/goediss-9841

uli, flexibly acquired through our novel learning paradigm, could influence face processing even in a valence-implicit task, suggesting a fast prioritization of learned negative context as a protection against potential threats (e.g., Lundqvist & Öhman, 2005; Öhman et al., 2001) largely independent of goal-directed attention. Moreover, positive associations were learned faster and affected later processing, but only in the presence of goal-directed attention toward valence.

Funding: This work was supported by Deutsche Forschungsgemeinschaft, Grant/Award Number: 254142454 / GRK 2070. **Acknowledgements**: We thank Emma Koch, Hanna Uphaus, Yasmin Fiedler and Lina Meiners for their help with recruitment and data collection. **Author contributions**: Annika Ziereis: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing - Original Draft, Writing - Review & Editing; Anne Schacht: Conceptualization, Methodology, Writing - Review & Editing, Supervision.

doi: 10.53846/goediss-9841

Chapter 5

Validation of scrambling methods for vocal affect bursts

Abstract

Neuroscientific studies require methods to disentangle basic sensory processing of physical stimulus properties and the recognition and processing of the semantics of a stimulus. Similar to image scrambling, the scrambling of auditory signals aims at creating stimulus instances that are not recognizable but have comparable low-level features. In the present study, we generated scrambled stimuli with four different methods (frequency-, phase-, and two time-scrambling transformations) of short vocalizations taken from the Montreal Affective Voices database (Belin et al., 2008). The original stimuli and scrambled versions were validated by $N = 60$ participants for the apparency of a human voice, gender, and valence of the expressions, or if no human voice was detected, for the valence of the subjective response toward the stimulus. The human-likeness was reduced for all scrambled versions, albeit to a lesser degree for phase-scrambled versions of neutral bursts. For phase-scrambled neutral bursts, valence ratings were equivalent to the ratings of the original neutral burst. All other scrambled versions were rated as slightly unpleasant, suggesting that although valence differences between emotions were reduced, they should be used with caution due to their potential aversiveness.

doi: 10.53846/goediss-9841

5.1 Introduction

Are emotional stimuli processed differently from non-emotional stimuli? As simple as this question appears, and irrespective of the stimulus modality, it is methodologically challenging to disentangle the impacts of emotional quality and of other stimulus properties. For example, the image of a smiling person and an image of the same person with a neutral facial expression differ in terms of emotional valence but also, to some degree, in their low-level properties. In this example, local differences in low-level properties might even increase with the intensity of emotional expressions shown, e.g., smiling with the mouth open and showing teeth will result in a higher number of bright pixels at the mouth region compared to the corresponding closed mouth or a neutral expression. Especially physical stimulus features like luminance, size, and contrast impact early visual processing (e.g., Bobak et al., 1987; Johannes et al., 1995; Korth & Nguyen, 1997; Marcar & Wolf, 2021), leading to problematic confounds of emotion-related effects and other stimulus effects, particularly relevant for electrophysiological and imaging research. Furthermore, any modification of a stimulus will result in changes of both the stimulus and its processing. However, not all stimulus properties are related to the inherent emotional meaning of a stimulus. Thus, to differentiate between emotion-sensitive and emotion-insensitive functional processing units (e.g., single neurons or larger spatial and temporal regions of interest), one would need to keep low-level properties comparable but eliminate the properties related to the emotional value of an image. For instance, the face-sensitive N170 event-related potential (ERP) component (e.g., Bentin et al., 1996) has been suggested to be already sensitive to emotional expressions (for reviews, see Hinojosa et al., 2015; Schindler & Bublatzky, 2020; but see Rellecke et al., 2012b). Whether such early differentiation of a signal is based on a functional detection of an emotional quality or on the reactivity to the confounded low-level features has important implications for theoretical models of face perception (Bruce & Young, 1986). One frequently used method is to compare the processing of intact images with scrambled versions of the images. There are different forms of visual scrambling, for example, shuffling individual or chunks of pixels (used in, e.g., Cano et al., 2009; George et al., 1996; Herrmann et al., 2004; Latinus & Taylor, 2006; Linkenkaer-Hansen et al., 1998), or shuffling windows in the frequency or phase domain (used in, e.g., Jacques & Rossion, 2004; Rossion & Caharel, 2011; Schindler, Bruchmann, Gathmann, et al., 2021), combinations thereof (used in, e.g., Coggan et al., 2017; Sadr & Sinha, 2004), using cyclic wavelet transformations (Koenig-Robert & VanRullen, 2013), or computational models of object recognition (Stojanoski & Cusack, 2014). All methods have in common that they are implemented with the aim to preserve some of the low-level properties (e.g., luminance, color histograms, frequency spectrum, contrast) and, at the same time, eliminate the identifiability or the semantic properties of a stimulus.

doi: 10.53846/goediss-9841

Analogously to the visual domain, the same potential confounds apply in the context of auditory processing. Thus, the investigation of emotional sounds and affective prosody in speech or non-speech vocalizations requires methods to create non-emotional references with comparable low-level properties (Jürgens et al., 2018; Lausen & Hammerschmidt, 2020). Scrambled versions of auditory stimuli have been particularly implemented to detect voice-sensitive and voice-selective areas in the human auditory cortex (e.g., Belin et al., 2002). Scrambling has also been used in investigations on the sensitivity of the amygdala, insula, and superior temporal sulcus to emotional sounds and human vocalizations (Zhao et al., 2016) and in research on music (Menon & Levitin, 2005). Similar to the visual domain, auditory scrambling involves procedures such as time scrambling, i.e., cutting the signal into time bins and shuffling them (used in, e.g., Angulo-Perkins & Concha, 2019; Jiang et al., 2013; Menon & Levitin, 2005; Wilf et al., 2016), phase scrambling (e.g., Gazzola et al., 2006; Y. Zhang et al., 2021), frequency scrambling (Barbero et al., 2021; e.g., Belin et al., 2002), gammatone filter banks (Minagawa-Kawai et al., 2010; Patterson et al., 1995), or combinations of methods (Coggan et al., 2016; e.g., Dormal et al., 2018), preserving different kinds of low-level features of the stimulus. The emerging question is, which method (incl. specific parameters) shall be used? In some cases, the method of choice might be directly related to the to-be-preserved stimulus features. However, in other cases, one might investigate which low-level features elicit the (neural or behavioral) effects of interest.

The aim of the present validation study was to compare different scrambling methods to create neutrally valenced instances of human non-speech vocalizations (affect bursts). Moreover, we were interested in how the scrambled versions would be rated in terms of their valence and whether certain levels of stimulus semantics would be preserved, e.g., if a human voice or/and the speaker's gender would be detectable. The two main reasons for this were to a) find stimuli for experimental tasks (e.g., a gender decision task) other than passive-listening or no-go tasks and b) investigate potential valence effects of stimuli as a result of the scrambling procedures.

5.2 Method

Participants

Data was collected from 62 participants, of which two were excluded from the following analysis as they did not differentiate between the original stimuli. The remaining 60 participants (41 female, 19 male, 0 diverse; M_{age} = 29.2 years, range$_{\text{age}}$ = [18; 70]) reported normal or corrected to normal hearing. Participants were informed about the procedure of the study and data policy,

doi: 10.53846/goediss-9841

and informed written consent was obtained. For reimbursement, participants could choose between course credit or a sound file of a scrambled version of their voice.

Stimuli

Original sound stimuli were short affect bursts, i.e., non-speech vocalizations, of a validated database, ("Montreal Affective Voices," Belin et al., 2008). We selected affect bursts of ten different speakers, half of which were female and half male (ID6m, ID42m, ID45f, ID46f, ID53f, ID55m, ID58f, ID59m, ID60f, and ID61m). From every identity, we included bursts expressing anger, happiness or a sustained neutral tone. The stimuli were of variable duration, ranging from 0.24 to 2.61 seconds.

Scrambling

We used four different scrambling methods to manipulate the original stimuli: frequency scrambling, phase scrambling, and two versions of time scrambling. All methods resulted in different acoustic aspects. The amplitude envelope remained similar for the frequency sampling, whereas all other methods changed the envelope towards a more uniform shape, with more spiky envelopes for the time-scrambled versions. Frequency and phase scrambling preserved the overall energy, which was to some degree reduced for the time scrambling due to the implementation of amplitude ramps (see below). The Python code for the different scrambling methods and scrambled versions of the stimuli can be found at https://osf.io/uat6m. An exemplary visualization of the sound envelopes and frequency spectra of one original and its manipulations are shown in Figure 5.1.

Frequency scrambling. We used an adapted version of the frequency scrambling of (Belin et al., 2002): after importing the audio files of the original stimuli and normalizing the amplitudes, we trimmed the array of samples to obtain full-sized windows (1024 samples per window) for the Fourier transformation. In incremental steps of 512 samples, we used the real fast Fourier transformation and shuffled the respective frequencies (by shuffling the positions of the Fourier transformed values) while keeping the amplitude as of the original window. After applying the inverse fast Fourier transformation, all windows were combined and normalized.

Phase scrambling. The phase scrambling was oriented to the idea by Gazzola et al. (2006). Instead of using an arbitrary threshold, we used stimulus-specific frequency thresholds to account for

doi: 10.53846/goediss-9841

gender and valence-specific differences. The median pitch of every stimulus was extracted with the software Praat (Boersma & Weenink, 2018). Based on descriptions by Belin et al. (2008), we used a larger pitch analysis range (75 to 2000 Hz) for the pitch extraction to account for female and male affective bursts. Differently from the frequency scrambling method, after Fourier transformation, frequencies were separated based on the threshold frequency (pitch). We scrambled the phases of the higher frequencies by power-transforming the amplitudes, taking the arc tangent, and shuffling the array. The inverse transformed array was then merged with the unshuffled values, back-transformed into the time domain, and normalized.

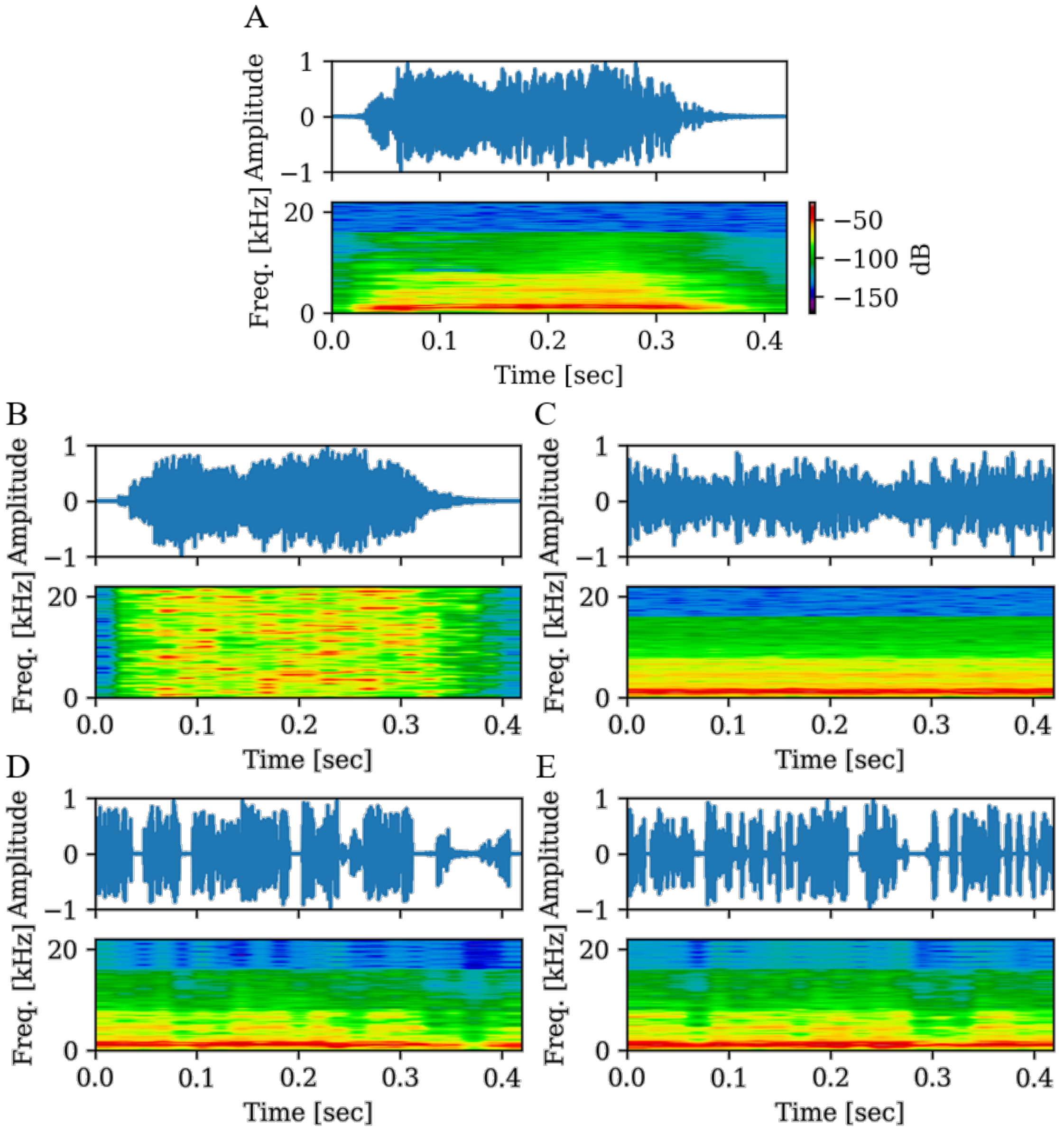
Time scrambling. The resolution for temporal differences in human hearing is around 4 ms (Samelli & Schochat, 2008). Based on this, we cut the normalized sound files into 6 ms (and 12 ms) windows, shuffled them, and added a one ms amplitude ramp at the beginning and end of each bin to erase crackling noise between recomposed windows. The sound files were normalized before their export.

Procedure of the validation study

We tested a maximum of ten participants at the same time in a group laboratory. All participants were seated in front of separated test cubicals provided with headphones (Beyerdynamic DT 770 PRO) and laptops (Dell Notebook E5530), all set at a constant, medium volume level. For the stimulus presentation and ratings, we used the survey tool formR (Arslan et al., 2019). After general information about the study and after having provided written consent and sociodemographic information, participants were presented with an example sound stimulus together with the respective rating scales. Before starting with the main validation, open questions about the procedure could be clarified with the experimenter. Participants were instructed to assess the presented sounds along different dimensions. There was a total of 150 stimuli (10 identities $\times$ 3 valences $\times$ 5 manipulations) to be rated by each participant. The stimuli were presented in random order and in individual trials. The questions and rating scales of the validation study are shown in Figure 5.1. Every trial started with the sound file played once automatically and the question, whether a human voice was apparent in the audio sample together with a four-point Likert scale, of which the extremes were labelled with “not at all apparent” (1) to “clearly apparent” (4). Depending on the response provided in this rating, different follow-up questions were presented. If participants indicated (1) or (2) in the initial rating concerning the presence of a human voice they judged, “what effect does the audio example have on you personally?”. If participants rated the presence of a human voice with (3) or (4), they were asked about the speaker’s gender (“not identifiable”,

doi: 10.53846/goediss-9841

Figure 5.1
Example of the amplitudes (envelope) and the power spectral density of the original stimulus and scrambled versions.

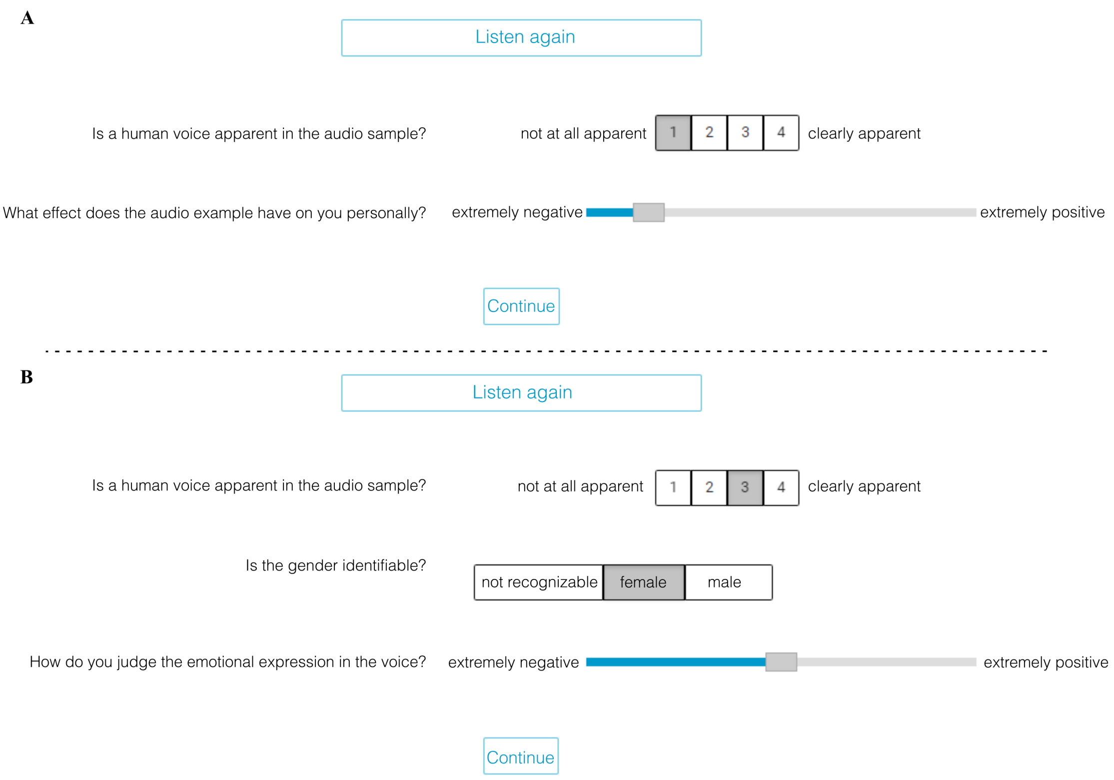


Notes: **A** original, angry bursts stimulus of ID46. **B** Frequency-scrambled version, **C** phase-scrambled version with scrambling frequencies above the median pitch of the stimulus. **D** Time-scrambling of 12 ms windows, **E** Time-scrambling of 6 ms windows. Both time-scrambling versions include amplitude ramps of 1 ms at the beginning and end of each window and thus differed from the original stimulus also in the overall energy.

doi: 10.53846/goediss-9841

"female", "male") and the emotional expression of the voice on a slider with labeled poles (left: "extremely negative" and right: "extremely positive"; as in Belin et al., 2008). The same response slider as for the expression rating was shown. Only the poles of the response sliders were shown and no ticks. However, internally, values were recorded from 0 to 100 in steps of 1. Participants could listen again to the audio file by clicking on a button presented centrally at the top of the window. Before submitting ratings and continuing with the next audio sample, answers could be changed. After responding, it was not possible to return to previous audio samples.

Figure 5.2
Procedure of the valence rating.



Notes: Two example trials of the valence ratings are shown. Auditory stimuli are played automatically at the beginning of a trial. However, participants could listen more often to a stimulus by clicking on 'listen again'. The first question was always about the identifiability of a human voice in the stimulus. In **A**, a participant indicated that no human voice was recognizable. Consequently, they rated the stimulus on their subjective reaction. However, in **B**, participants indicated that a human voice was present in the stimulus. In this case, they were asked whether the speaker's gender was identifiable and how they would judge the valence of the speaker's expression. Note that the slider poles and appearance were the same for both valence ratings, but the questions differed. After submitting their answers by clicking on 'continue' participants could not go back to previous stimuli.

doi: 10.53846/goediss-9841

5.3 Results

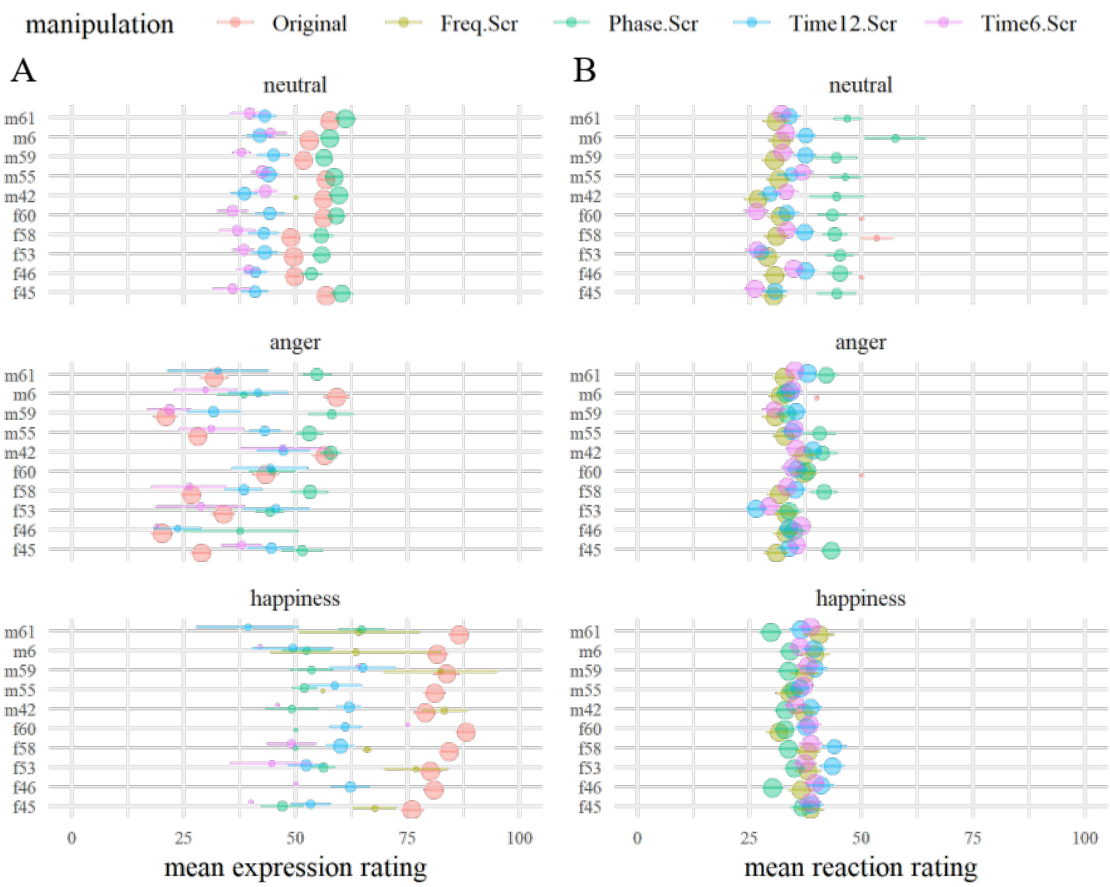
Valence ratings. The type of valence rating depended on whether participants detected a human voice in the sound file. For the original samples, almost all participants detected human voices, whereas, for the manipulated stimuli, participants varied in terms of the human voice categorization, which led to unbalanced group sizes of the ratings and rating types. Panel A of Figure 5.3 shows the mean ratings of the speaker's expression in case of a detected human voice. Analogously, panel B shows the mean ratings of the participant's reaction toward the stimulus in case no human voice was detected in the sample.

Gender classification and accuracy. If participants indicated the presence of a human voice in a stimulus, they were asked to categorize the speaker's gender. Accuracy of the gender decision was highest for the original stimuli, although for female anger and female neutral stimuli, there existed some uncertainty. Although overall scrambling introduced more uncertainty, more correct compared to incorrect, and more correct than unsure gender ratings were obtained (see Table 5.1).

doi: 10.53846/goediss-9841

Figure 5.3

Mean valence ratings by rating type, stimulus ID, valence, and manipulation method.



Notes: Expression and reaction ratings are shown separately for every speaker ID (y-axis) and emotion category. **A** displays mean valence ratings of stimuli, in which a human voice was detected. Values represent the rated valence of the speaker's expression. **B** displays mean valence ratings of stimuli in which *no* human voice was detected. Here, valence ratings refer to the participants' reported reaction toward the stimulus. Both sliders' poles included the labels 0 = 'extremely negative' and 100 = 'extremely positive'. Error bars show ±1 SE.
As unequal numbers of ratings contributed to the valence rating means, we included dot size as a proxy for the number of ratings on which the mean was calculated. Smaller dots indicate fewer ratings, i.e., fewer participants rating the stimulus with regard to the respective rating type (expression vs. reaction).

doi: 10.53846/goediss-9841

Table 5.1
Accuracy of gender ratings of the voices in case a voice was detected.

		Original				Phase.Scr				Freq.Scr				Time12.Scr.				Time6.Scr			
Valence	StimID	total	corr.	wrong	unsure	total	corr.	wrong	unsure	total	corr.	wrong	unsure	total	corr.	wrong	unsure	total	corr.	wrong	unsure
neutral	f45	60	55		5	47	32	2	13					17	13	1	3	9	8		1
	f46	58	53		5	21	13		8					12	9		3	8	7		1
	f53	60	52	2	6	42	30	3	9					26	21	1	4	15	12		3
	f58	57	42	3	12	33	21	1	11					17	16		1	11	9	1	1
	f60	59	51		8	43	25	2	16					21	16		5	14	12		2
	m42	58	58			52	47		5	1			1	24	17	1	6	20	14	3	3
	m55	59	58		1	55	51		4					33	29		4	20	19		1
	m59	59	59			47	46		1					20	20			7	5		2
	m6	60	60			55	53		2					28	26		2	9	8		1
	m61	60	59	1		55	53		2					21	19		2	14	11	1	2
anger	f45	59	53	1	5	12	5	1	6					12	11		1	7	6		1
	f46	60	49	1	10	4	2		2					3	3			1	1		
	f53	59	47	3	9	8	3		5					7	7			5	4		1
	f58	59	42	8	9	25	8	4	13					13	11		2	4	3		1
	f60	58	34	12	12	5		2	3					5	2	1	2				
	m42	58	58			28	18		10					8	5		3	4	1	2	1
	m55	59	59			32	29		3					10	10			4	4		
	m59	59	59			8	4	1	3					12	8	1	3	12	11		1
	m6	59	59			3	2		1					5	4		1	2	2		
	m61	59	59			20	13		7					3	1		2				
happiness	f45	60	57	1	2	10	5	1	4	3		1	2	7	5		2	1	1		
	f46	59	58		1									11	10		1	1	1		
	f53	58	57	1		8	3	1	4	2			2	14	12		2	3	3		
	f58	59	59			2	1	1		2			2	30	29		1	6	5		1
	f60	59	59			1			1					6	3		3	1	1		
	m42	58	57	1		5	3		2	4			4	13	9		4	1	1		
	m55	59	59			6	3		3	1			1	5	3		2				
	m59	59	56	2	1	5	2		3	2			2	7	2	3	2	1		1	
	m6	59	58		1	5	3		2	4	2		2	7	3		4	1		1	
	m61	59	59			5	2		3	4	1		3	3	2		1				

Notes: Counts of correct, wrong and unsure answers per stimulus ID, valence and manipulation. Gender classifications were only obtained from participants if they classified a stimulus as entailing a human voice (total). The maximum number of counts equals the number of participants (N = 60)

doi: 10.53846/goediss-9841

Emotional valence of the scrambled affect bursts. To investigate how the scrambled stimuli were perceived in terms of their emotional valence, we decided to collapse ratings irrespectively of whether participants rated their emotional reaction to the voice or whether they rated the valence of the speakers' expression (see Figure 5.2), although we are aware that ratings differ in their meaning. Since the original stimuli were almost exclusively rated regarding the valence of the speaker's expression, the participant's personal reaction can not be inferred from these types of ratings. The opposite applies to the frequency-scrambled stimuli. On the one hand, by collapsing ratings, we decided on a relatively liberal criterion for being categorized as neutral, e.g., if participants were unsure about the valence of the speaker's expression, they might have been more likely to categorize them as neutral. On the other hand, in order for a stimulus to be "truly" neutral, neither the participant's reaction toward the stimulus nor the speaker's expression should be identified as very negative or positive. Instead of comparing ratings to a fixed point (e.g., the center of the scale), we compared ratings of all manipulations and valence categories with the original, neutral stimulus ratings, see Figure 5.4.

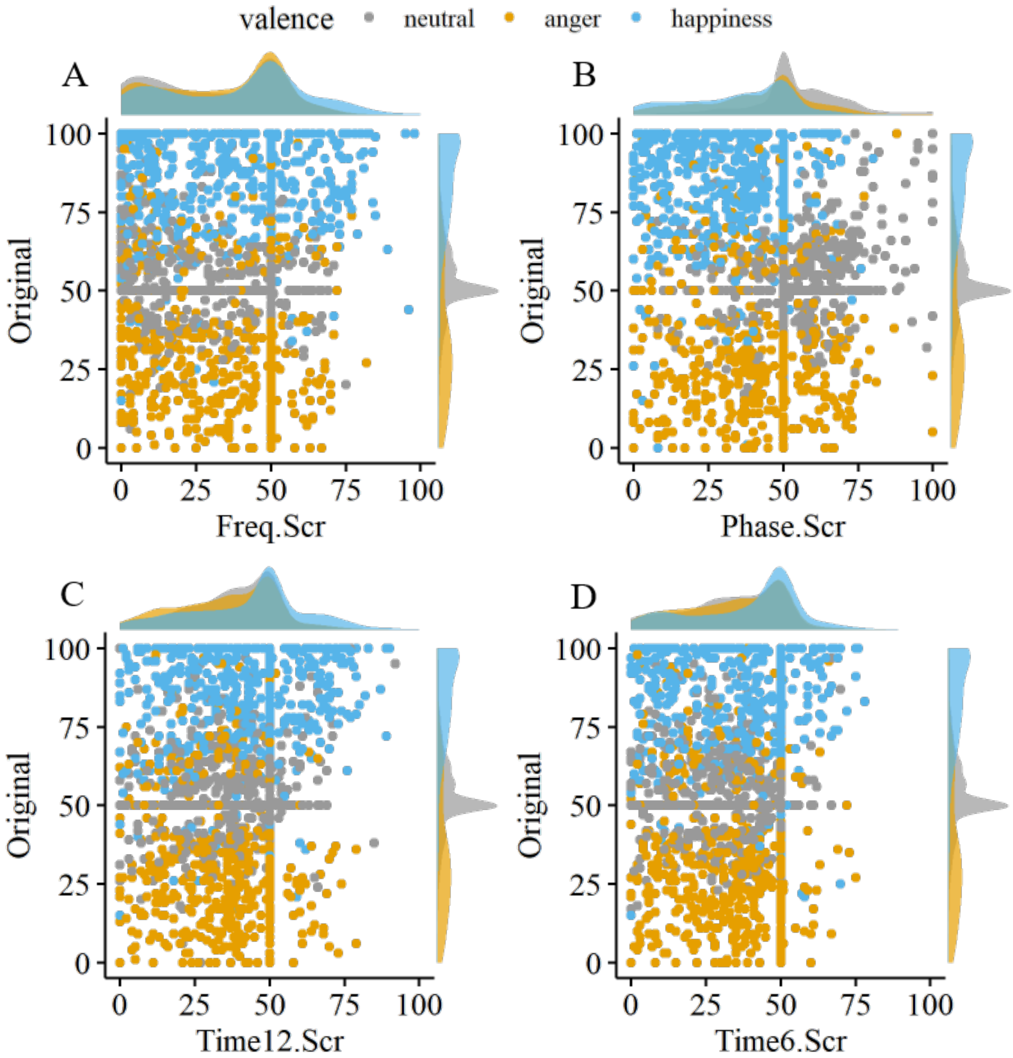
Equivalence tests of valence ratings on the scrambled voices. We conducted two one-sided tests of equivalence for paired samples to test whether the mean of differences between the scrambled and original stimuli of neutral valence are statistically equivalent. The hypothesis testing of this approach is different from normal paired sample tests, in which the null hypothesis states that the mean of the differences between two samples that are paired is zero. The null hypothesis of equivalence tests for paired samples states that the mean of differences is outside the equivalence interval $(-\delta, \delta)$, of which δs have to be chosen a priori. When the null hypotheses $H_{0(1)}$: $\mu_1 - \mu_2 \geq \delta$ and $H_{0(2)}$: $\mu_1 - \mu_2 \leq -\delta$ can be rejected, it can be inferred that the mean of the differences lays within the equivalence interval. Due to the non-normality of the voice ratings, we used a non-parametric version of the two one-sided test of equivalence for paired samples (NPAR, Mara & Cribbie, 2012). We chose δ as the standard deviation of ratings of the original, neutral burst (δ = 8.63) and compared it to each manipulation and each valence. We estimated 95% non-parametric bootstrapped confidence intervals (n_{boot} = 10000) around the differences.

Only phase-scrambled versions of neutral affect bursts were equivalent to the original neutral affect burst ratings (est = 1, *CI* = [-2.15, 4.35]). No other combination of scrambling method and original valence could be regarded as equivalent based on the ratings we obtained. Moreover, differences were negative throughout manipulations and valences, indicating a shift towards negative ratings compared to the original, neutral stimuli. The results of the model are shown in Table 5.2.

doi: 10.53846/goediss-9841

Figure 5.4

Scatterplots of individual valence ratings of the scrambled vs. the original stimuli.



Notes: Every dot represents the rating per stimulus ID and participant, where on the x-axis, the respective manipulated version is plotted against the rating of the unmanipulated, i.e., the original version on the y-axis. **A** shows ratings of the frequency scrambled, **B** of the phase scrambled, **C** of the 12 ms time scrambled, and **D** of the 6 ms time scrambled stimuli. Colors represent the valence category of the original stimulus. Densities of the valence ratings per valence categories are displayed at the top and right sides of the scatterplots.

doi: 10.53846/goediss-9841

Table 5.2
Results of the two one-sided equivalent tests for the scrambled stimuli.

Comparison	Difference	CI
Freq.Scr neutral	-17.55	[-32.95,-11.25]
Freq.Scr anger	-15.55	[-26.45, -7.40]
Freq.Scr happiness	-6.60	[-17.85, -3.90]
Phase.Scr neutral	1.00	[-2.15, 4.35]
Phase.Scr anger	-8.20	[-13.60, -5.45]
Phase.Scr happiness	-15.45	[-18.30,-12.75]
Time12.Scr neutral	-14.65	[-18.75,-10.85]
Time12.Scr anger	-16.25	[-18.70,-12.05]
Time12.Scr happiness	-7.05	[-9.75, -4.35]
Time6.Scr neutral	-18.90	[-21.40,-14.85]
Time6.Scr anger	-16.00	[-18.70,-12.00]
Time6.Scr happiness	-8.75	[-15.75, -5.25]

Notes:
All stimuli were compared to the ratings on the original version of neutral stimuli. CI = 95% non-parametric bootstraped confidence intervals.

5.4 Discussion

The present study compared valence ratings for auditory affect bursts and for different types of their scrambled versions, namely frequency, phase, and two time-scrambling approaches, with the aim of finding neutralized versions of affective stimuli while preserving some of their low-level features. All scrambling approaches decreased overall valence differences which were present between originally happy, neutral and angry affect bursts. However, none of the scrambling methods used in this study was able to create "neutralized" versions of all stimuli due to differential effects scrambling methods had on the original valence categories.

In addition to valence ratings, we were interested in whether stimuli still were perceived as entailing a human voice and gender information depending on the level of distortion of the scrambling methods. Both the judgments on how human-like the stimulus sounded and about the speaker's gender were affected by the scrambling method and by the valence category. Thus, none of the scrambling methods preserved gender information in all stimuli. Phase scrambled versions of neutral but not happy bursts tended to be rather classified as entailing a voice compared to not entailing a voice. The rate was also higher for the 12 ms time scrambling as for the 6 ms time scrambling and overall more pronounced for bursts that were of neutral valence originally. Possibly, this was due to the monotonous melody of neutral bursts, which did not change with the destruction of the temporal coherence. Although the frequency scrambling led to the lowest rate of recognizing a human voice in the stimulus descriptively, of these stimuli, the happy frequency-scrambled bursts

doi: 10.53846/goediss-9841

had the highest rate for detecting a human voice, probably, due to the very characteristic sound envelope of happy bursts (piecewise melody with many brief pauses in between). We observed that in cases a human voice was detected, gender information was still preserved to some degree, although scrambling increased the perceiver's uncertainty about the speaker's gender, shown by the accuracy of gender categorizations.

The applied scrambling methods failed to create truly neutral versions of the affect bursts. The clear separation between valence categories that was observable for the original stimuli was diminished but not completely eliminated for the scrambled versions. The largest difference between valence categories was found for the phase-scrambling. Participants' reaction ratings of phase-scrambled versions of originally neutral stimuli were overall closest to the center of the rating scale, i.e., "neutral" and thus descriptively more positive compared to phase-scrambled versions of happy and angry stimuli (a few participants mentioned that the neutral phase-scrambled stimuli sounded like synthesized sounds of a choir). Particularly phase-scrambled versions of happy bursts were rated descriptively as the most unpleasant of all manipulated happy stimuli. Nevertheless, when testing whether scrambling-valence combinations were equivalent to the original neutral stimulus category, only the phase-scrambled versions of originally neutral stimuli could be regarded as equivalent in terms of valence ratings. Moreover, other stimulus properties, such as gender information, were detected in phase-scrambled neutral stimuli to a higher degree.

Notably, there was a tendency for scrambled stimuli to be rated as more unpleasant than their original versions. To our knowledge, only a few studies included explicit valence and arousal ratings of scrambled stimuli. In contrast to our findings, Zhao et al. (2016) presented frequency-scrambled sounds and reported comparable valence and arousal ratings for scrambled and neutral sounds, besides that stimuli were rated as meaningless. However, time-scrambled classical music excerpts in Menon & Levitin (2005) were both rated as less pleasant compared to the original stimuli and as rather unpleasant. To detect potential response tendencies, we investigated the whole rating distributions (Figure 5.4) of the original stimuli and scrambled versions. Several things were noteworthy: there was some asymmetry observable for the original stimuli with more extreme (positive) valence ratings for happy stimuli, compared to the angry stimuli and neutral, original stimuli tended to be rated slightly positive. For all scrambling methods and the neutral, original stimuli, there were inflated ratings for the midpoint of the response scale. Due to the nature of the slider responses with initial thumb values, the resolution around the center is low, as participants tend to leave the slider with its default value if it is subjectively close to their (latent) rating. Notably, frequency-scrambled versions seemed bimodally distributed with a second peak at the negative end of the rating scale, i.e., some participants rated them as highly unpleasant.

doi: 10.53846/goediss-9841

Implications

We used explicit valence ratings in our study. Explicit ratings or categorizations can be seen as the integrated and cumulative outcome of encoding and appraisal processes and do not necessarily correspond to valence-driven effects at earlier, automatic processing stages (e.g., Hammerschmidt et al., 2017; Rossi et al., 2017; Roux et al., 2010; Walla et al., 2013; M. Wieser et al., 2006). Thus, our findings do not suggest that scrambled versions of auditory stimuli should not be used in investigations of auditory (emotion) processing. However, the assumption of using them as a neutral control may be flawed and may overshadow emotion effects in processing phases that are sensitive to general valence or arousal effects. Moreover, it might be problematic to use scrambled versions as references for difference measures (e.g., negative-scrambled vs. positive-scrambled), as sometimes used in imaging studies. In the case of valence differences between scrambled versions, the valence effects of interest might be falsely detected or not detected at all. If it cannot be excluded that measures of interest are insensitive to valence differences, it might be beneficial to test the homogeneity of scrambled stimulus responses beforehand.

Based on our findings, the question emerges whether there is a fundamental difference between visual and auditory scrambling. Visual scrambling methods have been criticized (e.g., Dakin et al., 2002; Stojanoski & Cusack, 2014) mainly for maintaining or non-maintaining important low-level features. However, only a few studies included assessments of valence and arousal for scrambled images, possibly due to the intuitive assumption that without recognizability of affective stimuli, there would be no valence effects (e.g., Braly et al., 2021). Another important aspect is that different low-level features might serve as general valence cues for recognizing emotional stimuli. For example, Delplanque et al. (2007) reported a confound of spatial frequencies and emotion effects for pictures selected from the IAPS database (Lang et al., 2005). Thus, valence effects can remain even in the absence (or reduction) of object recognition and even in the case of earlier processing. For example, arousal and valence of the original stimuli affected mid-latency event-related potentials (ERPs) and their spatially scrambled versions in Rozenkrants et al. (2007). In contrast, no valence effects on mid-latency ERPs of spatially-scrambled emotional pictures were reported by Cano et al. (2009).

This study does not come without its limitations. Due to the choice of different types of valence rating (i.e., rating the valence of the expression vs. the subjective reaction towards a stimulus), we could not directly compare ratings of all scrambled versions with the original stimuli. It would have been interesting to test the correlation between the rated valence of the expression and the personal reaction to the original stimuli (one might find a laughter highly unpleasant and

doi: 10.53846/goediss-9841

accurately classify the speaker's expression as positive). However, by including the expression ratings, we identified lower accuracies of the gender classification in some of the original stimuli and more variability in valence ratings for angry bursts. Thus, these stimuli might be problematic for certain experimental tasks. As our participants' age range was larger than the one of the original validation study by Belin et al. (2008), we checked whether the valence effects were related to the participants' age or gender, which was not the case.

Outlook

Different stimulus categories are potentially affected in different ways when being scrambled. Social stimuli like faces and voices might form special categories due to their high biological relevance (e.g., Belin, 2017) and typicality, e.g., faces have been shown to require a higher degree of scrambling before becoming unrecognizable (Stojanoski & Cusack, 2014). There might be a modality-specific divergence of scrambling effects between visual and auditory stimuli. Different from uncanny-valley effects (for a review, see Kätsyri et al., 2015) for only slight modifications of a facial stimulus (e.g., preserving external facial features but scrambling the eye and/or mouth region), strongly distorted auditory stimuli potentially become more aversive. For example, bursts of white noise are an effective aversive stimulus in fear-conditioning research (Sperl et al., 2016). The specific (non-linear) function of valence effects of visual and auditory scrambling is an interesting field for future research, especially in the context of research with artificial agents (e.g., Meah & Moore, 2014). A systematic comparison of scrambling methods on various distortion/preservation levels for social stimuli could help find adequate comparator stimuli and, at the same time, give insight into which low-level properties are relevant cues for social (re-)cognition and their subdomains, including the identification of emotional expression, gender, age, and identity.

Conclusion

Despite their benefits and intuitive use for being used as baseline or reference stimuli, scrambled versions of stimuli should be used with caution. The scrambling method should be selected based on specific hypotheses about which relevant low-level properties should be preserved or eliminated. If these are unknown, different scrambling methods can be compared. In the present study, we showed that for the auditory domain, scrambling methods could interact with the underlying stimulus category and produce potentially aversive stimuli. At least for emotion-related research, valence effects of scrambled stimuli should be explicitly tested and controlled for and not merely presumed to be "neutral".

doi: 10.53846/goediss-9841

Chapter 6

Additive effects of emotional expression and stimulus size on the perception of genuine and artificial facial expressions: An ERP study

Abstract

Seeing an angry person in close physical proximity will not only result in a larger retinal representation of this person but might also increase the motivation for rapid visual processing and action preparation. The present study examined the effects of stimulus size and emotional expressions on the perception of faces expressing happiness, anger, non-expressive faces, and scrambled faces. We analyzed event-related potentials (ERPs) and behavioral responses from N = 40 participants who performed a naturalness-classification task on genuine or artificially-created facial expressions. Whereas the difference in accuracy to detect artificial faces for emotional expressions was modulated by stimulus size, ERPs only showed additive effects of stimulus size and expression but no interaction with size, unlike previous research on emotional scenes and words. Effects of size were present in all included ERPs, whereas emotional expressions affected the N170, EPN, and the LPC, but not the P1, irrespective of size. The present findings suggest that the decoding of emotional valence in faces can already occur for small stimuli. Supra-additive effects in faces might require larger size ranges or dynamic stimuli that increase arousal.

doi: 10.53846/goediss-9841

6.1 Introduction

Being approached by a person staring angrily at you requires fast detection, interpretation of the situation, and prompt response - even more so if the person is already close by. Over decades, research has accumulated evidence in favor of the preferential processing of emotional facial expressions compared to neutral expressions (e.g., Bublatzky et al., 2014; Eastwood et al., 2003; Esteves, Dimberg, et al., 1994; Frischen et al., 2008; for a review, see Schindler & Bublatzky, 2020), likely due to their biological relevance (e.g., Dimberg & Öhman, 1996). However, also boundary conditions could be identified which moderated the emotion effects, such as task requirements (competing/high demanding tasks, e.g., Rellecke et al., 2012a; Schindler et al., 2020; Tannert & Rothermund, 2020) or re-appraisal (Bublatzky et al., 2018; e.g., Herbert et al., 2013). As face perception is strongly context-dependent (for a review, see M. J. Wieser & Brosch, 2012), also the (perceived) physical distance might modulate how we process faces and their relevance. The main goal of the present study was to systematically investigate the influence of stimulus size as a proxy for physical distance on the perception and processing of emotional facial expressions. To this end, we presented faces with real or manipulated happy, angry, or neutral expressions of different sizes and implemented a go/no-go task, in which participants decided on the naturalness of the emotional expression. Moreover, we included scrambled faces of different sizes as emotionally-meaningless control stimuli (no-go condition).

There are at least two different ways in which stimulus size can affect the processing of faces: First, especially early visual processing is affected by several dimensions, such as luminance (Bieniek et al., 2013), contrast (Bobak et al., 1987), and spatial frequencies (Zani & Proverbio, 1995), which can be indirectly also affected by a stimulus' retinal size (Loftus & Harley, 2005). Thus, stimuli of the same physical size but at close proximity will result in different visual processing, irrespective of their semantic meaning. Second, stimulus size correlates with the perceived physical proximity and distance, at least for stimuli of which the real size is known. The question arises whether the perception of biologically relevant stimuli, such as faces, would be intensified with increasing stimulus size, resulting in stronger emotion-based effects for emotional facial expressions compared to neutral expressions. There is some evidence in favor of a modulating effect of physical proximity. For instance, pictures of emotional and neutral scenes resulted in differences in autonomic responses, depending on the presented picture size (Codispoti & Cesarei, 2007). Similarly, skin conductance effects were larger between arousing and non-arousing video clips presented on big compared to small screens (Reeves et al., 1999). Noticeably, only a few studies have investigated the influence of stimulus size on the neurophysiological processing of emotional stimuli

doi: 10.53846/goediss-9841

using event-related potentials (ERPs) (e.g., words: Bayer et al., 2012; emotional scenes: Cesarei & Codispoti, 2006; feedback processing: Pfabigan et al., 2015; faces with administered pain: Lomoriello et al., 2018; and peripheral looming fearful and neutral faces Martin et al., 2021) and not all investigated both early and late processing. Bayer et al. (2012) and Cesarei & Codispoti (2006) reported interactions of stimulus size and valence condition at mid-latency occipito-temporal components (for pleasant compared to neutral scenes during 150-300 ms Cesarei & Codispoti, 2006; for positive and negative compared to neutral words from 340-480 ms Bayer et al., 2012), whereas both studies did not find interactions between emotion and size at earlier and late processing. In contrast, the face-sensitive N170 distinctively increased for looming fearful faces, and increased P3 effects of larger faces in different (administered) pain conditions were reported by Lomoriello et al. (2018), suggesting that for faces, early and late processing of emotion might be affected by size.

Particularly for early visual components, which are known to be sensitive to size, evidence for sensitivity to emotional expressions is inconclusive (Schindler & Bublatzky, 2020), leaving the question open about potential supra-additive effects of emotion and stimulus size, e.g., emotion effects might only become detectable when the size of a face exceeds a certain threshold. Moreover, despite the large body of research on emotional face perception, to the best of our knowledge, a systematic investigation of potential interactions between emotional expressions and the size of the face presentation is still outstanding. With the present study, we aimed to contribute to a better understanding of the emotion specificity of commonly reported ERPs in face perception and to identify one potential reason for the heterogeneous findings of early emotion effects in the literature.

To investigate the effects of emotion and size on face processing, we tested ERP components that have been related to the visual processing of faces and facial expressions of emotion: The **P1** is an occipital positivity with a bilateral distribution, usually peaking around 100 ms after the onset of a visual stimulus. Only a minority of studies reported modulations of the P1 by facial expressions (e.g., Bublatzky et al., 2014; Foti et al., 2010; Hammerschmidt et al., 2017; Rellecke et al., 2011). Its sensitivity for size has been reported in studies on the perception of other, more abstract stimuli (e.g., Busch et al., 2004; Kornmeier et al., 2011). The **N170**, a negative deflection over occipito-temporal regions, peaking around 170 ms, has been typically related to face perception due to its enhancement for faces compared to other objects (Bentin et al., 1996; Rossion et al., 2000). Emotion-related effects on the **EPN**, the early posterior negativity, have been reported not only for faces but also for other stimulus domains (e.g., Bayer & Schacht, 2014; Schacht & Sommer, 2009), which is why it is generally assumed to reflect selective attention to hedonic and arousing stimuli

doi: 10.53846/goediss-9841

(Schupp et al., 2006). The late positive complex (**LPC**, sometimes also late positive potential or *LPP*) appears to be sensitive to task requirements (e.g., Rellecke et al., 2012a), and emotion effects have been reported particularly for tasks that involved attending toward the affective content of stimuli (Schindler & Bublatzky, 2020).

Hypotheses

Overall, we hypothesized that different stages of face processing would be differentially affected by stimulus size, with stronger size effects irrespective of the emotional content in early processing and size-emotion interactions appearing rather at those processing stages which are sensitive for relevance detection. Moreover, due to the general task dependency of later processing, we hypothesized that the influence of size would depend on whether its internal representation is beneficial for the current task goals. Based on the biological relevance hypothesis, all size $\times$ emotion interactions should result in greater differences between emotional and neutral expressions in larger faces compared to smaller faces. More specifically, for the **P1**, we expected main effects of size with larger peak amplitudes and shorter latencies for larger stimuli (similar to Kornmeier et al., 2011) and differences between scrambled and intact faces due to the different contrast configurations. We were specifically interested in whether emotion effects on the P1 would be present for large faces, compared to small, to identify a potential reason for the inconclusive findings of emotion effects on the P1. Regarding the **N170**, we expected to replicate the face vs. non-face effect (e.g., Bentin et al., 1996) between intact and scrambled stimuli, which we also tested for differential modulations by size. As latency effects might carry over from earlier processing, we expected a latency effect of size on the N170 peak latency but not on amplitude. Moreover, faces with emotional expressions, especially angry faces, should produce a more pronounced N170 compared to faces with neutral expressions (Schindler & Bublatzky, 2020). In line with Bayer et al. (2012) and Cesarei & Codispoti (2006), we predicted for the **EPN** mean amplitude, a main effect of size, with more negative amplitudes for larger stimuli, a main effect of emotion with more negative amplitudes for happy and angry compared to neutral faces, and an interaction between emotion and size with pronounced emotion effects for larger stimuli. In addition, we predicted that intact faces would result in more negative amplitudes compared to scrambled stimuli. The **LPC** component should be modulated by size with larger stimuli leading to larger positive amplitudes. We expected that the naturalness decision for large faces would be performed with higher motivation, as facial features were presented with more detail, leading to a sustained motivation of large intact faces. We also expected a main effect of emotion with larger amplitudes for happy and angry compared to neutral expressions, as our task required focussing and evaluating the facial expressions, although

doi: 10.53846/goediss-9841

explicit emotional valence was not task-relevant. We have had no directed hypothesis regarding the influence of stimulus size on response times (RT), as studies were inconclusive regarding the direction (Busch et al., 2004; Cesarei & Codispoti, 2006). Similarly, emotion effects on RTs seem to depend on the specific task requirements. A number of studies reported processing advantages of angry faces, leading to faster responses (e.g., Öhman et al., 2001; Valk et al., 2015), whereas delayed disengagement from angry faces might result in slower responses (e.g., Eastwood et al., 2003; Schacht & Sommer, 2009). Regarding emotion, the naturalness decision for neutral faces should be rather difficult due to only subtle changes to the original stimuli. Thus, we expected happy faces to be responded to fastest and show the highest accuracy, but we had no clear prediction regarding the differences between neutral and angry faces. Due to the nature of the task, we did not expect a high rate of false alarms to scrambled images and therefore disregarded no-go trials for the behavioral data analysis.

6.2 Method

We preregistered this study on https://osf.io/7eyz6.

Participants

Data was analyzed from 40 participants (29 female, 11 male, 0 diverse; M_{age}= 22.98 years, SD_{age}= 3.23), our preregistered sample size. Of the originally recorded 43 participants, three datasets had to be excluded due to excessive artifacts resulting in less than 30 trials per condition. All participants had a good command of German, were right-handers (according to Oldfield, 1971), and reported no (neuro-) psychiatric disorders. We only included participants with normal or within plus/minus one diopter corrected-to-normal vision. Participants were recruited through the department's participant recruitment database, flyers distributed on campus, postings on social media (Twitter, Facebook), the online notice board of the university, and the website of the Institute of Psychology. Participants were reimbursed at an hourly rate (8.50 EUR) for the EEG session in the lab or course credits.

Stimuli

Faces were selected from the Göttingen Faces Database (Kulke et al., 2017) and the Radboud Faces Database (Langner et al., 2010). The faces were presented in greyscale on a light grey background. The face stimuli were edited and combined with a transparency mask that covers the

doi: 10.53846/goediss-9841

hairline, ears, and neck. Of 90 face identities, 45 displayed the original facial expression, whereas the other 45 (all of which had a neutral expression in their original version) were manipulated using a generative adversarial network (GAN) to create instances of neutral, angry, and happy expressions by scaling the intensities of selected action units (see below). Stimuli were presented in three different sizes at a viewing distance of approximately 78 cm. Measured sizes were for large stimuli 8.3 $\times$ 5.8 cm (6.09 $\times$ 4.26 visual degree (vd)), medium stimuli 5.5 $\times$ 3.9 cm (4.04 $\times$ 2.86 vd), and small stimuli 2.7 $\times$ 1.9 cm (1.98 $\times$ 1.40 vd). Stimuli had a resolution of 261 $\times$ 353, 172 $\times$ 232, and 86 $\times$ 116 pixels, respectively. Participants only saw one facial expression per identity and size. In addition, scrambled versions of a subset of faces were created by shuffling squares of pixels of the area of the face and adding a mask to account for the “ziggy”-edges. Due to the different lighting, images differed in terms of their luminance and contrast between databases, which we could not totally diminish without creating visible artifacts. Also, scrambling made stimuli slightly brighter despite scrambling within the facial area, probably due to the masking and edge effects. Quantiles of luminance per group and condition are shown in Figure D2 of Appendix D. Due to this confound, we refrained from directly comparing the effects of face artificiality but collapsed fake and real expressions per emotion and stimulus size.

Creating fake expressions. Although we were curious whether participants would be sensitive to real and manipulated facial expressions, the main reason for creating these stimuli was to have a larger stimulus set of different identities with comparable attributes. Classic face databases in neuro/physiological research, which include emotional expressions of the face, often have a limited number of identities and can not easily be combined due to their big differences in brightness, color, and contrast. Our study design required a large number of individual faces of different expressions to avoid memory and transfer effects of seeing the same face in different sizes (e.g., processing of a small face might be facilitated if one has already seen the same face in a larger version). A high-quality face database with many different identities is the Göttingen faces database (Kulke et al., 2017), with the limitation that it only includes faces with neutral expressions. With advances in the field of computer vision and artificial intelligence, fast and easily accessible applications for face processing, including facial expressions, have become available. Although they provided impressive results when manipulating single faces, the obtained results were highly susceptible to artifacts. Importantly, when including a range of stimuli, it became apparent that facial features were homogenized, and identities became poorly distinguishable once hair and image background was removed. Both freely available apps and commercial tools had a further disadvantage, such that image rights were not cleared. For some tools, image information would be used for the train-

doi: 10.53846/goediss-9841

ing and optimization of the algorithms. Since the licenses of the databases we were interested in using did not allow the transfer of data to third parties and facial information counts as sensitive data, we decided for an in-house solution to create happy and angry and neutral(-ized) expressions from previously non-expressing faces. An illustration of the procedure can be found in Figure D1 in Appendix D. In the first step, we used *OpenFace landmark detections* (Baltrusaitis et al., 2016) to automatically align, rotate and scale faces in an image. Images were then cropped and downsized to 128 $\times$ 128 pixels for the GAN. Expression manipulations were performed using the publicly available GAN *Ganimation-replicate* (Pumarola et al., 2018), which includes the pre-trained model *EmotionNet* and allows to customize action units intensities (AUs, Baltrusaitis et al., 2015; Ekman & Friesen, 1978). As the resulting images had a resolution of 128 $\times$ 128 pixels and contained face distortions, we used a different GAN devoted to restoring and upscaling small images containing faces and images of low quality, *GFPGAN* (Wang et al., 2021). Although preserving crucial identity information in the face, faces appeared a little posterized and "shiny". Thus, we also resized faces with real expressions to 128 $\times$ 128 pixels and upscaled them with GF-PGAN to diminish differences due to image restoration. Next, images were scanned for remaining strong artifacts/distortions, and a subset of identities and expressions were selected. An oval mask was then applied to all images to exclude extra-facial features, and all stimuli were converted to greyscale and normalized. To create scrambled versions of images, we used *scrambpy* (v0.5.0, *GitHub - Snekbeater/Scrambpy*, n.d.). We only shuffled chunks of pixels within the facial area to avoid including pixels from the image background and to keep the amount of bright and dark pixels constant but still be differentiable from the background. In the last step, images were resized to the respective small, medium, and large presentation sizes for our study.

Randomization

The stimulus set included 90 different face identities $\times$ 3 sizes + 36 scrambled images $\times$ 3 sizes. To ensure an equal number of images for each size and no image is presented in different sizes, we pseudo-randomized the size of each face and scrambled stimulus for each participant at the beginning of their session. All participants saw all face identities and all scrambled images; however, one respective stimulus only in one size. Participants performed 7 repetitions of the respective 90 + 36 = 126 images. For each block of repetitions, the 126 images were shuffled, resulting in 882 trials for the whole session.

doi: 10.53846/goediss-9841

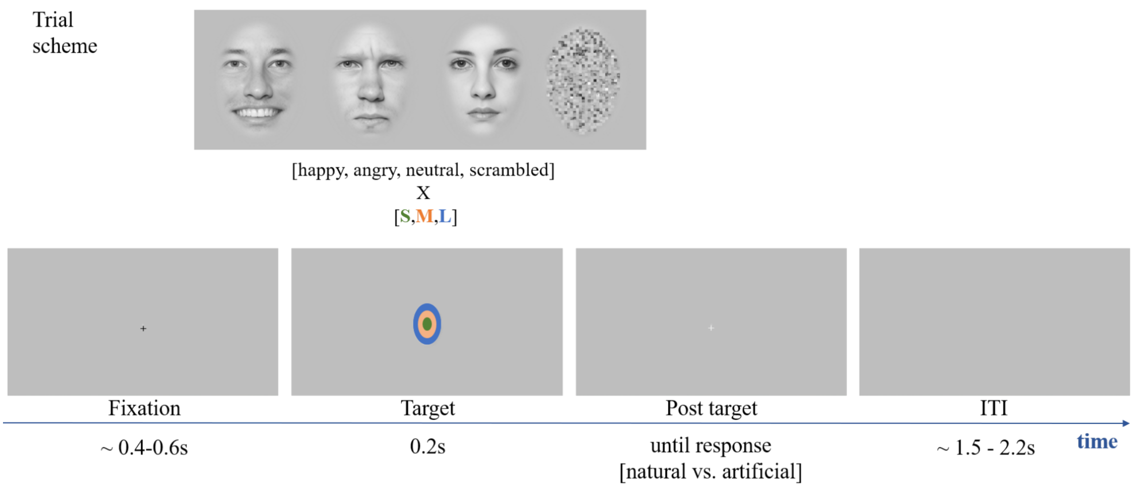
Procedure

The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the Institute of Psychology at the University of Göttingen. Participants were fully briefed on the study process (all study steps, compensation, and approximate time required) at recruitment when signing up and at the beginning of the experimental session. Sociodemographic information was obtained when participants signed up for the study to assess their eligibility, and written informed consent was obtained at the beginning of the experimental session. Participants were seated in a dimly-lit and electrically shielded room in front of a computer screen at a distance of approx. 78 cm between eyes and presented face stimuli. To avoid head movements, participants positioned their chins in a height-adjustable chin rest. The experiment was presented using functions of PsychoPy (Peirce, 2009) in Python (v2.7). We used PyGaze (v0.6.5, Dalmaijer et al., 2013) for communication with the eye-tracker. After calibrating the eye-tracker (9-point calibration), participants were instructed about the task and performed four example trials, on which they obtained feedback on whether they responded correctly. Participants were instructed at the beginning that they would see faces with happy, neutral, or angry expressions, of which some were manipulated, whereas others would be real. We explained that manipulation was performed with a "neural network", which could alter the facial expression (i.e., the person depicted might have shown a different expression at the time the picture was taken). The participant's task was to indicate via key press whether the presented face showed a natural or an artificial expression. In case a pixelated (i.e., scrambled) image was shown, no answers should be given (no-go condition). An illustration of the procedure and example stimuli can be found in Figure 6.1. No information about how many images were manipulated (50%) or hints about how to detect them were given. Feedback about correct and wrong answers was not given throughout the main task, but participants could choose to see their performance in the end. After ensuring that the task was understood, participants started with the experiment. Behavioral data (response times, hit rate) and psychophysiological data (electroencephalography (EEG), eye gaze and pupil size) was collected during the main experiment. A trial started with a black fixation cross presented at the center of the screen for 0.4 - 0.6 sec. (uniformly distributed). Following that, a face or scrambled stimulus was presented for 0.2 sec, which was replaced by a white fixation cross displayed until response, or for 0.2 sec in case of no-go trials. The next trial started after an inter-trial interval of 1.5 - 2.2 sec (uniformly distributed), showing a blank screen. During the main part, after every 100 trials, there was a break for recovery. Before resuming the task, we re-calibrated the eye-tracker (1-point calibration/drift correction). The experiment could be paused and resumed, e.g., to fix noisy electrodes. In the end, participants were debriefed about the main aims and background of the study

doi: 10.53846/goediss-9841

(presented on the computer screen) and could ask the experimenters for additional information.

Figure 6.1
Procedure of the naturalness-classification task of the facial expression.



Notes: After a black fixation cross with variable duration, an individual face or scrambled stimulus is presented in the center of the screen. The different sizes in which stimuli could be presented are indicated as the coloured ovals (only for illustrational purposes). Size ratios between ovals and the box correspond to the presented stimuli and the display size of the monitor. The size of the fixation crosses were increased in this figure for visibility. All exemplary faces show manipulated, i.e., fake expressions.

EEG recording and processing

Electrophysiological data were recorded with a sampling rate of 512 Hz and a bandwidth of 102.4 Hz using a Biosemi ActiveTwoEEG AD-Box with 128 active electrodes (AgAgCl) mounted in an electrode cap (Easy Cap™) and the software ActiView. The arrangement was based on the ABC radial layout electrode positioning. Additionally, six external electrodes were used, one each for the left and right mastoids, one below and next to the left and right eye. The common mode sense (CMS) active electrode and as ground electrode the driven right leg (DLR) passive electrode served as reference electrodes. We determined the time windows and regions of interest (ROIs) electrodes for the ERP components of interest based on previous studies in the lab (Hammerschmidt et al., 2017) which is in line with typical ERP studies on emotional face processing (Schindler & Bublatzky, 2020). We confirmed the time windows and topographical ROIs with pilot data (*N*=4) which was not included in the analysis. The following ERPs for the visual (face-locked) components were extracted: **P1**: Peak amplitudes and peak latency, 80-120 ms; occipital electrode cluster: A8, A9, A10, A15 (O1), A16, A17, A28 (O2), A29, A30, B5, B6, B7; **N170**: Peak amplitudes and peak latency, 130 - 200 ms; occipitotemporal electrode cluster: D32 (P9), A10 (PO7),

doi: 10.53846/goediss-9841

A11, A12, B10 (P10), B7 (PO8), B8, B9; **EPN**: mean amplitudes, 250-300 ms; occipito-temporal cluster: A10 (PO7), A11, A12, A14, A15 (O1), D32 (P9), A24, A25, A26, A27, A28 (O2), B7 (PO8), B8, B9, B10(P10); **LPC**: mean amplitudes, 400 - 600 ms; occipito-parietal electrode cluster: A4, A19 (Pz), A20, A21 (POz), A5, A32, A18, A31, A17 (PO3), A30 (PO4).

Preprocessing. The continuous EEG was pre-processed offline using functions of EEGLAB (v2019.0, Delorme & Makeig, 2004) in MATLAB (2018). Event triggers were shifted by a constant of 26 ms to account for the systematic delay of stimulus appearance on the monitor. We re-referenced data to average, excluding external electrodes. The continuous data was 0.01 Hz high-pass filtered, and 50 Hz line noise was reduced by using "CleanLine" (v1.04, Mullen, 2012), an EEGLAB plugin, to remove sinusoidal noise. Then, data was epoched from -500 to 1000 ms around stimulus onset and corrected to a 200 ms pre-stimulus baseline. For artifact correction, we performed Independent Component Analysis (ICA) on a 1 Hz high-pass filtered version of the data and transferred ICA weights on the original 0.01 Hz filtered data. Independent Components were removed if classified as eye, muscle, or channel noise components with a probability of >90% using "ICLabel" (v1.2.4, Pion-Tonachini et al., 2017). The remaining noisy channels were interpolated. We performed trial-wise rejections of epochs trimmed to -200 to 1000 ms with rejecting amplitudes exceeding -100/100 μV during -200 and 600 ms (avr. 5.3%), steep amplitude changes (>100 μV within the epoch; avr. 5.9%) and improbable activation (deviation >3 of the mean distribution for every time point; 0%). For the final subsample of participants, there was an overall mean rejection rate of 7.8% (range 0.1% - 38.2%) of trials due to these artifact rejection methods. Information about eye blinks was obtained by analysis of the pupil data and independently applied to exclude ERP trials with blinks during baseline and time windows of interest. To monitor the pupil size and gaze, we used a desktop-mounted eye tracker (EyeLink 1000 CL 1 - AAD01, SR Research) and the corresponding software in version 4.56. We measured pupil size and gaze (in pixels) to detect blinks and fixation deviations from the target stimulus and included these measures in exploratory analysis, which is not part of this manuscript.

Statistical analysis

All statistical analysis was conducted in R (v4.0, R Core Team, 2020). To analyze the behavioral and ERP data, we used (generalized) linear mixed models and used the maximum likelihood (ML) estimator to estimate parameters. Separate models for each response variable (AV) were conducted. Statical significance was inferred through likelihood ratio tests (LRT), by testing a model against a reduced model which does not include the predictor of interest. In the case

doi: 10.53846/goediss-9841

of significant LRTs, we reported post-hoc contrasts for the difference between the factor levels. The conventional significance level α = 0.05 (two-sided) was applied, and post-hoc tests were Šid'ak'-corrected to adjust for multiple comparisons. We inspected the models for potential variance inflation and model residuals for potential misspecification of the model. Tables of results for all models, including regression coefficients β, Standard Errors (SE), 95 % nonparametric bootstrapped Confidence Intervals (CI), and stability of the coefficients (leave-one(participant)-out), are included in Appendix D. We preregistered two models for each ERP component of interest: a) a model including stimulus size (with three levels) $\times$ Emotion$_{\text{+Scram}}$ (with four levels: scrambled (reference), happy, angry, neutral) and b) stimulus size (with three levels) $\times$ Face$_{\text{+Scram}}$ (with two levels: scrambled (reference), intact faces). However, as the ERPs of scrambled stimuli were very distinct from all face categories, we dropped this factor in model a) as it is already included in b) and to not confuse a significant effect of the "Emotion$_{\text{+Scram}}$" factor with differences between emotion levels. The originally preregistered model with all levels can be found in Appendix D. Behavioral analysis was only applied to intact stimuli, as scrambled versions were "no-go" conditions. We included correct and incorrect answers and trimmed RTs to an upper threshold of 5000 ms. A skewness-adjusted boxplot method was applied to exclude extreme values separately for every participant and condition (Hubert & Vandervieren, 2008; "adjbox" function from the R package "robustbase," Maechler et al., 2021). RT-model estimation was based on mean responses per condition and participant. We used a linear mixed model to estimate the RT effects of emotion, stimulus size, and their interaction and included a random intercept for participant ID. The analysis of hit rates (accuracy) was exploratory and not a priori preregistered. We conducted a mixed logistic regression with emotion, stimulus size, and their interaction as fixed effects. In addition to the random intercept of participant ID, we included random slopes of emotion and stimulus size to reduce the overdispersion of the model.

6.3 Results

ERP results

Emotion by stimulus size. **P1**: Mean amplitudes were not modulated by emotion ($\chi^2(2)$ = 3.62, *p* = .163) but there was a main effect of stimulus size ($\chi^2(2)$ = 8.83, *p* = .012), with larger mean amplitudes for medium and large faces compared to small faces (diff$_{\text{medium-small}}$ = 0.31, *p* = .028; diff$_{\text{large-small}}$ = 0.30, *p* = .038). Medium and large faces did not differ significantly (diff$_{\text{medium-large}}$ = 0.01, *p* = .999). There was no interaction between emotion and stimulus size ($\chi^2(4)$ = 2.50, *p* = .645). Similarly, peak amplitudes were not modulated by emotion ($\chi^2(2)$ = 4.14, *p* = .126) but

doi: 10.53846/goediss-9841

by stimulus size ($\chi^2(2) = 42.68$, p <.001) with larger mean amplitudes for medium and large faces compared to small faces ($diff_{medium\text{-}small} = 0.65$, p <.001; $diff_{large\text{-}small} = 0.83$, p <.001), similarly to P1 mean amplitudes. Medium and large faces did not differ significantly ($diff_{medium\text{-}large} = -0.18$, $p = .413$) and the interaction between emotion and stimulus size was not significant ($\chi^2(4) = 0.82$, $p = .935$). P1 peak latency showed a main effect of emotion ($\chi^2(2) = 6.94$, $p = .031$). However, adjusted post hoc contrasts showed only a trend for happy faces having shorter latencies compared to neutral and angry faces ($diff_{hap\text{-}ang} = -2.08$, $p = .054$; $diff_{hap\text{-}neu} = -1.89$, $p = .095$). There was a main effect of stimulus size ($\chi^2(2) = 40.24$, p <.001) with shorter latencies for large faces compared to medium ($diff_{large\text{-}medium} = -2.23$, p .035) and shorter latencies for medium compared to small faces ($diff_{medium\text{-}small} = -3.42$, p = <.001). The interaction between emotion and stimulus size was not significant ($\chi^2(4) = 0.52$, $p = .971$).

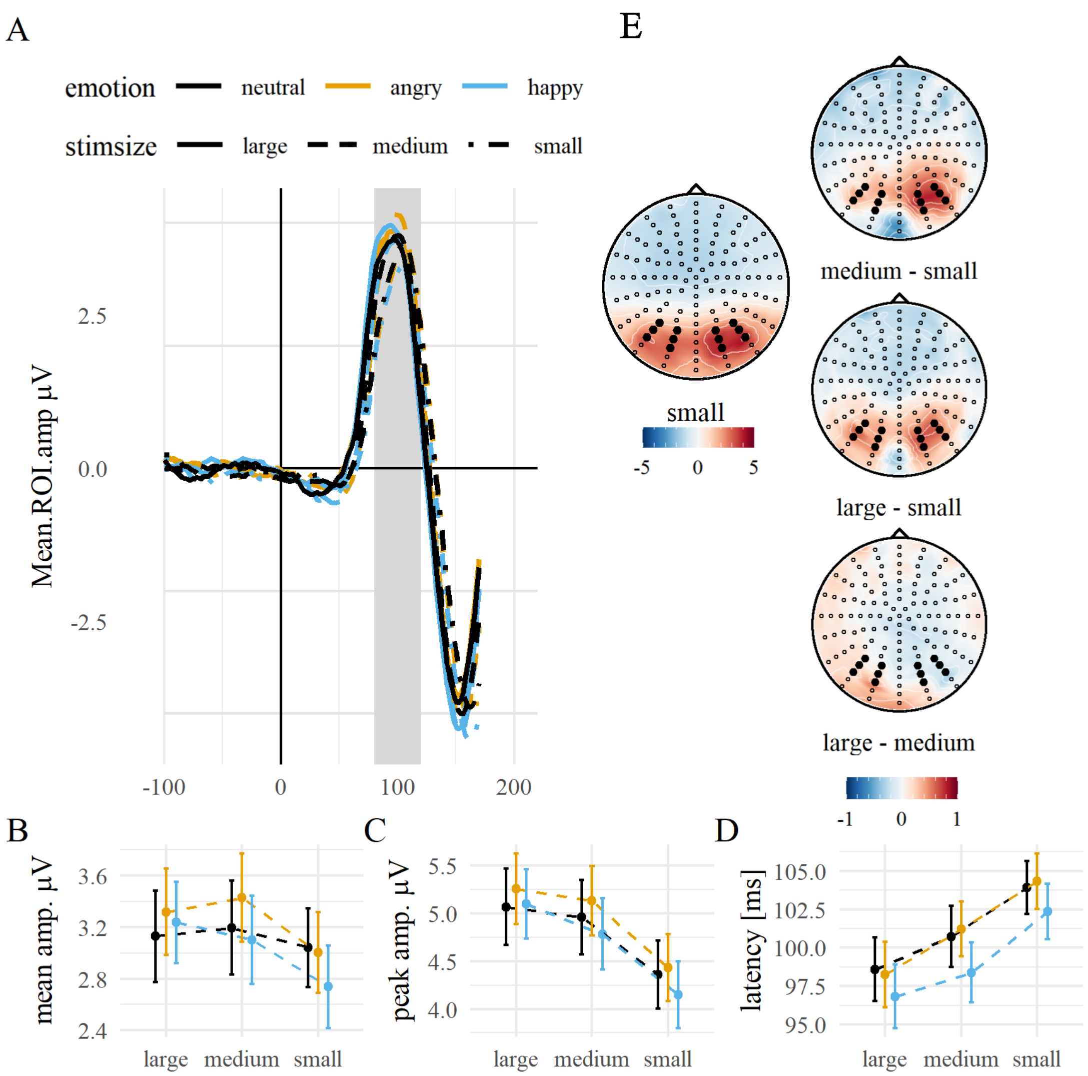
N170: N170 mean amplitudes were modulated by emotion ($\chi^2(2) = 19.78$, p <.001), and by stimulus size ($\chi^2(2) = 40.72$, p <.001), but not by their interaction ($\chi^2(4) = 0.46$, $p = .977$). Across stimulus sizes, happy and neutral faces ($diff_{hap\text{-}neu} = -0.41$, $p = .016$), as well as angry and neutral faces ($diff_{ang\text{-}neu} = -0.65$, p <.001) differed significantly, with the emotional expressions eliciting more pronounced negative amplitudes. Stimulus size was negatively related to amplitude. Small faces elicited more negative amplitudes than medium ($diff_{small\text{-}medium} = -0.52$, $p = .001$), and medium faces than large faces ($diff_{medium\text{-}large} = 0.44$, $p = .010$). Also, peak amplitudes were modulated by emotion ($\chi^2(2) = 30.63$, p <.001), and by stimulus size ($\chi^2(2) = 11.36$, $p = .003$), but not by their interaction ($\chi^2(4) = 0.93$, $p = .920$). Like mean amplitudes, peak amplitudes differed between happy and neutral faces ($diff_{hap\text{-}neu} = -0.46$, p <.001), and angry and neutral faces ($diff_{ang\text{-}neu} = -0.67$, p <.001), with the emotional expressions eliciting pronounced negative amplitudes. Regarding stimulus size, peak amplitudes differed only between medium and small ($diff_{medium\text{-}small} = -0.36$, $p = .010$) and between medium and large faces ($diff_{medium\text{-}large} = -0.35$, $p = .014$) but not between large and small faces ($diff_{large\text{-}small} = -0.02$, $p = .999$). N170 peak latency was not modulated by emotion ($\chi^2(2) = 3.66$, $p = .160$) but there was a main effect of stimulus size ($\chi^2(2) = 337.03$, p <.001) with shorter latencies for large faces compared to medium ($diff_{large\text{-}medium} = -2.41$, p <.001) and shorter latencies for medium compared to small faces ($diff_{medium\text{-}small} = -8.12$, p <.001). The interaction between emotion and stimulus size was not significant ($\chi^2(4) = 2.21$, $p = .697$).

EPN:There was a main effect of emotion on EPN amplitudes ($\chi^2(2) = 29.42$, p <.001), with larger negative amplitudes for happy faces ($diff_{hap\text{-}neu} = -0.69$, p <.001) and angry faces ($diff_{ang\text{-}neu} = -0.54$, p <.001) compared to neutral faces. Amplitudes between happy and angry faces did not differ significantly ($diff_{hap\text{-}ang} = -0.15$, $p = .615$). Also stimulus size modulated EPN amplitudes

doi: 10.53846/goediss-9841

Figure 6.2

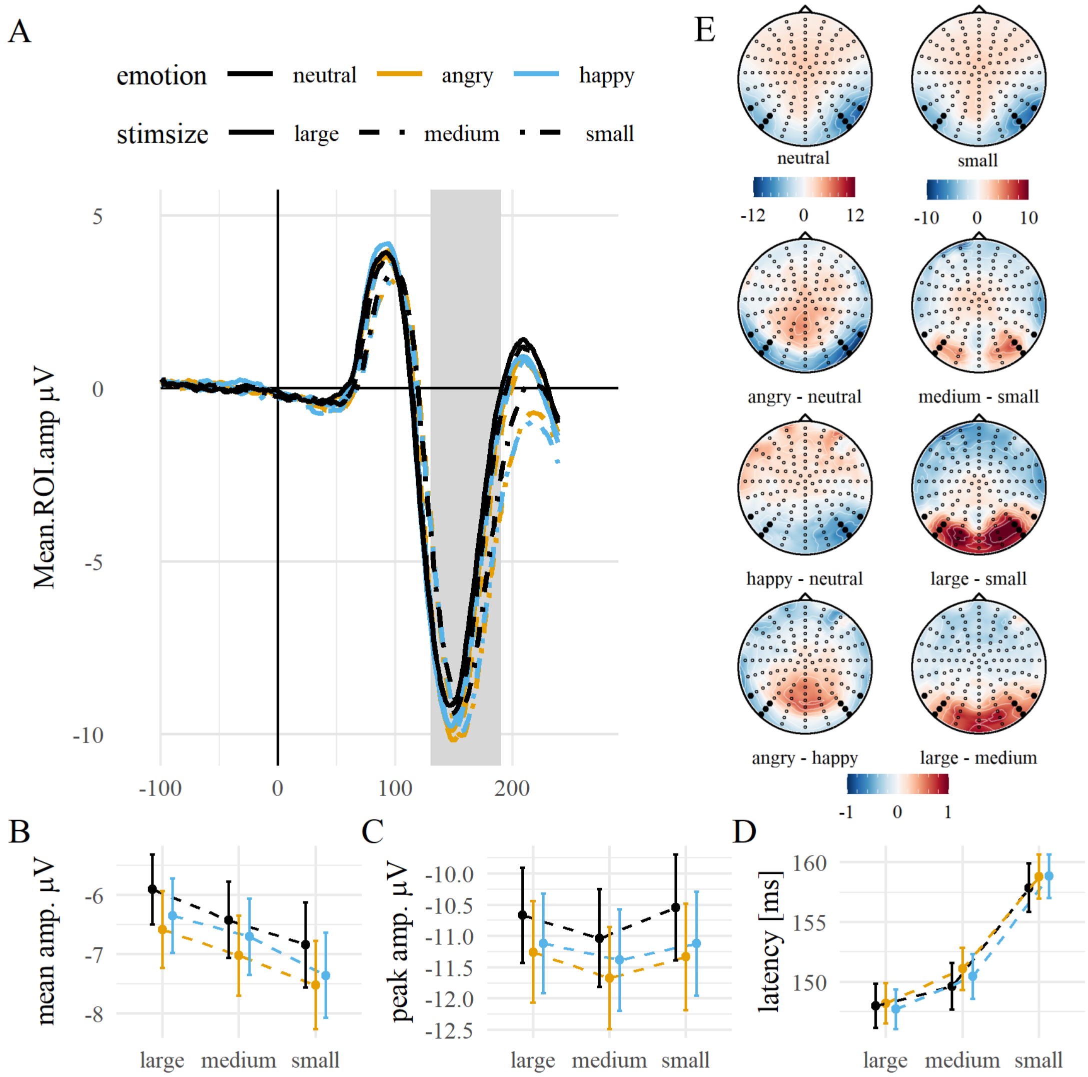
P1 by emotion and stimulus size.



Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand averages of the ROI mean amplitudes and **C** peak amplitudes, and **D** peak latencies, contrasted for stimulus sizes and all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **E** Topographies of the ERP distribution for small faces and pairwise differences of size levels, averaged over emotion levels. The highlighted channels depict the ROI channels.

doi: 10.53846/goediss-9841

Figure 6.3
N170 by emotion and stimulus size.



Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand averages of the ROI mean amplitudes and **C** peak amplitudes, and **D** peak latencies, contrasted for stimulus sizes and all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **E** Topographies of the ERP distributions. The left column shows the main effect of emotion, with the topography for neutral expressions, and the pairwise differences between emotion levels (all averaged over sizes). The right column shows the main effect of size, with the topography of small faces and pairwise differences of size levels (all averaged over emotion levels). The highlighted channels depict the ROI channels.

doi: 10.53846/goediss-9841

($\chi^2(2) = 64.14$, p <.001). Stimulus size was negatively related to EPN amplitudes with the most negative amplitudes for small faces, followed by medium and large stimuli ($diff_{medium\text{-}small} = 0.60$, p <.001; $diff_{large\text{-}medium} = 0.50$, p <.001). However, there was no interaction between emotion and stimulus size ($\chi^2(4) = 5.00$, $p = .287$).

LPC: LPC amplitudes were modulated by emotion ($\chi^2(2) = 17.32$, p <.001) and stimulus size ($\chi^2(2) = 23.90$, p <.001). As for the other ERP components, the interaction between emotion and stimulus size was not significant ($\chi^2(4) = 2.20$, $p = .700$). Happy faces elicited larger amplitudes compared to neutral faces ($diff_{hap\text{-}neu} = 0.41$, p <.001) and angry faces ($diff_{hap\text{-}ang} = 0.32$, $p = .006$). Amplitudes between angry and neutral faces did not differ ($diff_{ang\text{-}neu} = 0.09$, $p = .778$). Amplitudes for small stimuli differed from medium ($diff_{medium\text{-}small} = 0.29$, $p = .016$) and large ($diff_{large\text{-}small} = 0.50$, p <.001), whereas they did not differ between medium and large ($diff_{medium\text{-}large} = -0.21$, $p = .107$).

Face intactness (faces vs. scrambled stimuli) and stimulus size. ERP results of face intactness (faces vs. scrambled stimuli) and stimulus size are shown in Figure 6.6.

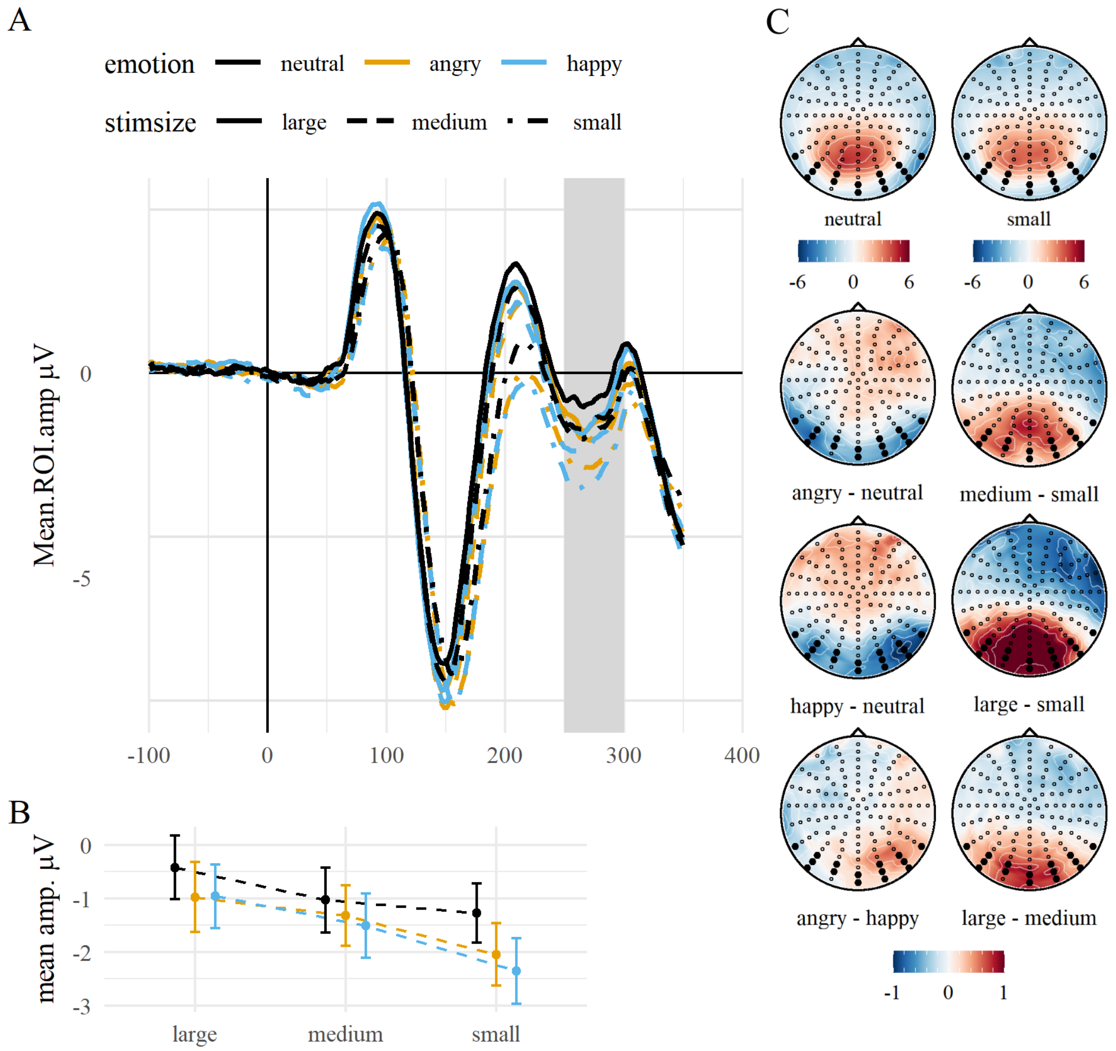
P1: Mean amplitudes differed between faces and scrambled stimuli ($\chi^2(1) = 21.42$, p <.001), with smaller amplitudes for faces ($diff_{face\text{-}scrb} = -0.72$). There was also an effect of stimulus size ($\chi^2(2) = 14.03$, p <.001), with differences between medium and small, large and small, but not between medium and large stimuli ($diff_{medium\text{-}small} = 0.55$, $p = .011$; $diff_{large\text{-}small} = 0.66$, $p = .002$; $diff_{medium\text{-}large} = -0.11$, $p = .916$). The interaction with stimulus size was not significant ($\chi^2(2) = 3.80$, $p = .149$). Peak amplitudes were modulated by intactness of the face ($\chi^2(1) = 27.55$, p <.001), stimulus size ($\chi^2(2) = 44.34$, p <.001) and their interaction ($\chi^2(2) = 6.84$, $p = .033$). Peak amplitudes differed significantly between face and scrambled stimuli of medium sizes ($diff_{m.face\text{-}m.scrb} = -0.88$, $p = .002$) and large sizes ($diff_{l.face\text{-}l.scrb} = -1.37$, p <.001) but not of small sizes ($diff_{s.face\text{-}s.scrb} = -0.34$, $p = .222$). Peak latency showed only a trend modulation of the face intactness ($\chi^2(1) = 2.92$, $p = .087$), with longer latencies for faces compared to scrambled stimuli ($diff_{face\text{-}scrb} = 1.73$). There was a main effect of stimulus size ($\chi^2(2) = 37.11$, p <.001), analogously to the emotion model. No interaction between face intactness and size was present ($\chi^2(2) = 3.26$, $p = .196$).

N170: N170 mean amplitudes were only significantly modulated by intactness of the face ($\chi^2(1) = 302.79$, p <.001), with more pronounced negative amplitudes for intact faces ($diff_{face\text{-}scrb} = -6.89$). Neither stimulus size ($\chi^2(2) = 1.86$, $p = .395$), nor the interaction with intactness was significant ($\chi^2(2) = 3.93$, $p = .140$). However, peak amplitudes were affected both by intactness ($\chi^2(1) = 310.77$, p <.001), with more negative amplitudes for intact faces ($diff_{face\text{-}scrb} = -8.13$), and by stimulus size ($\chi^2(2) = 8.01$, $p = .018$). There was a trend for an interaction ($\chi^2(2) = 5.38$, p =

doi: 10.53846/goediss-9841

Figure 6.4

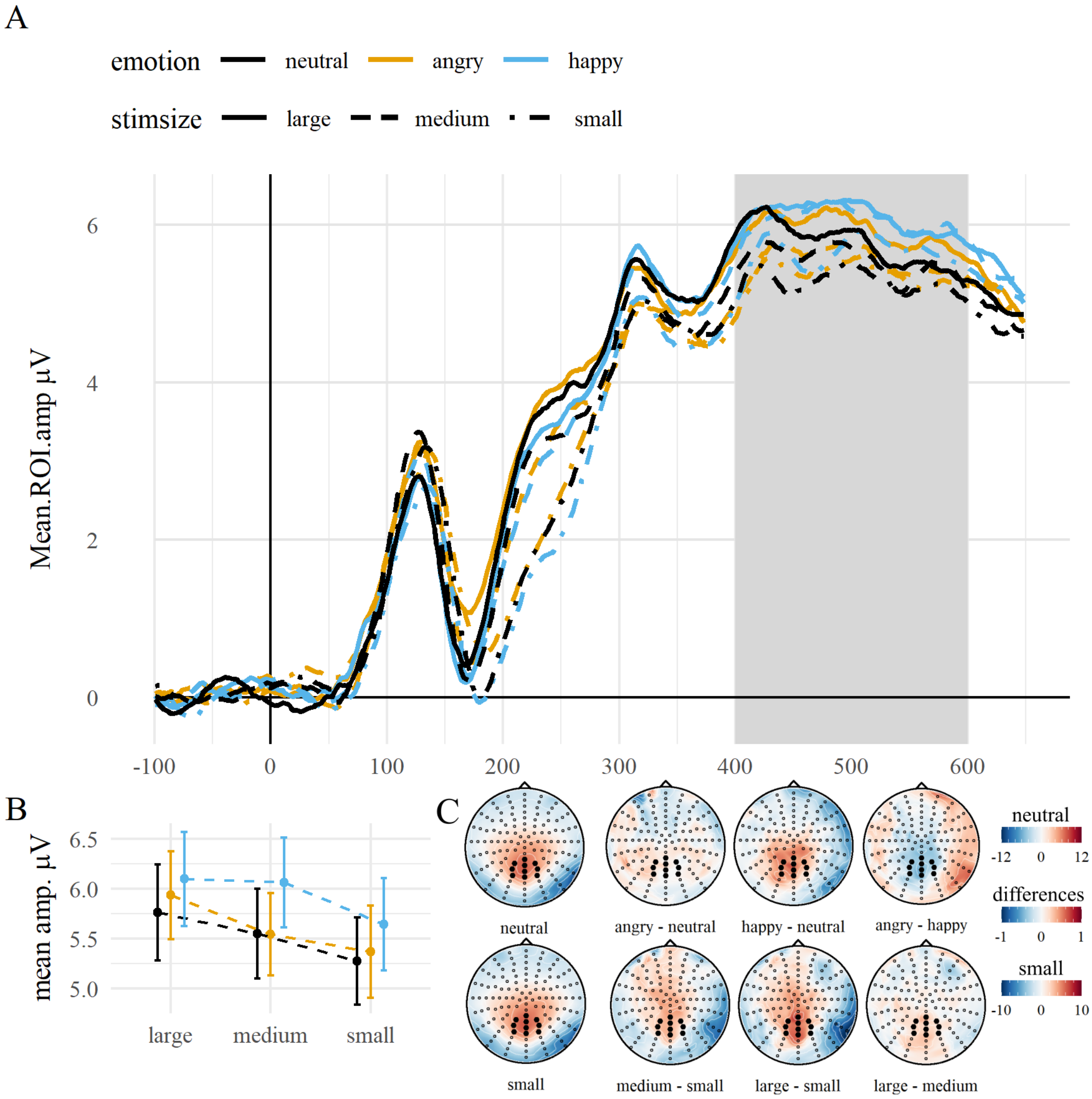
EPN by emotion and stimulus size.



Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand averages of the ROI mean amplitudes contrasted for stimulus sizes and all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **C** Topographies of the ERP distributions. The left column shows the main effect of emotion, with the topography for neutral expressions, and the pairwise differences between emotion levels (all averaged over sizes). The right column shows the main effect of size, with the topography of small faces and pairwise differences of size levels (all averaged over emotion levels). The highlighted channels depict the ROI channels.

doi: 10.53846/goediss-9841

Figure 6.5
LPC by emotion and stimulus size.

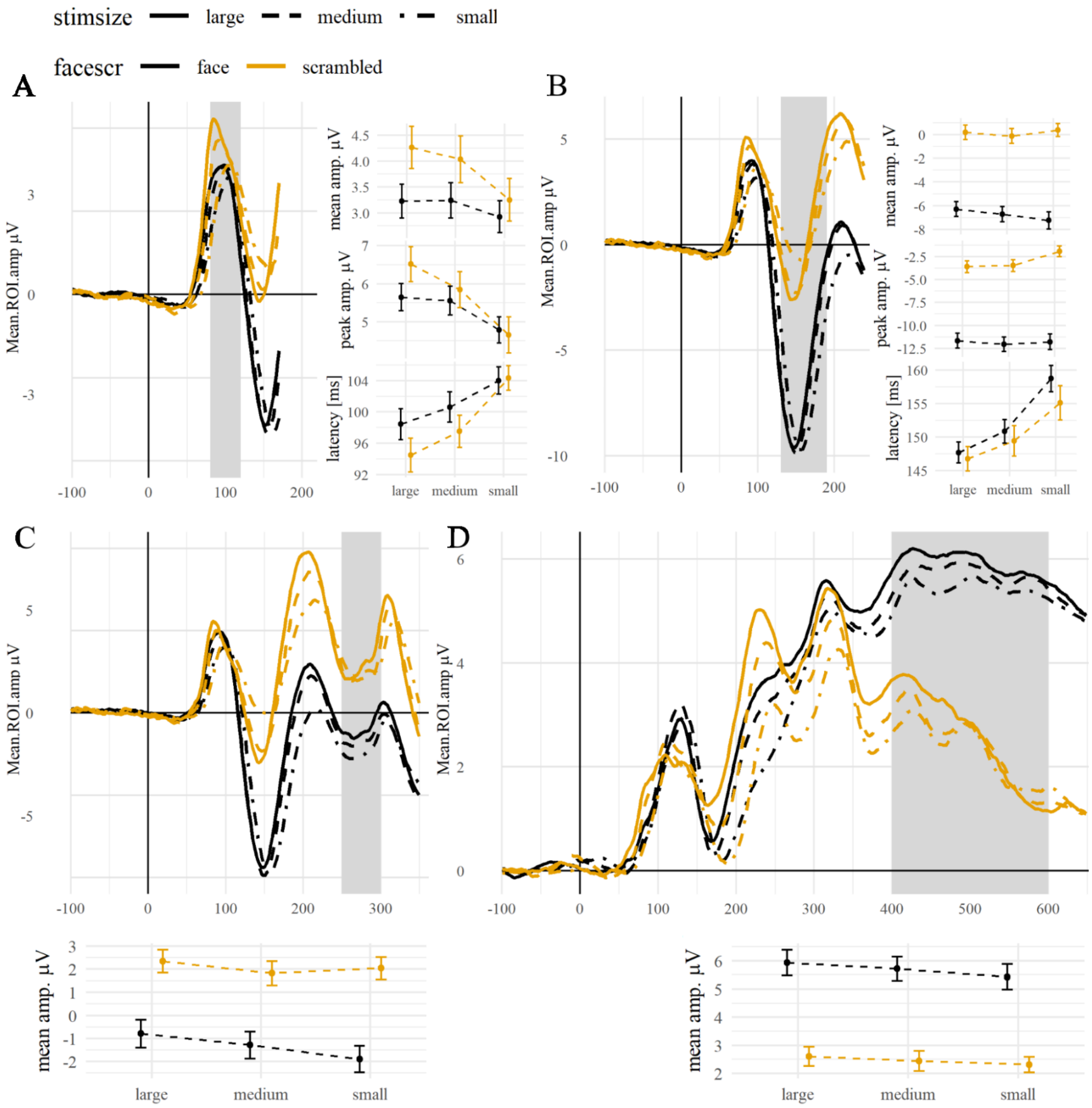


Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand averages of the ROI mean amplitudes contrasted for stimulus sizes and all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **C** Topographies of the ERP distributions. The top row shows the main effect of emotion, with the topography for neutral expressions, and the pairwise differences between emotion levels (all averaged over sizes). The bottom row shows the main effect of size, with the topography of small faces and pairwise differences of size levels (all averaged over emotion levels). The highlighted channels depict the ROI channels.

doi: 10.53846/goediss-9841

Figure 6.6

ERPs (P1, N170, EPN, and LPC) of scrambled vs. intact faces by stimulus size.



Notes: Grand average ERP time series of the averaged ROI channels. The highlighted areas display the respective ROI time window. Dotplots show the grand averages, contrasted for stimulus sizes and all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **A** left: P1 ERP time series, right: grand averages of the ROI mean and peak amplitudes and peak latencies. **B** left: N170 ERP time series, right: grand averages of the ROI mean and peak amplitudes and peak latencies. **C** top: EPN ERP time series, bottom: grand averages of the ROI mean amplitude. **D** top: LPC ERP time series, bottom: grand averages of the ROI mean amplitude.

doi: 10.53846/goediss-9841

.068). Post-hoc contrasts showed that the size effect was only apparent within scrambled stimuli, with differences between medium and small ($\text{diff}_{\text{scrb.medium-small}}$ = -1.54, p = .010), as well as large and small stimuli ($\text{diff}_{\text{scrb.large-small}}$ = -1.67, p = .005) but not between medium and large stimuli ($\text{diff}_{\text{scrb.medium-large}}$ = 0.13, p = .967). Within faces, sizes effects were all insignificant (all p >.05). Similar to P1 peak latencies, N170 peak latencies showed a trend for face intactness ($\chi^2(1)$ = 3.69, p = .055), with longer latencies for intact faces ($\text{diff}_{\text{face-scrb}}$ = 1.36). However, N170 peak latency was affected by stimulus size ($\chi^2(2)$ = 104.57, p <.001), with shorter latencies for large stimuli, compared to medium and medium compared to small stimuli ($\text{diff}_{\text{large-medium}}$ = -2.22, p = .036; $\text{diff}_{\text{medium-small}}$ = -7.50, p <.001; $\text{diff}_{\text{large-small}}$ = -9.72, p <.001;). No interaction between face intactness and size was present ($\chi^2(2)$ = 0.53, p = .766).

EPN: There was a difference between faces and scrambled stimuli for EPN amplitudes ($\chi^2(1)$ = 156.68, p <.001), with more negative amplitudes for faces ($\text{diff}_{\text{face-scrb}}$ = -3.40). Also stimulus size modulated EPN amplitudes ($\chi^2(2)$ = 7.76, p = .021). Stimulus size only differed between large and small stimuli ($\text{diff}_{\text{large-small}}$ = 0.74, p = .022) and there was no interaction between intactness and stimulus size ($\chi^2(2)$ = 2.52, p = .283).

LPC: LPC amplitudes were only affected by intactness of the face ($\chi^2(1)$ = 184.64, p <.001) with larger amplitudes for faces compared to scrambled stimuli ($\text{diff}_{\text{face-scrb}}$ = 3.24). Neither stimulus size ($\chi^2(2)$ = 2.94, p = .230) nor the interaction between intactness and stimulus size were significant ($\chi^2(2)$ = 0.25, p = .881).

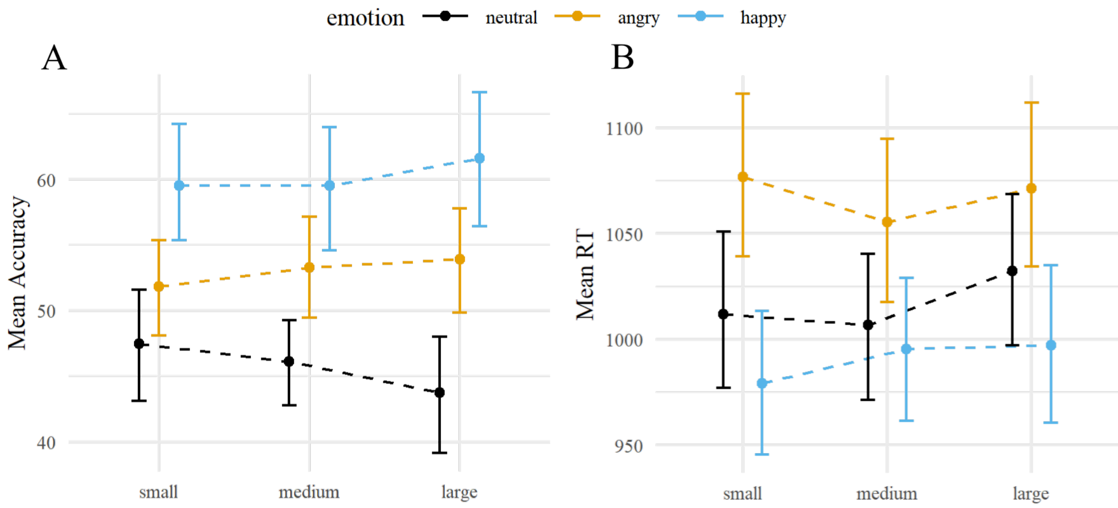
Behavioral outcomes

Descriptively, accuracy was highest for happy expressions, followed by angry and neutral expressions. Whereas for angry and happy expressions, accuracy increased with stimulus size, for neutral expressions, it decreased. Responses were descriptively slowest for angry and fastest for happy expressions. Small stimuli showed the largest variability in RTs between emotion levels. Figure 6.7 shows accuracy and RT results of the naturalness classification task.

Accuracy. There was a main effect of emotion ($\chi^2(2)$ = 26.90, p <.001) on accuracy, with higher accuracy for happy compared to angry faces ($OR_{\text{hap/ang}}$ = 1.39, p = <.001), and angry compared to neutral faces ($OR_{\text{ang/neu}}$ = 1.36, p = <.001). Stimulus size did not affect accuracy ($\chi^2(2)$ = 0.13, p = .938), but there was a significant emotion $\times$ stimulus size interaction ($\chi^2(4)$ = 16.17, p = .003). Within a stimulus size, accuracy differences were significant between all emotion levels but it reached only a trend for the difference between small angry and small neutral faces ($OR_{\text{s.ang/s.neu}}$

doi: 10.53846/goediss-9841

Figure 6.7
Accuracy and response times in the naturalness classification task.



Notes: **A** shows the average accuracy in percent and **B** the response times in ms. Mean values were averaged per emotion and stimulus size over participants. Errorbars depict 95% non-parametric bootstrapped Confidence Intervals.

= 1.2, p = .096).

Response times. Response times were modulated by emotion ($\chi^2(2) = 85.57$, p <.001) with happy faces being responded to fastest (est_{hap} = 1002 ms), followed by neutral (est_{neu} = 1032 ms) and angry faces (est_{ang} = 1090 ms). All pairwise comparisons between emotion levels were significant on $p < .01$ (adjusted) . There was also an effect of stimulus size ($\chi^2(2) = 7.37$, $p = .025$) with large stimuli being significantly slower than medium-sized stimuli ($diff_{large\text{-}medium}$ = 24.14, p = .025), but not between large and small ($diff_{large\text{-}small}$ = 15.79, p = .230) or medium and small ($diff_{medium\text{-}small}$ = -8.35, p = .737). The interaction between stimulus size and emotion was not significant ($\chi^2(4) = 5.25$, $p = .263$).

6.4 Discussion

This study investigated the potential influences of stimulus size on the perception of emotional expressions in faces. We presented a large set of faces with angry, happy, or neutral expressions in three different sizes and tested size, emotion, and interaction effects systematically on early, mid-latency, and long-latency event-related potentials (ERPs). The experimental task was to decide about the naturalness of briefly presented faces, of which the expression could be real or manipulated. We expected general effects of size and emotion on ERP components but were specifically interested in whether emotion effects would be enhanced with increasing face

doi: 10.53846/goediss-9841

sizes. We found greater differences in accuracy among emotion levels for large images, suggesting, indeed, an interaction of facial expression and stimulus size on how accurate real and artificial expressions could be detected. However, electrophysiological responses suggested an overall additive and not interactive effect of emotion and stimulus size. Whereas stimulus size affected all ERP components, effects of emotion were observable for the N170, Early Posterior Negativity (EPN), and Late Positive Complex (LPC). Importantly, for none of the ERP components, the interaction between emotion and stimulus size was significant, at variance with the findings of Bayer et al. (2012) and Cesarei & Codispoti (2006). However, when aggregating intact faces and comparing them with scrambled stimuli, early visual processing was differently affected by stimulus size, suggesting distinct effects of size on (specific) low-level stimulus properties.

Effects of emotional expression and stimulus size

As we hypothesized, P1 amplitudes were affected by stimulus size, with larger amplitudes and shorter latencies for larger stimuli, as in (Busch et al., 2004; Kornmeier et al., 2011). However, emotional expressions of the face did not affect P1 amplitudes, although there was a trend to shorter peak latencies for faces with happy expressions. The earliest significant emotion effect in our study was apparent in the N170 component. N170 peak amplitudes were both affected by emotion and size, but there was no interaction between both. In accordance with our hypotheses, negative expressions, and to a lesser degree also positive expressions, elicited enhanced negative amplitudes compared to neutral expressions, adding evidence that facial expressions of anger and threat-related stimuli (see, e.g., Hammerschmidt, Kagan, et al., 2018; Schindler & Bublatzky, 2020) and to a lesser degree for happy faces (for a review, see Hinojosa et al., 2015) affect the N170. As N170 peak latencies differed between sizes, possibly as a continuation of earlier processing, the effect of size on mean amplitudes should be interpreted accordingly. Noticeably, N170 peak amplitudes were most pronounced for medium-sized faces across expression levels, suggesting size-invariance of emotion effects of the N170 but not size-invariance of the N170 in general. Rolls and Baylis (1986) reported size-invariance for the majority of face-selective neurons in the macaques' superior temporal sulcus, but a minority of neurons responded distinctively to retinal angle and absolute size, for which the authors suggested both contribute to size-invariant recognition of faces, including descriptors for plausible absolute sizes of faces.

As hypothesized, emotion effects also extended to the EPN component (for a discussion on the functional distinction between the N170 and EPN, see Rellecke, 2012). However, the EPN was modulated by both happy and angry faces (e.g., Hammerschmidt, Kagan, et al., 2018; Sommer et al., 2013; Wronka & Walentowska, 2011) and did not differ between emotional expressions (Recio

doi: 10.53846/goediss-9841

et al., 2011; cf. Valdés-Conroy et al., 2014). Contrary to our prediction, size did not modulate emotion effects of the EPN, but there was a general effect of size on the EPN. Noticeably, small stimuli resulted in enhanced negative amplitudes (cf. Bayer et al., 2012), and descriptively, also the difference between emotional and neutral expressions was greatest within small stimuli, although the interaction was not significant. This was opposite to our prediction about the potential relevance effect transmitted through size and requires further investigation, particularly because the effects of size were consistent with our predictions for the other ERPs. In alignment with our predictions, the LPC was modulated by stimulus size and emotion but not their interaction (similar to Bayer et al., 2012; Cesarei & Codispoti, 2006). Medium and large stimuli elicited more pronounced LPC amplitudes than small stimuli when testing across emotion levels, indicating sustained attention as a function of stimulus size. LPC amplitudes were the largest for happy expressions across all stimulus sizes. Descriptively, also amplitudes for angry faces were larger than for neutral faces. However, the difference between neutral and angry faces was not significant, contradicting, in particular, our prediction about a general negativity bias for anger stimuli (Schacht & Sommer, 2009; Schupp et al., 2004).

Effects of face intactness and stimulus size

We included scrambled versions of faces as control stimuli to estimate whether stimulus size would differentially affect stimuli of the same retinal size but a different configuration. Throughout processing, ERPs differed between scrambled and intact faces. P1 mean and peak amplitudes were larger for scrambled stimuli (see Gliga & Dehaene-Lambertz, 2005; Schindler, Tirloni, et al., 2021; cf. Herrmann et al., 2004; Schindler, Bruchmann, Gathmann, et al., 2021; Zion-Golumbic & Bentin, 2006). Probably, this effect occurred due to a reduced perceivable contrast of stimuli due to the scrambling method. However, we also measured a small but systematic difference in luminance for scrambled images (see Figure D2), which might have contributed to increased P1 amplitudes (Johannes et al., 1995). Remarkably, P1 peak amplitudes showed an interaction between stimulus size and face intactness, such that only medium and large scrambled stimuli and faces differed, but not small versions. A similar effect was found for peak latencies, which were shortest for large scrambled stimuli and longest for small scrambled and intact stimuli. The random allocation of pixel chunks in scrambled stimuli reduced particularly low spatial frequencies but increased medium frequencies (see Figure D3). Although both low and high spatial frequencies have critical functions in face perception and were shown to be selectively used depending on the current task (Cesarei & Codispoti ((2013), especially low frequencies have been suspected to affect coarse and fast initial visual processing (e.g., Bar et al., 2006; Willenbockel et al., 2012). As edge contrasts

doi: 10.53846/goediss-9841

were relatively well preserved in small faces, the low contrast of scrambled stimuli would make them blend more into the background (individual dark and bright pixels would be perceived as grey). The well-replicated face-sensitivity effect on the N170 (e.g., Bentin et al., 1996; Herrmann et al., 2004) was also corroborated by our results that showed enhanced negative amplitudes for intact faces compared to scrambled stimuli. Also, the EPN differed between scrambled and intact stimuli, possibly as attention to scrambled faces should drastically decrease as soon as they are identified as such, as no decision or response was required. The influence of size on the EPN was similar between intact and scrambled faces, and thus potentially still being affected by low-level stimulus features (e.g., luminance). Whether the (insignificant) larger effect of size for intact faces reflected attentional effects of the face or task-specific relevance is unclear. Similar to the findings of the EPN. Also, the effects of size on the LPC were diminished when collapsing face stimuli and contrasting them against scrambled stimuli. Inspecting the ERP time series, the size effect appeared to be most pronounced for all stimuli (before and) at the beginning of the LPC window and to decrease over time.

Our findings on emotion-related and face-intactness effects should not be seen independently of the particular task participants performed on the faces. We chose the naturalness-decision task to motivate the processing of the facial expressions without explicitly deciding about their valence. Since emotion effects on later ERPs have been shown to be task-sensitive (e.g., Rellecke et al., 2012a), we were interested in detecting emotion effects also when not performing a typical valence-classification of the expression but still aiming for deeper processing compared to tasks like passive-viewing or gender-decisions. Behavioral results of the naturalness classification task revealed both emotional expression and stimulus size effects on response times, as well as an interaction effect of emotion and stimulus size on accuracy. As expected, participants were overall more accurate when deciding about the artificiality of happy expressions, and their decisions were also faster than for the other emotion categories, suggesting an overall facilitated information processing of relevant facial characteristics, which possibly was also reflected by the LPC amplitudes. Furthermore, angry stimuli slowed response times, as similarly shown in previous research for negative or threat-related information where the valence detection was not task-relevant (e.g., Cesarei & Codispoti, 2006; Eastwood et al., 2003; Schacht & Sommer, 2009; Sommer et al., 2013). However, it is possible that the slowing of responses was at least partly caused by decision difficulty. When accounting for both accuracy and response times, angry faces were responded to slowest, but importantly, accuracy was only slightly above chance level, indicating that participants were unsure about whether angry expressions were real or artificial. Interestingly, neutral expressions, for which the decision should have been rather difficult as their difference from the original was

doi: 10.53846/goediss-9841

marginal, overall resulting in below chance accuracy levels.

Stimulus size overall did not affect accuracy, but there was an effect on RTs with slower responses for large compared to medium-sized stimuli. On average, response times were relatively slow compared to two-forced choice tasks, e.g., gender or valence classification (e.g., Hammerschmidt et al., 2017; Rellecke et al., 2012a). Possibly, participants focussed rather on accuracy than on speed, trying to detect fake expressions. However, we cannot differentiate whether participants actually *made use* of the higher resolution of facial details in larger stimuli and intentionally processed the faces longer or whether expressions in small faces were processed more efficiently due foveal presentation of relevant facial features. There might be an optimal retinal stimulus size for information extraction in experimental tasks. For example, Busch et al. (2004) reported the fastest responses to medium-sized stimuli. However, we suspect that speed measures for faces might be less affected by size due to the relatively high size-invariance for faces compared to less familiar objects. Nevertheless, our results show that the discriminability of facial expressions was still sufficient also in small faces.

Because the individual effects of the actual and perceived artificiality of faces (see, e.g., Tauscher et al., 2021; Tucciarelli et al., 2022) is beyond the focus of this study but a time-relevant and interesting research question itself, we planned to investigate and discuss them separately. However, we will briefly discuss potentially relevant factors that might have contributed to the present findings: Real and manipulated happy (more than angry and neutral) faces differed in several dimensions, which participants might have learned to use as discriminative cues over the course of the experiment. Among these potential cues for being a real face were a) the overall variability of expression intensities, b) the variability of the mouth region (particularly of teeth, as the GAN had no information about how the person's teeth actually look), and eye region with stronger activation of the orbicularis oculi (cheek raiser) affecting the perceived intensity of the smile (Gunnery & Ruben, 2015). Expression-unrelated cues were the illumination of the face due to the different lighting conditions between databases that partly remained in the processed stimuli. Learning about these cues might have triggered focusing on specific facial features, potentially biasing not only behavioral responses. For example, LPC effects might have been partly caused by different processing demands and motivation to integrate indicative facial cues, which was probably easier for happy and more difficult for neutral expressions.

doi: 10.53846/goediss-9841

Summary and Outlook

We found additive effects of emotion and stimulus size in face-sensitive ERP components, suggesting that the differential processing of facial expressions of emotion is relatively size-invariant. Both the face-sensitive N170 and EPN were enhanced for happy and angry compared to neutral expressions. LPC amplitudes and the accuracy of naturalness classification were greatest for happy faces. Moreover, stimulus size modulated all ERP amplitudes, with more substantial effects on early components. However, unlike what studies on interactive effects of stimulus size and emotional words or scenes predicted, we did not find emotion effects moderated by stimulus size. At least within the range of sizes, our findings suggest that discriminability between facial expressions was sufficient to produce typical emotion effects. The inclusion of extremely large stimuli might lead to higher arousal and stronger emotion effects. However, side effects such as eye movements are likely to happen and require different, sensible methods. In our study, there were stronger size effects on scrambled compared to intact faces. The specificity of the relative size-invariance for faces resembles an interesting target for future research. Moreover, another approach to test the interactive effects of emotion and physical size might induce perceived physical proximity through (additional) contextual cues and not exclusively by retinal size. Lastly, although stimuli were homogenized in some of their low-level features, physical differences in emotional expressions are inherent to the stimulus (equal stimuli can not simultaneously express different emotions). However, faces have been shown to elicit emotion-based effects when associated with hedonic stimuli. Full control over physical stimulus features between valence categories could be achieved by presenting associated faces in different sizes.

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Chapter 7

General Discussion

Psychological research has ever since its beginning aimed at understanding the interplay between inherent and learned processes, especially in the context of the perception of emotional stimuli (Cardinal et al., 2002; Gore et al., 2015). Seemingly inconsistent findings on the prioritization of emotional cues in social stimuli such as faces and voices (at what processing stage, for which stimuli/features) have motivated, and will continue to motivate, further research under which boundary conditions effects of prioritization occur. With this dissertation, I aimed at contributing to the understanding of how emotional relevance shapes learning and retrieval processes in the perception of two social cues: faces and voices. The studies included in this thesis particularly targeted the questions of whether affective information would be preferentially selected, and how stimulus-external factors, such as motivation and current goals, influence the learning and retrieval processes of cross-modally associated faces. In the following discussion, I first present the extent to which these studies provide evidence for the preferential processing of inherently affective stimuli, such as affect bursts, which served as unconditioned stimuli in *Studies 1, 2,* and *3*, and emotional expressions of faces in *Study 5*. In the second section, I review the results on the acquisition and extinction of valence-based associations and how the task might have influenced their processing and learning (*Studies 1, 2,* and *3*). In the third section, I present open questions and suggestions for further research.

The automaticity of processing inherent emotional social stimuli

Several studies have shown differential processing for emotional and neutral stimuli, in visual search tasks (e.g., Becker et al., 2017), concurrent tasks with emotional distractors (e.g., Deweese et al., 2016), and even emotion-implicit tasks (e.g., Gutiérrez-Cobo et al., 2019; Rellecke et al., 2011; Valdés-Conroy et al., 2014; for a review, see Carretié, 2014). Early emotion effects were

doi: 10.53846/goediss-9841

explained by reciprocal projections of sensory regions and the amygdala, such that sensory regions can be influenced by the amygdala before a cortical representation is completed (LeDoux, 2000; Vuilleumier et al., 2004). Similar activation of the superior temporal sulcus and amygdala for emotion-explicit and -implicit tasks suggests that processing of emotion in the auditory domain might occur involuntarily to some degree (Grandjean et al., 2005; Sander et al., 2005). Moreover, effects of emotion were also found in cross-modal priming (Doi & Shinohara, 2013) and cross-modal binding studies (Maiworm et al., 2012), which speaks to a rather seamless integration of emotional information from different modalities. Whereas the majority of studies show that emotional stimuli *can* be prioritized under different conditions, there is also evidence that the effects of emotion are not immune against current task demands and perceptual load (for a review, see Straube et al., 2011). Several studies have been aimed at approaching the threshold for emotion effects with different tasks and types of distractors (e.g., Lim et al., 2008; Mitchell et al., 2007; Pessoa, McKenna, et al., 2002; Puls & Rothermund, 2017; Rellecke et al., 2012a; Tannert & Rothermund, 2020; Victeur et al., 2019) and have tested modality-specific attention effects. For example, emotion presented in the auditory modality has been shown to be particularly affected by cross-modally presented distractors (Johnson & Zatorre, 2006), although intra-modal competition seems generally stronger than cross-modal competition (Zeelenberg & Bocanegra, 2010).

Given the mixed findings on automatic processing in the presence of distracting stimuli or distracting tasks, in our studies we either circumvented or explicitly addressed this issue by choosing an experimental design, in which all stimuli were consistently relevant for the task (i.e., there were no intramodal or crossmodal distractor stimuli included). Instead, we manipulated which property of the stimulus would be task-relevant. In *Study 1*, we applied an emotion-implicit associative learning paradigm in which participants had to decide about the gender-congruence of face-voice pairs. The conditioning procedure encompassed perfect contingency between a specific face, voice, and correct key to answer. Thus, with learning, faces gained the predictive value of the answer (gender-matching/mismatching) which made voices gradually obsolete for this decision. However, by including a go-no/go condition, the processing of the voice was necessary to decide whether to execute or inhibit a response. During both decisions, the gender matching and response execution can be considered as emotion-implicit. Yet, the emotion of the bursts modulated the auditory P2 component (similar to T. Liu, Pinheiro, Zhao, et al., 2012), suggesting an early relevance detection of the burst's valence (T. Liu, Pinheiro, Deng, et al., 2012; Paulmann et al., 2013; Sauter & Eimer, 2010; Schirmer et al., 2012; Spreckelmeyer et al., 2009). In contrast, the auditory N1 was not significantly modulated by emotion, indicating that not only low-level stimulus differences (Pell et al., 2015) elicited the P2 effect of valence. Moreover, pupil dilation during

doi: 10.53846/goediss-9841

learning was larger for neutral (mismatching) face-voice pairs compared to angry face-voice pairs (cf., Cosme et al., 2021). Potential explanations for the (unexpected) increase in pupil size for the gender-mismatching face-voice pairs will be discussed separately below. The emotional valence of the auditory stimuli also influenced the task performance; differences in response times between emotional and neutral face-voice pairs were present in *Study 1* and replicated in a larger sample as well as in an independent experiment (*Study 2*). In both studies, happy and angry affect bursts led to slower responses, suggesting a distraction or impairment of the processing of the task-relevant information (similar to Schacht & Sommer, 2009).

Taken together, the findings of *Studies 1 and 2* suggest that affect bursts attracted attention independent of the current task goals, i.e., the gender-matching task of the face-voice pairs did not inhibit emotional processing. To investigate whether the affective processing of the auditory bursts would be sufficient for the faces' acquisition of emotional valence, we tested whether responses to the conditioned faces would show modulations of associated valence in response times in *Study 2*, and in face-locked event-related potentials (ERPs) and pupil size both during the acquisition and extinction phases in *Study 1* and during the extinction phase in *Study 3*.

The automaticity of learning and retrieval of valence associations in social stimuli

The processing of inherent emotional cues can be affected by current task goals. But how much attention towards the conditioned and unconditioned stimulus (features) is needed to *associate* valence to a second stimulus? Studies on implicit conditioning addressed the question of whether stimulus awareness is even necessary to form associations between the unconditioned stimulus (US) and the conditioned stimulus (CS) (e.g., Balderston et al., 2014; Balderston & Helmstetter, 2010; Schultz & Helmstetter, 2010). In a fear-conditioning study by Balderston et al. (2014), masked faces (CS) were associated with electric shock in a trace conditioning paradigm (i.e., without timely overlap between the presentation of the face and the shock). The authors argued that the amygdala activity triggered by faces might have supported learning, suggesting the facilitating role of biologically relevant stimuli like faces for learning. Another debated factor in the literature is contingency awareness, with evidence against the necessity (Brockelmann et al., 2011; Houwer et al., 1997; Junghöfer et al., 2016; Roesmann et al., 2020; Schultz & Helmstetter, 2010; Steinberg et al., 2012; Steinberg, Bröckelmann, Rehbein, et al., 2013) and evidence speaking against conditioning when fully unaware of the CS-US contingency (Dawson et al., 2007; Hur, Iordan, Berenbaum, et al., 2016; Marcos & Marcos, 2021; for a review, see Mertens & Engelhard, 2020). Other studies reported associated valence effects when implementing tasks in which the valence of the US was irrelevant for learning (Abdel Rahman, 2011; Hammerschmidt, Kagan, et

doi: 10.53846/goediss-9841

al., 2018) or retrieval (Bruchmann et al., 2021; Hammerschmidt et al., 2017; Luo et al., 2016; Pooresmaeili et al., 2014; Rossi et al., 2017). However, it was not only the type of hedonic stimuli (e.g., monetary reward, person knowledge) but also the temporal course of valence effects that differed between these studies (early vs. later processing).

In *Studies 1 and 2*, we investigated whether faces would acquire additional relevance when conditioned with affect bursts of which the valence was not task-relevant. In both studies we found effects of emotion for affect burst despite them being task-irrelevant. Little evidence in favor of automatic valence associations to the faces was shown by the electrophysiological results of *Study 1*. Neither during learning nor extinction were ERPs modulated by emotion, except for a significant interaction between emotion and congruence of the N170 mean (but not peak) amplitudes in the test session (amplitudes were smaller for neutral mismatching faces, whereas the other conditions showed similar values). This was in contrast to our prediction since we expected effects of associated valence to occur either independently of the conditioned gender-congruence, or, in the case of an interaction effect, to be more pronounced in the gender-matching condition, since there would be less interference by gender-incongruence. A critical question to address is whether our conditioning paradigm might have failed. Did participants not acquire associations between the faces and voices?

The overall results from *Studies 1 and 2* provide strong evidence against this notion. Participants did not only attend to the faces and the voices but also learned about their contingencies. During learning in *Study 1*, conditioned gender-congruence modulated the face-locked Early Posterior Negativity (EPN) and Late Positive Complex (LPC), and there was a trend for a P1 modulation by gender-congruence during learning. Remarkably, effects of conditioned gender-congruence on the P1 and EPN were also present during extinction, i.e., on the subsequent day, where only the face's gender was task-relevant and no sounds were played. Moreover, response times of the test session in Exp.2 of *Study 2* indicated a behavioral advantage for previously gender-congruent faces. The increased allocation of attention to gender-related cues would fit into the framework of attentional effects driven by selection history (B. A. Anderson et al., 2021; Failing & Theeuwes, 2017) and perceptual learning theories (Seitz & Watanabe, 2005). Although gender-related cues were elements of the presented visual and auditory stimuli (which also entailed valence information), selection history might have enhanced the discriminatory control over gender-related cues within the stimulus.

Assuming that this was true, further explanation for the differential processing of gender-congruent and -incongruent information is required. In Studies 1 and 2, gender-congruent and -incongruent face-voice pairs were presented equally often, thus an effect of the probability of

doi: 10.53846/goediss-9841

occurrence can be excluded. Nevertheless, behaviorally, gender-matching face-voice pairs might have resembled the “default” option in the gender-matching task of the face-voice pairs, and thus responses for “matching” were overall faster compared to mismatching. Studies have shown facilitated processing of congruent information (e.g., Laeng et al., 2010; Sim et al., 2020) which is corroborated by the observation that not only gender-congruence but also the emotional congruence of neutral faces and neutral voices facilitated processing during learning, reflected by the fastest responses for neutral gender-matching face-voice pairs. A further possibility is that gender-mismatching pairs have not only introduced general interference, but also uncertainty about the face’s gender, changing the internal representation of the face. This second option might be supported by the effects in pupil dilation, which was enhanced for neutral gender-mismatching face-voice pairs. An increase in pupil size was reported for changes to a different representation of a bistable image, such as a Necker cube (Einhäuser et al., 2008). If gender information was detected best in neutral face-voice pairs, gender-mismatching voices should produce the strongest representational changes.

Potential explanations for the null effects of associated valence are as follows. First, the *emotional valence* of the affect bursts was not associated to the face because emotional cues did not attract *enough* attention. It is possible that while emotional cues sufficed for the momentary processing of the emotion in the voice, by directing the participant’s attention to the gender information, gender-informative cues might have overshadowed the emotional cues in the cross-modal associations. Stimulus salience is a crucial factor for associative learning (e.g., Bevins, 1997; Bradfield & McNally, 2008; Odling-Smee, 1975). However, participants might have habituated faster to the affect bursts (irrespective of their emotional quality) than they would have to more potent and arousing US, such as electric shock (Glenn et al., 2012) or aversive white-noise bursts (Sperl et al., 2016). The regulating function of the amygdala on the cortex facilitates highly danger-signaling stimuli even at less attended spatial locations (J. Armony, 2002; J. L. Armony et al., 1997) or possibly also in competing tasks. Thus, it is likely that effects of associated valence would be found with the same paradigm but different stimulus material, such as panicking screams (e.g., Bruchmann et al., 2021), making the affective quality of the US relatively more salient over the gender information. However, highly aversive US are targeting different types of social affective stimuli, which are not directly comparable to angry or happy bursts that resemble stimuli that people encounter on a daily basis.

Second, participants might have learned about the emotional valence of the stimulus but our selected measures did not capture it. The processing of emotional and social stimuli has been related to a distributed set of brain regions (e.g., Haxby et al., 2000; Ishai, 2008). Although we

doi: 10.53846/goediss-9841

based our focus of investigation on the most frequently investigated and reported ERPs evoked by visually presented affective scenes and faces (Schindler & Bublatzky, 2020; Schupp et al., 2006), there is a high chance that effects of emotion effects were expressed in other areas which were not captured by the ERPs selected in this project. Nevertheless, with our task and our measures, typical emotion-related ERPs in face processing were more affected by the task-relevant gender-congruence condition than by the emotional valence of the voice, suggesting a non-negligible influence of task-relevance on face processing during learning and extinction.

Third, participants might have learned about the emotional valence of the stimulus but the associated effects required tasks that activate valence-related memory traces to become measurable. A dissociation between learning and performance has been addressed in the comparator hypothesis (e.g., Blaisdell et al., 1999; Matzel et al., 1985; McConnell et al., 2010), in which the recovery of a blocked or overshadowed stimulus can retrospectively be achieved by extinguishing the blocking or overshadowing stimulus. In a similar way, we tested whether a specific task during retrieval would uncover effects of associated valence, although further excitatory learning about the stimulus is not possible, i.e., associations would have to be already acquired. To this end, we conducted *Study 3,* which also served to exclude the alternative explanation that happy and angry affect bursts were generally too weak for association and retrieval. In this study, we employed a different learning paradigm in which participants had to study face-voice pairs with the aim of being able to match the faces and voices. Although instructions did not prompt participants to focus on specific stimulus features, they reported categorizing the stimuli into separate gender and emotion levels to better discriminate the face-voice pairs. Moreover, they acquired explicit knowledge about the valence associations, although not instructed to, reflected by the high accuracy in the valence-classification task of the test sessions, in which participants had to retrieve the valence of the previously associated voices. The valence associations were similarly reflected in the likability ratings of the faces, which clustered in valence categories and less in individual emotion levels.

The explicit knowledge about valence contingencies was accompanied by effects in the EPN and a trending LPC effect in the explicit valence classification task (similar to Bruchmann et al., 2021). Importantly, valence effects in the EPN were similarly present in the preceding valence-implicit (old/new) task, suggesting that, once learned, valence associations in faces could attract attention, although no focus on valence or emotion was prompted by the task (similar to Abdel Rahman, 2011; Luo et al., 2016). This might particularly apply to tasks in which no other stimulus features compete strongly for attention as it is in the old-new task. Notably, our implementation of the old-new task in *Study 3* was comparatively difficult for participants due to the stimulus material and task-design since only the inner facial features served to distinguish between iden-

doi: 10.53846/goediss-9841

tities and since non-associated ("new") faces were also repeatedly presented. In contrast, later, more elaborate processing, as measured in the LPC, seemed to more strongly depend on intention. This was indicated by substantially larger differences between valence categories in the valence-classification task (cf. Bruchmann et al., 2021).

Contrary to our prediction, associated valence did not modulate early processing (cf. Aguado et al., 2012; Hammerschmidt et al., 2017; Muench et al., 2016; Rehbein et al., 2014; Schacht et al., 2012; Sperl et al., 2021; Steinberg et al., 2012). Unlike in *Study 1*, in which faces were repeatedly and systematically conditioned with the US in a similar context, learning in *Study 3* showed more variation in individual learning styles, as learning was distributed over days and contexts and, importantly, appeared to be more explicit since participants invented mnemonics to study the face-voice pairs. Possibly, the effects of associated valence in *Study 3* occurred after the structural encoding of the face and might have been bound to the recognition of the individual face. How these differences in learning procedures modulate the effects of perceptual learning and interact with stimulus potency would be an interesting target for further research, particularly since already very few conditioning trials in the context of fear conditioning were shown to affect visual processing at an early stage (e.g., Rehbein et al., 2015; Rehbein et al., 2014).

In summary, we demonstrated that faces can be associated even with stimuli of moderate intensity if the task focus is not directed to another emotion-independent feature during learning. The dissociation between valence effects in the EPN and LPC depending on the task show a multifaceted interplay of task relevance and experience-driven attention for retrieval of associated valence. Moreover, with extensive training on specific CS-US features, such as face-voice gender-congruence in *Study 1*, we potentially identified signs of perceptual learning and selection history.

Superiority of valence-specific information

So far, general effects of valence in comparison with non-expressive ("neutral") stimuli have been presented. However, it has not been discussed whether there were differences between non-expressive stimuli or effects of associated valence that differed between valences. All studies in this thesis included stimuli of both positive and negative affect (*Studies 1, 2, and 4* included happy and angry affect bursts; *Study 3*, happiness, elation, anger, and disgust expressing vocal bursts; *Study 5* showed happy and angry facial expressions). However, not all measures were equally affected by positively and negatively valenced stimuli.

Behavioral differences for angry and happy affect bursts were apparent in response times of the *Studies 1 and 2* during learning, with the slowest responses for happy bursts, largely sta-

doi: 10.53846/goediss-9841

ble throughout the learning sessions, and a steeper learning curve for angry bursts. Although it is possible that angry bursts were learned faster due to their biological relevance (e.g., Siegman et al., 1990), the most parsimonious explanation would be that gender information unfolds at different speed in happy and angry affect bursts (compare with emotional unfolding Schaerlaeken & Grandjean, 2018; cf. Fecteau et al., 2007). This alternative explanation could be excluded with a replication implementing a task independent of other stimulus features that potentially differ between bursts. Moreover, in the more flexible learning task of *Study 3*, positive face-voice pairs were learned faster than negative face-voice pairs, i.e., they required fewer repetitions to correctly match faces and voices. A learning bias for positive over neutral or negative outcomes has also been reported in reward-learning studies (Bayer et al., 2018; e.g., Hammerschmidt, Kulke, et al., 2018; Hammerschmidt et al., 2017; Kulke et al., 2019; Rossi et al., 2017).

Electrophysiological effects differed for mid- and long latencies between positive and negative valences. For the affect bursts in *Study 1*, the auditory P2 (i.e., mid-latency ERP) showed no difference between happy and angry affect bursts (as in Pell et al., 2015). Similarly to the auditory modality, both happy and angry facial expressions in *Study 5* were enhanced for the face-locked N170 (Aguado et al., 2012; Bublatzky et al., 2014; Müller-Bardorff et al., 2016; cf. Aguado et al., 2012; Hammerschmidt, Kagan, et al., 2018) and EPN (e.g., Aguado et al., 2012; Bublatzky et al., 2014; Calvo & Beltrán, 2013; Kulke et al., 2021; Recio et al., 2014; Rellecke et al., 2012a; cf. Valdés-Conroy et al., 2014). In contrast, the effects of associated valence on the face-locked EPN (*Study 3*) were enhanced for negative associations (similar to Abdel Rahman, 2011; Baum & Rahman, 2021; Luo et al., 2016; Suess et al., 2014; M. J. Wieser, Gerdes, et al., 2014), but not for positive associations (cf. Abdel Rahman, 2011). Moreover, in *Study 3* associated disgust affected already the N170, which would require further research on whether this effect is due to the special role of disgust-related cues or whether it might resemble a potential threat of social exclusion. This was different for the associated valence effects on the LPC in *Study 3*, with no differences between positive and negative associations, whereas the effects of inherent expressions on the LPC in Study 5 were only significantly enlarged for happy facial expressions (cf. Calvo & Beltrán, 2013; Hammerschmidt, Kulke, et al., 2018; Schacht & Sommer, 2009). Possibly, the superiority of happy faces in *Study 5* was driven by searching for artifacts in the artificial happy faces.

In sum, faces and voices expressing emotion showed similar effects for positive and negative valences during mid-latency processing stages but differed at later processing and in the behavioral measures, possibly due to interactions with the task. In contrast, the effects of associated valence were more pronounced for negative valences, evident in the N170 and EPN. Since later processing (LPC) included effects of both valences, we conclude that also positive valence was associated to

doi: 10.53846/goediss-9841

the faces but required intentional retrieval to become effective whereas associated negative information was effective irrespective of the task, suggesting a threat-related bias (Dimberg & Öhman, 1996; Esteves, Dimberg, et al., 1994; Esteves, Parra, et al., 1994; Öhman, 2005).

Future directions

Whereas specific limitations have been discussed in the previous chapters, there are also general limitations of this dissertation. With the experimental approaches implemented, we took neither a molecular nor a molar perspective on the fundamental research question about the processing of social cues: We tried to aim for high experimental control while at the same time including complex stimuli such as faces and voices that naturally differ on several dimensions. The choice of reducing faces to their inner part or even presenting them in greyscale (*Study 5*) might raise questions about the generalizability to the naturalistic context in which usually faces and voices are perceived (but see Schindler et al., 2019 for manipulations of face naturalness). However, the general mechanism behind prioritization and attention effects are often tested with more controlled visual and auditory stimuli (e.g., grating patterns and sine waves), which allows for isolating the processes of interest. By using social stimuli like faces and voices of different identities, the stimulus variability requires relatively large numbers of trials to un-blur the targeted processes, while increasing the number of trials poses further issues of habituation effects, specifically for affective content. The biggest challenge might be to find the balance between isolating the processes of interest through experimental control and at the same time not experimentally eliminating them.

Several follow-up studies might be conceived to address open questions that emerged from the findings of the presented studies: We observed additive effects of valence and stimulus size on the processing of inherent facial expressions (*Study 5*). In contrast, stimuli like emotional words which at some point acquired valence indicated interaction effects of stimulus size and valence, it would be interesting to compare size effects in various domains, i.e., with stimuli of inherent and acquired valence (e.g., words, associated faces, faces with inherent expressions, and scenes).

The main aim of the thesis was to test whether we could associate the valence of affect bursts cross-modally to faces. In *Study 3*, we confirmed that affect bursts could be associated to faces and that faces associated with cross-modal valence showed prioritization also during extinction. However, to test the comparator hypothesis that also in the emotion-implicit learning task (*Studies 1 and 2*), valence associations in the face were acquired but not activated due to persistent focus on gender-information, it would be interesting to replicate *Study 1* with the identical learning task but a test task which is valence specific, e.g., the valence-classification task of *Study 3*.

doi: 10.53846/goediss-9841

A further remarkable finding was the conditioned gender-congruence effects on early visual processing in the test session of *Study 1*. The prioritization of task-relevant features in the face during learning was continued in the test session, despite the slightly different task setting. I proposed that it was the sensory learning that drove the effect, although this would imply that not only a specific previously trained stimulus or its location can lead to sensory learning but also target-specific features within one stimulus (such as gender-related cues). It would be interesting to replicate this effect for other facial features, such as age-related cues.

An open question is to what extent sensory learning is facilitated by biologically relevant information since it also has been shown in studies with arbitrary stimuli (e.g., Kim & Anderson, 2019). Results from *Study 1* indicated that the crossmodal associations of affective bursts were not sufficient to elicit effects of associated valence during learning and test when valence was task-irrelevant. Nevertheless, gender information resembles no completely arbitrary feature but can be regarded as biologically and socially relevant. Contrasting the association of faces with arbitrary congruent or incongruent information or biologically/socially relevant information might uncover the domain generality of this learning mechanism.

Furthermore, the relationship between sensory preconditioning and sensory learning would be an interesting topic for future research to test especially the early effects of acquired valence in more naturalistic settings, where exactly the same perceptual stimulus configuration of the learning situation will unlikely occur twice. Research on the generalizability of sensory learning effects in faces might address issues on the prioritization of identity vs. general stimulus similarity. One option would be to test to which extent sensory learning generalizes to pre-conditioned alternative representations of a face (e.g., slightly different lighting, view, etc., cf. Unkelbach et al., 2012) or only to stimuli that share general perceptual similarity (Haddad et al., 2013; Whalley, 2015).

Conclusion

In this dissertation, I showed how task-relevance and emotional valence interact in the learning and retrieval of socially relevant stimuli, such as faces and voices, in a multi-method approach including behavioral and (electro-)physiological measures. The present research provides evidence for both preferential processing of emotional over neutral information on different time scales, ranging from early to later processing, and, importantly, that attention effects of emotional information are not immune against task requirements. The extensive training of discriminating task-relevant stimulus features produced effects of perceptual learning that were evident after overnight consolidation in early visual processing, in line with the proposed selection-history account by B.

doi: 10.53846/goediss-9841

A. Anderson et al. (2011). Furthermore, emotional expressions in the face and voice of both positive and negative valence impacted the EPN and auditory P2, respectively, indicating preferential processing of biologically prepared and socially relevant stimuli. Moreover, it could be shown that also non-expressive faces can gain additional relevance when cross-modally associated with affective vocalizations, with EPN effects for associations of negative valence, irrespective of whether valence was task-relevant. Since attentional effects of associated valence were tested in a controlled experimental laboratory setting, I suspect that further empirical examinations in a more natural and dynamic environment will bring out boosted effects of social relevance, as it has been shown similarly with dynamic expressions of emotion (e.g., Recio et al., 2011, 2014). In summary, there is evidence for both the prioritization of inherent emotional expressions and faces having acquired relevance through their association with affective social cues, possibly for resembling an adaptive mechanism to navigate through the social environment.

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

Abdel Rahman, R. (2011). Facing good and evil: Early brain signatures of affective biographical knowledge in face recognition. *Emotion*, *11*(6), 1397–1405. https://doi.org/10.1037/a0024717

Aguado, L., Valdés-Conroy, B., Rodríguez, S., Román, F. J., Diéguez-Risco, T., & Fernández-Cahill, M. (2012). Modulation of Early Perceptual Processing by Emotional Expression and Acquired Valence of Faces. *Journal of Psychophysiology*, *26*(1), 29–41. https://doi.org/10.1027/0269-8803/a000065

Akaike, H. (1983). Information measures and model selection. *Int Stat Inst*, *44*, 277–291.

Allison, T. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, *9*(5), 415–430. https://doi.org/10.1093/cercor/9.5.415

Amir, N., Klumpp, H., Elias, J., Bedwell, J. S., Yanasak, N., & Miller, L. S. (2005). Increased activation of the anterior cingulate cortex during processing of disgust faces in individuals with social phobia. *Biological Psychiatry*, *57*(9), 975–981. https://doi.org/10.1016/j.biopsych.2005.01.044

An, J., & Hsiao, J. H. (2021). Modulation of mood on eye movement and face recognition performance. *Emotion*, *21*(3), 617–630. https://doi.org/10.1037/emo0000724

Anderson, A. K. (2005). Affective Influences on the Attentional Dynamics Supporting Awareness. *Journal of Experimental Psychology: General*, *134*(2), 258–281. https://doi.org/10.1037/0096-3445.134.2.258

Anderson, B. (2021). Stop paying attention to “attention”. *WIREs Cognitive Science*. https://doi.org/10.1002/wcs.1574

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00246

Anderson, B. A. (2021). Time to stop calling it attentional “capture” and embrace a mechanistic understanding of attentional priority. *Visual Cognition*, *29*(9), 537–540. https://doi.org/10.1080/13506285.2021.1892894

Anderson, B. A., & Halpern, M. (2017). On the value-dependence of value-driven attentional

doi: 10.53846/goediss-9841

capture. *Attention, Perception, & Psychophysics*, *79*(4), 1001–1011. https://doi.org/10.3758/s13414-017-1289-6

Anderson, B. A., Kim, H., Kim, A. J., Liao, M.-R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience & Biobehavioral Reviews*, *130*, 326–350. https://doi.org/10.1016/j.neubiorev.2021.09.004

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

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451–474. https://doi.org/10.1016/0010-0285(74)90021-8

Angulo-Perkins, A., & Concha, L. (2019). Discerning the functional networks behind processing of music and speech through human vocalizations. *PLOS ONE*, *14*(10), e0222796. https://doi.org/10.1371/journal.pone.0222796

Anikin, A. (2019). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods*, *51*(2), 778–792. https://doi.org/10.3758/s13428-018-1095-7

Antono, J. E., Vakhrushev, R., & Pooresmaeili, A. (2022). Value-driven modulation of visual perception by visual and auditory reward cues: The role of performance-contingent delivery of reward. *Frontiers in Human Neuroscience*, *16*. https://doi.org/10.3389/fnhum.2022.1062168

Antov, M. I., Plog, E., Bierwirth, P., Keil, A., & Stockhorst, U. (2020). Visuocortical tuning to a threat-related feature persists after extinction and consolidation of conditioned fear. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60597-z

Apergis-Schoute, A. M., Schiller, D., LeDoux, J. E., & Phelps, E. A. (2014). Extinction resistant changes in the human auditory association cortex following threat learning. *Neurobiology of Learning and Memory*, *113*, 109–114. https://doi.org/10.1016/j.nlm.2014.01.016

Armony, J. (2002). Modulation of spatial attention by fear-conditioned stimuli: An event-related fMRI study. *Neuropsychologia*, *40*(7), 817–826. https://doi.org/10.1016/s0028-3932(01)00178-6

Armony, J. L., Servan-Schreiber, D., Cohen, J. D., & LeDoux, J. E. (1997). Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning. *Trends in Cognitive Sciences*, *1*(1), 28–34. https://doi.org/10.1016/S1364-6613(97)01007-3

Arslan, R. C., Walther, M. P., & Tata, C. S. (2019). Formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using r. *Behavior Research Methods*, *52*(1), 376–387. https://doi.org/10.3758/s13428-019-01236-y

Atkinson, A. P., Tipples, J., Burt, D. M., & Young, A. W. (2005). Asymmetric interference between

doi: 10.53846/goediss-9841

sex and emotion in face perception. *Perception & Psychophysics*, *67*(7), 1199–1213. https://doi.org/10.3758/bf03193553

Audacity Team. (2021).

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion*, *11*(6), 1406–1414. https://doi.org/10.1037/a0023578

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage*, *42*(2), 919–927. https://doi.org/10.1016/j.neuroimage.2008.05.034

Balderston, N. L., & Helmstetter, F. J. (2010). Conditioning with masked stimuli affects the timecourse of skin conductance responses. *Behavioral Neuroscience*, *124*(4), 478–489. https://doi.org/10.1037/a0019927

Balderston, N. L., Schultz, D. H., Baillet, S., & Helmstetter, F. J. (2014). Rapid Amygdala Responses during Trace Fear Conditioning without Awareness. *PLoS ONE*, *9*(5), e96803. https://doi.org/10.1371/journal.pone.0096803

Balsam, P., Drew, M., & Gallistel, C. (2010). Time and associative learning. *Comparative Cognition & Behavior Reviews*, *5*, 1–22. https://doi.org/10.3819/ccbr.2010.50001

Baltrusaitis, T., Mahmoud, M., & Robinson, P. (2015, May). Cross-dataset learning and person-specific normalisation for automatic action unit detection. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. https://doi.org/10.1109/fg.2015.7284869

Baltrusaitis, T., Robinson, P., & Morency, L.-P. (2016, March). OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. https://doi.org/10.1109/wacv.2016.7477553

Banko, E. M., Gal, V., Kortvelyes, J., Kovacs, G., & Vidnyanszky, Z. (2011). Dissociating the effect of noise on sensory processing and overall decision difficulty. *Journal of Neuroscience*, *31*(7), 2663–2674. https://doi.org/10.1523/JNEUROSCI.2725-10.2011

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Barbero, F. M., Calce, R. P., Talwar, S., Rossion, B., & Collignon, O. (2021). Fast periodic auditory stimulation reveals a robust categorical response to voices in the human brain. *Eneuro*, *8*(3),

doi: 10.53846/goediss-9841

ENEURO.0471–20.2021. https://doi.org/10.1523/eneuro.0471-20.2021

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, *17*(3), 613–620. https://doi.org/10.1016/s0926-6410(03)00174-5

Baum, J., Rabovsky, M., Rose, S. B., & Rahman, R. A. (2020). Clear judgments based on unclear evidence: Person evaluation is strongly influenced by untrustworthy gossip. *Emotion*, *20*(2), 248–260. https://doi.org/10.1037/emo0000545

Baum, J., & Rahman, R. A. (2021). Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation. *NeuroImage*, *244*, 118572. https://doi.org/10.1016/j.neuroimage.2021.118572

Bayer, M., Grass, A., & Schacht, A. (2018). Associated valence impacts early visual processing of letter strings: Evidence from ERPs in a cross-modal learning paradigm. *Cognitive, Affective and Behavioral Neuroscience*, *19*(1), 98–108. https://doi.org/10.3758/s13415-018-00647-2

Bayer, M., & Schacht, A. (2014). Event-related brain responses to emotional words, pictures, and faces – a cross-domain comparison. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01106

Bayer, M., Sommer, W., & Schacht, A. (2012). Font Size Matters—Emotion and Attention in Cortical Responses to Written Words. *PLoS ONE*, *7*(5), e36042. https://doi.org/10.1371/journal.pone.0036042

Becker, S. I., Dutt, N., Vromen, J. M. G., & Horstmann, G. (2017). The capture of attention and gaze in the search for emotional photographic faces. *Visual Cognition*, *25*(1-3), 241–261. https://doi.org/10.1080/13506285.2017.1333182

Beckes, L., Coan, J. A., & Morris, J. P. (2013). Implicit conditioning of faces via the social regulation of emotion: ERP evidence of early attentional biases for security conditioned faces. *Psychophysiology*, *50*(8), 734–742. https://doi.org/10.1111/psyp.12056

Belin, P. (2017). Similarities in face and voice cerebral processing. *Visual Cognition*, *25*(4-6), 658–665. https://doi.org/10.1080/13506285.2017.1339156

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. https://doi.org/10.3758/brm.40.2.531

doi: 10.53846/goediss-9841

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. https://doi.org/10.1016/s0926-6410(01)00084-2

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identifcation in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, *17*(1-3), 35–55. https://doi.org/10.1080/026432900380472

Bevins, M., Rick A. (1997). Converging evidence for one-trial context fear conditioning with an immediate shock: Importance of shock potency. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*(3), 312–324.

Bieniek, M. M., Frei, L. S., & Rousselet, G. A. (2013). Early ERPs to faces: Aging, luminance, and individual differences. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00268

Björkstrand, J., Pine, D. S., & Frick, A. (2022). Evaluating an internet-delivered fear conditioning and extinction protocol using response times and affective ratings. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-07999-3

Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior*, *27*(1), 63–76. https://doi.org/10.3758/bf03199432

Blechert, J., Testa, G., Georgii, C., Klimesch, W., & Wilhelm, F. H. (2016). The pavlovian craver: Neural and experiential correlates of single trial naturalistic food conditioning in humans. *Physiology & Behavior*, *158*, 18–25. https://doi.org/10.1016/j.physbeh.2016.02.028

Bobak, P., Bodis-Wollner, I., & Guillory, S. (1987). The effect of blur and contrast of VEP latency: Comparison between check and sinusoidal grating patterns. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(4), 247–255. https://doi.org/10.1016/0168-5597(87)90045-1

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer (Computer program). Version 6.0.37*. http://www.praat.org/

Borghesani, V., Narvid, J., Battistella, G., Shwe, W., Watson, C., Binney, R. J., Sturm, V., Miller, Z., Mandelli, M. L., Miller, B., & Gorno-Tempini, M. L. (2019). "Looks familiar, but i do not know who she is": The role of the anterior right temporal lobe in famous face recognition. *Cortex*, *115*, 72–85. https://doi.org/10.1016/j.cortex.2019.01.006

Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects.

doi: 10.53846/goediss-9841

Journal of Affective Disorders, *33*(4), 215–226. https://doi.org/10.1016/0165-0327(94)00092-n

Bradfield, L., & McNally, G. P. (2008). Unblocking in pavlovian fear conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*(2), 256–265. https://doi.org/10.1037/0097-7403.34.2.256

Bradley, M. M., Keil, A., & Lang, P. J. (2012). Orienting and emotional perception: Facilitation, attenuation, and interference. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00493

Braly, A. M., DeLucia, P., & Oberfeld, D. (2021). Does affective content of sounds affect auditory time-to-collision estimation? *Auditory Perception & Cognition*, *4*(3-4), 212–234. https://doi.org/10.1080/25742442.2021.1997064

Brewer, G. A. (2011). Analyzing response time distributions. *Zeitschrift Für Psychologie*, *219*(2), 117–124. https://doi.org/10.1027/2151-2604/a000056

Breyer, B., & Bluemke, M. (2016). Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel). In *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. ZIS - GESIS Leibniz Institute for the Social Sciences. https://doi.org/10.6102/ZIS242

Brockelmann, A.-K., Steinberg, C., Elling, L., Zwanzger, P., Pantev, C., & Junghofer, M. (2011). Emotion-Associated Tones Attract Enhanced Attention at Early Auditory Processing: Magnetoencephalographic Correlates. *Journal of Neuroscience*, *31*(21), 7801–7810. https://doi.org/10.1523/jneurosci.6236-10.2011

Brown, A. S. (2020). The butcher on the bus experience. In *Memory quirks* (pp. 224–247). Routledge. https://doi.org/10.4324/9780429264498-17

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Bruchmann, M., Schindler, S., Heinemann, J., Moeck, R., & Straube, T. (2021). Increased early and late neuronal responses to aversively conditioned faces across different attentional conditions. *Cortex*, *142*, 332–341. https://doi.org/10.1016/j.cortex.2021.07.003

Bruchmann, M., Schindler, S., & Straube, T. (2020). The spatial frequency spectrum of fearful faces modulates early and mid-latency ERPs but not the N170. *Psychophysiology*, *57*(9). https://doi.org/10.1111/psyp.13597

Brückner, S. (2019). *The impact of emotional sounds on the processing and perception of neutral faces. A classical conditioning paradigm* [Master's thesis]. Georg-August-Universität Goettingen.

Bublatzky, F., Gerdes, A. B. M., White, A. J., Riemer, M., & Alpers, G. W. (2014). Social and

doi: 10.53846/goediss-9841

emotional relevance in face processing: Happy faces of future interaction partners enhance the late positive potential. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00493

Bublatzky, F., Guerra, P., & Alpers, G. W. (2018). Verbal instructions override the meaning of facial expressions. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-33269-2

Bublatzky, F., Kavcıoğlu, F., Guerra, P., Doll, S., & Junghoefer, M. (2020). Contextual information resolves uncertainty about ambiguous facial emotions: Behavioral and magnetoencephalographic correlates. *NeuroImage*, *215*, 116814. https://doi.org/10.1016/j.neuroimage.2020.116814

Bublatzky, F., Riemer, M., & Guerra, P. (2019). Reversing Threat to Safety: Incongruence of Facial Emotions and Instructed Threat Modulates Conscious Perception but Not Physiological Responding. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02091

Burklund, L. J., Eisenberger, N. I., & Lieberman, M. D. (2007). The face of rejection: Rejection sensitivity moderates dorsal anterior cingulate activity to disapproving facial expressions. *Social Neuroscience*, *2*(3-4), 238–253. https://doi.org/10.1080/17470910701391711

Bürkner, P.-C. (2019). *Bayesian Item Response Modeling in R with brms and Stan*. arXiv. https://doi.org/10.48550/ARXIV.1905.09501

Burley, D. T., Gray, N. S., & Snowden, R. J. (2017). As Far as the Eye Can See: Relationship between Psychopathic Traits and Pupil Response to Affective Stimuli. *PLOS ONE*, *12*(1), 1–22. https://doi.org/10.1371/journal.pone.0167436

Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., & Herrmann, C. S. (2004). Size matters: Effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clinical Neurophysiology*, *115*(8), 1810–1820. https://doi.org/10.1016/j.clinph.2004.03.015

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Cabanac, M. (2002). What is emotion? *Behavioural Processes*, *60*(2), 69–83. https://doi.org/10.1016/S0376-6357(02)00078-5

Caharel, S., & Rossion, B. (2021). The N170 is Sensitive to Long-term (Personal) Familiarity of a Face Identity. *Neuroscience*, *458*, 244–255. https://doi.org/10.1016/j.neuroscience.2020.12.036

Calvo, M. G., & Beltrán, D. (2013). Recognition advantage of happy faces: Tracing the neurocognitive processes. *Neuropsychologia*, *51*(11), 2051–2060. https://doi.org/10.1016/j.neuropsychologia.2013.07.010

doi: 10.53846/goediss-9841

Camfield, D. A., Mills, J., Kornfeld, E. J., & Croft, R. J. (2016). Modulation of the N170 with Classical Conditioning: The Use of Emotional Imagery and Acoustic Startle in Healthy and Depressed Participants. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00337

Cano, M. E., Class, Q. A., & Polich, J. (2009). Affective valence, stimulus attributes, and P300: Color vs. Black/white and normal vs. Scrambled images. *International Journal of Psychophysiology*, *71*(1), 17–24. https://doi.org/10.1016/j.ijpsycho.2008.07.016

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, *26*(3), 321–352. https://doi.org/10.1016/s0149-7634(02)00007-6

Carretié, L. (2014). Exogenous (automatic) attention to emotional stimuli: A review. *Cognitive, Affective and Behavioral Neuroscience*, *14*(4), 1228–1258. https://doi.org/10.3758/s13415-014-0270-2

Cesarei, A. D., & Codispoti, M. (2013). Spatial frequencies and emotional perception. *Reviews in the Neurosciences*, *24*(1). https://doi.org/10.1515/revneuro-2012-0053

Cesarei, A. D., & Codispoti, M. (2006). When does size not matter? Effects of stimulus size on affective modulation. *Psychophysiology*, *43*(2), 207–215. https://doi.org/10.1111/j.1469-8986.2006.00392.x

Chen, S., Tan, Z., Xia, W., Gomes, C. A., Zhang, X., Zhou, W., Liang, S., Axmacher, N., & Wang, L. (2021). Theta oscillations synchronize human medial prefrontal cortex and amygdala during fear learning. *Science Advances*, *7*(34). https://doi.org/10.1126/sciadv.abf4198

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience*, *8*(5), 387–402. https://doi.org/10.1162/jocn.1996.8.5.387

Cochrane, A., Sims, C., Bejjanki, V., Green, S., & Bavelier, D. (2022). *Multiple timescales of learning within a single task: Continuous-time changes in evidence-accumulation processes during perceptual decision-making*. https://doi.org/10.21203/rs.3.rs-2057667/v1

Codispoti, M., & Cesarei, A. D. (2007). Arousal and attention: Picture size and emotional reactions. *Psychophysiology*, *44*(5), 680–686. https://doi.org/10.1111/j.1469-8986.2007.00545.x

Coggan, D. D., Allen, L. A., Farrar, O. R. H., Gouws, A. D., Morland, A. B., Baker, D. H., & Andrews, T. J. (2017). Differences in selectivity to natural images in early visual areas (V1–V3). *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-02569-4

Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). The role of visual and semantic properties in the emergence of category-specific patterns of neural response in the human brain. *Eneuro*,

doi: 10.53846/goediss-9841

3(4), ENEURO.0158–16.2016. https://doi.org/10.1523/eneuro.0158-16.2016

Cooney, R. E., Atlas, L. Y., Joormann, J., Eugène, F., & Gotlib, I. H. (2006). Amygdala activation in the processing of neutral faces in social anxiety disorder: Is neutral really neutral? *Psychiatry Research: Neuroimaging*, *148*(1), 55–59. https://doi.org/10.1016/j.pscychresns.2006.05.003

Cosme, G., Rosa, P. J., Lima, C. F., Tavares, V., Scott, S., Chen, S., Wilcockson, T. D. W., Crawford, T. J., & Prata, D. (2021). Pupil dilation reflects the authenticity of received nonverbal vocalizations. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-83070-x

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist*, *74*(6), 698–712. https://doi.org/10.1037/amp0000399

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear Conditioning in Humans: The Influenceof Awareness and Autonomic Arousalon Functional Neuroanatomy. *Neuron*. https://doi.org/10.1016/s0896-6273(02)00588-3

Curby, K. M., Johnson, K. J., & Tyson, A. (2012). Face to face with emotion: Holistic face processing is modulated by emotional state. *Cognition & Emotion*, *26*(1), 93–102. https://doi.org/10.1080/02699931.2011.555752

Curran, T., & Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. *NeuroImage*, *36*(2), 464–471. https://doi.org/10.1016/j.neuroimage.2006.12.016

Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology*, *12*(14), R476–R477. https://doi.org/10.1016/s0960-9822(02)00960-0

Dal Ben, R. (2019). *SHINE color and Lum_fun: A set of tools to control luminance of colorful images (Version 0.2). [Computer program]*. Open Science Framework. https://doi.org/10.17605/OSF.IO/AUZJY

Dalmaijer, E. S., Mathôt, S., & Stigchel, S. V. der. (2013). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

Davis, F. C., Johnstone, T., Mazzulla, E. C., Oler, J. A., & Whalen, P. J. (2009). Regional response differences across the human amygdaloid complex during social conditioning. *Cerebral Cortex*, *20*(3), 612–621. https://doi.org/10.1093/cercor/bhp126

Dawson, M. E., Rissling, A. J., Schell, A. M., & Wilcox, R. (2007). Under what conditions can human affective conditioning occur without contingency awareness? Test of the evaluative conditioning paradigm. *Emotion*, *7*(4), 755–766. https://doi.org/10.1037/1528-3542.7.4.755

De Cesarei, A., Mastria, S., & Codispoti, M. (2013). Early Spatial Frequency Processing of Natural

doi: 10.53846/goediss-9841

Images: An ERP Study. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0065103

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delplanque, S., N'diaye, K., Scherer, K., & Grandjean, D. (2007). Spatial frequencies or emotional effects? *Journal of Neuroscience Methods*, *165*(1), 144–150. https://doi.org/10.1016/j.jneumeth.2007.05.030

Dessel, P. V., & Vogt, J. (2012). When does hearing laughter draw attention to happy faces? Task relevance determines the influence of a crossmodal affective context on emotional attention. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00294

Deweese, M. M., Müller, M., & Keil, A. (2016). Extent and time-course of competition in visual cortex between emotionally arousing distractors and a concurrent task. *European Journal of Neuroscience*, *43*(7), 961–970. https://doi.org/10.1111/ejn.13180

Dimberg, U., & Öhman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion*, *20*(2), 149–182. https://doi.org/10.1007/BF02253869

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory*, *11*(5), 549–554. https://doi.org/10.1101/lm.78004

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09239-1

Doi, H., & Shinohara, K. (2013). Unconscious Presentation of Fearful Face Modulates Electrophysiological Responses to Emotional Prosody. *Cerebral Cortex*, *25*(3), 817–832. https://doi.org/10.1093/cercor/bht282

Dolan, R. J., Morris, J. S., & Gelder, B. de. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences*, *98*(17), 10006–10010. https://doi.org/10.1073/pnas.171288598

Dormal, G., Pelland, M., Rezk, M., Yakobov, E., Lepore, F., & Collignon, O. (2018). Functional preference for object sounds and voices in the brain of early blind and sighted individuals. *Journal of Cognitive Neuroscience*, *30*(1), 86–106. https://doi.org/10.1162/jocn_a_01186

Dresler, T., Mériau, K., Heekeren, H. R., & Meer, E. (2008). Emotional Stroop task: effect of word arousal and subject anxiety on emotional interference. *Psychological Research Psychologische Forschung*, *73*(3), 364–371. https://doi.org/10.1007/s00426-008-0154-6

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, *1*(1), 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

doi: 10.53846/goediss-9841

Duffy, E. (1941). An explanation of "emotional" phenomena without the use of the concept "emotion". *The Journal of General Psychology*, *25*(2), 283–293. https://doi.org/10.1080/00221309.1941.10544400

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*(6635), 808–810. https://doi.org/10.1038/42947

Dutilh, G., Krypotos, A.-M., & Wagenmakers, E.-J. (2011). Task-related versus stimulus-specific practice. *Experimental Psychology*, *58*(6), 434–442. https://doi.org/10.1027/1618-3169/a000111

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*(6), 1026–1036. https://doi.org/10.3758/16.6.1026

Eastwood, J., Smilek, D., & Merikle, P. (2003). *Negative facial expression captures attention and disrupts performance*.

Eilers, P. H. C., & Marx, B. D. (2010). Splines, knots, and penalties. *WIREs Computational Statistics*, *2*(6), 637–653. https://doi.org/10.1002/wics.125

Eimer, M. (2000). An ERP study of sustained spatial attention to stimulus eccentricity. *Biological Psychology*, *52*(3), 205–220. https://doi.org/10.1016/s0301-0511(00)00028-4

Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences*, *105*(5), 1704–1709. https://doi.org/10.1073/pnas.0707727105

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. American Psychological Association (APA). https://doi.org/10.1037/t27734-000

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Esteves, F., Dimberg, U., & Öhman, A. (1994). Automatically elicited fear: Conditioned skin conductance responses to masked facial expressions. *Cognition and Emotion*, *8*(5), 393–413. https://doi.org/10.1080/02699939408408949

Esteves, F., Parra, C., Dimberg, U., & Öhman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology*, *31*(4), 375–385. https://doi.org/10.1111/j.1469-8986.1994.tb02446.x

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., & Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *NeuroReport*, *17*(3), 249–253. https://doi.org/10.1097/01.wnr.0000199466.32036.5d

doi: 10.53846/goediss-9841

Failing, M., & Theeuwes, J. (2017). Selection history: How reward modulates selectivity of visual attention. *Psychonomic Bulletin & Review*, *25*(2), 514–538. https://doi.org/10.3758/s13423-017-1380-y

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage*, *36*(2), 480–487. https://doi.org/10.1016/j.neuroimage.2007.02.043

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00215

Finke, J. B., Roesmann, K., Stalder, T., & Klucken, T. (2021). Pupil dilation as an index of Pavlovian conditioning. A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *130*, 351–368. https://doi.org/10.1016/j.neubiorev.2021.09.005

Föcker, J., Gondan, M., & Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica*, *137*(1), 36–47. https://doi.org/10.1016/j.actpsy.2011.02.004

Fockert, J. W. de, Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*(5509), 1803–1806. https://doi.org/10.1126/science.1056496

Fontanesi, L., Gluth, S., Spektor, M. S., & Rieskamp, J. (2019). A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review*, *26*(4), 1099–1121. https://doi.org/10.3758/s13423-018-1554-2

Forgas, J. P., & East, R. (2008). How real is that smile? Mood effects on accepting or rejecting the veracity of emotional facial expressions. *Journal of Nonverbal Behavior*, *32*(3), 157–170. https://doi.org/10.1007/s10919-008-0050-1

Foti, D., Olvet, D. M., Klein, D. N., & Hajcak, G. (2010). Reduced electrocortical response to threatening faces in major depressive disorder. *Depression and Anxiety*, *27*(9), 813–820. https://doi.org/10.1002/da.20712

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637–1651. https://doi.org/10.1002/hbm.20630

Freeman, J. B., & Ambady, N. (2011a). When two become one: Temporally dynamic integration of the face and voice. *Journal of Experimental Social Psychology*, *47*(1), 259–263. https://doi.org/10.1016/j.jesp.2010.08.018

Freeman, J. B., & Ambady, N. (2011b). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279. https://doi.org/10.1037/a0022327

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

doi: 10.53846/goediss-9841

Frowd, C. D., Skelton, F., Atherton, C., Pitchford, M., Hepton, G., Holden, L., McIntyre, A. H., & Hancock, P. J. B. (2012). Recovering faces from memory: The distracting influence of external facial features. *Journal of Experimental Psychology: Applied*, *18*(2), 224–238. https://doi.org/10.1037/a0027393

Gan, X., Zhou, X., Li, J., Jiao, G., Jiang, X., Biswal, B., Yao, S., Klugah-Brown, B., & Becker, B. (2022). Common and distinct neurofunctional representations of core and social disgust in the brain: Coordinate-based and network meta-analyses. *Neuroscience & Biobehavioral Reviews*, *135*, 104553. https://doi.org/10.1016/j.neubiorev.2022.104553

Gao, C., Conte, S., Richards, J. E., Xie, W., & Hanayik, T. (2019). The neural sources of N170: Understanding timing of activation in face-selective areas. *Psychophysiology*, *56*(6), e13336. https://doi.org/10.1111/psyp.13336

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, *4*(1), 123–124. https://doi.org/10.3758/bf03342209

Garrido-Vásquez, P., Pell, M. D., Paulmann, S., & Kotz, S. A. (2018). Dynamic Facial Expressions Prime the Processing of Emotional Prosody. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00244

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, *16*(18), 1824–1829. https://doi.org/10.1016/j.cub.2006.07.072

Gelder, B. de, & Vroomen, J. (2000). The perception of emotion by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311. https://doi.org/10.1080/026999300378824

George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Cognitive Brain Research*, *4*(2), 65–76. https://doi.org/10.1016/0926-6410(95)00045-3

Gerdes, A. B. M., Wieser, M. J., Bublatzky, F., Kusay, A., Plichta, M. M., & Alpers, G. W. (2013). Emotional sounds modulate early neural processing of emotional pictures. *Frontiers in Psychology*, *4*(October), 1–12. https://doi.org/10.3389/fpsyg.2013.00741

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning & Behavior*, *40*(3), 255–268. https://doi.org/10.3758/s13420-012-0080-8

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*(5), 681–697. https://doi.org/10.1016/s0896-6273(01)00424-x

GitHub - snekbeater/scrambpy: Scramb.py is a region based JPEG image scrambler and descrambler written in python for end-to-end-encrypted (E2EE) image distribution through unaware channels. (n.d.). Retrieved January 12, 2023, from https://github.com/snekbeater/scrambpy#

doi: 10.53846/goediss-9841

whitepaper

Glenn, C. R., Lieberman, L., & Hajcak, G. (2012). Comparing electric shock and a fearful screaming face as unconditioned stimuli for fear learning. *International Journal of Psychophysiology*, *86*(3), 214–219. https://doi.org/10.1016/j.ijpsycho.2012.09.006

Gliga, T., & Dehaene-Lambertz, G. (2005). Structural encoding of body and face in human infants and adults. *Journal of Cognitive Neuroscience*, *17*(8), 1328–1340. https://doi.org/10.1162/0898929055002481

Gore, F., Schwartz, E. C., Brangers, B. C., Aladi, S., Stujenske, J. M., Likhtik, E., Russo, M. J., Gordon, J. A., Salzman, C. D., & Axel, R. (2015). Neural representations of unconditioned stimuli in basolateral amygdala mediate innate and learned responses. *Cell*, *162*(1), 134–145. https://doi.org/10.1016/j.cell.2015.06.027

Gottfried, J. A., ODoherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *The Journal of Neuroscience*, *22*(24), 10829–10837. https://doi.org/10.1523/JNEUROSCI.22-24-10829.2002

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146. https://doi.org/10.1038/nn1392

Green, D. M. (1971). Fourier analysis of reaction time data. *Behavior Research Methods & Instrumentation*, *3*(3), 121–125. https://doi.org/10.3758/bf03209931

Gunnery, S. D., & Ruben, M. A. (2015). Perceptions of duchenne and non-duchenne smiles: A meta-analysis. *Cognition and Emotion*, *30*(3), 501–515. https://doi.org/10.1080/02699931.2015.1018817

Gutiérrez-Cobo, M. J., Luque, D., Most, S. B., Fernández-Berrocal, P., & Pelley, M. E. L. (2019). Reward and emotion influence attentional bias in rapid serial visual presentation. *Quarterly Journal of Experimental Psychology*, *72*(9), 2155–2167. https://doi.org/10.1177/1747021819840615

Haddad, A. D. M., Xu, M., Raeder, S., & Lau, J. Y. F. (2013). Measuring the role of conditioning and stimulus generalisation in common fears and worries. *Cognition & Emotion*, *27*(5), 914–922. https://doi.org/10.1080/02699931.2012.747428

Hammerschmidt, W., Kagan, I., Kulke, L., & Schacht, A. (2018). Implicit reward associations impact face processing: Time-resolved evidence from event-related brain potentials and pupil dilations. *NeuroImage*, *179*, 557–569. https://doi.org/10.1016/j.neuroimage.2018.06.055

Hammerschmidt, W., Kulke, L., Broering, C., & Schacht, A. (2018). Money or smiles: Inde-

doi: 10.53846/goediss-9841

pendent ERP effects of associated monetary reward and happy faces. *PLOS ONE*, *13*(10), e0206142. https://doi.org/10.1371/journal.pone.0206142

Hammerschmidt, W., Sennhenn-Reulen, H., & Schacht, A. (2017). Associated motivational salience impacts early sensory processing of human faces. *NeuroImage*, *156*, 466–474. https://doi.org/10.1016/j.neuroimage.2017.04.032

Handy, T. C., Soltani, M., & Mangun, G. R. (2001). Perceptual load and visuocortical processing: Event-related potentials reveal sensory-level selection. *Psychological Science*, *12*(3), 213–218. https://doi.org/10.1111/1467-9280.00338

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*(6), 917–924. https://doi.org/10.1037/0022-3514.54.6.917

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently Ambiguous: Facial Expressions of Emotions, in Context. *Emotion Review*, *5*(1), 60–65. https://doi.org/10.1177/1754073912451331

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/s1364-6613(00)01482-0

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, *22*(1), 189–199. https://doi.org/10.1016/s0896-6273(00)80690-x

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207. https://doi.org/10.3758/bf03212979

Heisz, J. J., & Shedden, J. M. (2009). Semantic learning modifies perceptual face processing. *Journal of Cognitive Neuroscience*, *21*(6), 1127–1134. https://doi.org/10.1162/jocn.2009.21104

Herbert, C., Sfärlea, A., & Blumenthal, T. (2013). Your emotion or mine: Labeling feelings alters emotional face perception—an ERP study on automatic and intentional affect labeling. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00378

Herrmann, M. J., Ehlis, A.-C., Ellgring, H., & Fallgatter, A. J. (2004). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *Journal of Neural Transmission*, *112*(8), 1073–1081. https://doi.org/10.1007/s00702-004-0250-8

Hester, N. (2019). Perceived negative emotion in neutral faces: Gender-dependent effects on attractiveness and threat. *Emotion*, *19*(8), 1490–1494. https://doi.org/10.1037/emo0000525

Hills, P. J., & Lewis, M. B. (2011). Sad people avoid the eyes or happy people focus on the eyes?

doi: 10.53846/goediss-9841

Mood induction affects facial feature discrimination. *British Journal of Psychology*, *102*(2), 260–274. https://doi.org/10.1348/000712610x519314

Hills, P. J., Werno, M. A., & Lewis, M. B. (2011). Sad people are more accurate at face recognition than happy people. *Consciousness and Cognition*, *20*(4), 1502–1517. https://doi.org/10.1016/j.concog.2011.07.002

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787. https://doi.org/10.1073/pnas.95.3.781

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1257–1270. https://doi.org/10.1098/rstb.1998.0281

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *55*, 498–509. https://doi.org/10.1016/j.neubiorev.2015.06.002

Ho, H. T., Schröger, E., & Kotz, S. A. (2015). Selective Attention Modulates Early Human Evoked Potentials during Emotional Face-Voice Processing. *Journal of Cognitive Neuroscience*, *27*(4), 798–818. https://doi.org/10.1162/jocn_a_00734

Hofmann, W., Houwer, J. D., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*(4), 382–386. https://doi.org/10.1037/h0021740

Houwer, J. de, Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with “subliminally” presented stimuli. *Consciousness and Cognition*, *6*(1), 87–107. https://doi.org/10.1006/ccog.1996.0281

Hubert, M., & Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, *52*(12), 5186–5201. https://doi.org/10.1016/j.csda.2007.11.008

Hudson, A., Durston, A. J., McCrackin, S. D., & Itier, R. J. (2021). Emotion, Gender and Gaze Discrimination Tasks do not Differentially Impact the Neural Processing of Angry or Happy Facial Expressions—a Mass Univariate ERP Analysis. *Brain Topography*, *34*(6), 813–833. https://doi.org/10.1007/s10548-021-00873-x

Huestegge, S. M., Raettig, T., & Huestegge, L. (2019). Are Face-Incongruent Voices Harder to Pro-

doi: 10.53846/goediss-9841

cess? *Experimental Psychology*, *66*(2), 154–164. https://doi.org/10.1027/1618-3169/a000440

Hur, J., Iordan, A. D., Berenbaum, H., & Dolcos, F. (2016). Emotion–attention interactions in fear conditioning: Moderation by executive load, neuroticism, and awareness. *Biological Psychology*, *121*, 213–220. https://doi.org/10.1016/j.biopsycho.2015.10.007

Hur, J., Iordan, A. D., Dolcos, F., & Berenbaum, H. (2016). Emotional influences on perception and working memory. *Cognition and Emotion*, *31*(6), 1294–1302. https://doi.org/10.1080/02699931.2016.1213703

Iredale, J. M., Rushby, J. A., McDonald, S., Marco, A. D.-D., & Swift, J. (2013). Emotion in voice matters: Neural correlates of emotional prosody perception. *International Journal of Psychophysiology*, *89*(3), 483–490. https://doi.org/10.1016/j.ijpsycho.2013.06.025

Ishai, A. (2008). Let's face it: It's a cortical network. *NeuroImage*, *40*(2), 415–419. https://doi.org/10.1016/j.neuroimage.2007.10.040

Itier, R. J., & Neath-Tavares, K. N. (2017). Effects of task demands on the early neural processing of fearful and happy facial expressions. *Brain Research*, *1663*, 38–50. https://doi.org/10.1016/j.brainres.2017.03.013

Jacques, C., & Rossion, B. (2004). Concurrent processing reveals competition between visual representations of faces. *NeuroReport*, *15*(15), 2417–2421. https://doi.org/10.1097/00001756-200410250-00023

Jamal, Y., Lacey, S., Nygaard, L., & Sathian, K. (2017). Interactions Between Auditory Elevation, Auditory Pitch and Visual Elevation During Multisensory Perception. *Multisensory Research*, *30*(3-5), 287–306. https://doi.org/10.1163/22134808-00002553

Jessen, S., & Kotz, S. A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Frontiers in Human Neuroscience*, *7*(July), 1–7. https://doi.org/10.3389/fnhum.2013.00369

Jiang, F., Stecker, G. C., & Fine, I. (2013). Functional localization of the auditory thalamus in individual human subjects. *NeuroImage*, *78*, 295–304. https://doi.org/10.1016/j.neuroimage.2013.04.035

Johannes, S., Münte, T. F., Heinze, H. J., & Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. *Cognitive Brain Research*, *2*(3), 189–205. https://doi.org/10.1016/0926-6410(95)90008-x

Johnson, J. A., & Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *NeuroImage*, *31*(4), 1673–1681. https://doi.org/10.1016/j.neuroimage.2006.02.026

Johnstone, T., & Scherer, K. (2000). *The Handbook of Emotion* (M. Lewis & J. Haviland, Eds.).

doi: 10.53846/goediss-9841

Guilford.

Judah, M. R., Grant, D. M., & Carlisle, N. B. (2015). The effects of self-focus on attentional biases in social anxiety:an ERP study. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(3), 393–405. https://doi.org/10.3758/s13415-015-0398-8

Junghöfer, M., Rehbein, M. A., Maitzen, J., Schindler, S., & Kissler, J. (2016). An evil face? Verbal evaluative multi-CS conditioning enhances face-evoked mid-latency magnetoencephalographic responses. *Social Cognitive and Affective Neuroscience*, *12*(4), 695–705. https://doi.org/10.1093/scan/nsw179

Jürgens, R., Fischer, J., & Schacht, A. (2018). Hot speech and exploding bombs: Autonomic arousal during emotion classification of prosodic utterances and affective sounds. *Frontiers in Psychology*, *9*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2018.00228

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/jneurosci.17-11-04302.1997

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00390

Kaufmann, J. M., & Schweinberger, S. R. (2004). Expression influences the recognition of familiar faces. *Perception*, *33*(4), 399–408. https://doi.org/10.1068/p5083

Kesteren, M. T. R. van, Rijpkema, M., Ruiter, D. J., & Fernandez, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *Journal of Neuroscience*, *30*(47), 15888–15894. https://doi.org/10.1523/jneurosci.2674-10.2010

Kim, H., & Anderson, B. A. (2019). Dissociable neural mechanisms underlie value-driven and selection-driven attentional capture. *Brain Research*, *1708*, 109–115. https://doi.org/10.1016/j.brainres.2018.11.026

Kissler, J., & Strehlow, J. (2017). Something always sticks? How emotional language modulates neural processes involved in face encoding and recognition memory. *Poznan Studies in Contemporary Linguistics*, *53*(1). https://doi.org/10.1515/psicl-2017-0004

Klasen, M., Chen, Y.-H., & Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4). https://doi.org/10.1515/revneuro-2012-0040

Kneib, T. (2013). Beyond mean regression. *Statistical Modelling*, *13*(4), 275–303. https://doi.org/10.1177/1471082X13494159

doi: 10.53846/goediss-9841

Koenig-Robert, R., & VanRullen, R. (2013). SWIFT: A novel method to track the neural correlates of recognition. *NeuroImage*, *81*, 273–282. https://doi.org/10.1016/j.neuroimage.2013.04.116

Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2014). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognitive and Affective Neuroscience*, *10*(5), 713–720. https://doi.org/10.1093/scan/nsu105

Kornmeier, J., Pfaffle, M., & Bach, M. (2011). Necker cube: Stimulus-related (low-level) and percept-related (high-level) EEG signatures early in occipital cortex. *Journal of Vision*, *11*(9), 12–12. https://doi.org/10.1167/11.9.12

Korth, M., & Nguyen, N. X. (1997). The Effect of Stimulus Size on Human Cortical Potentials Evoked by Chromatic Patterns. *Vision Research*, *37*(5), 649–657. https://doi.org/10.1016/s0042-6989(96)00189-7

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107–118. https://doi.org/10.1016/j.brainres.2007.03.015

Kret, M. E., Roelofs, K., Stekelenburg, J. J., & Gelder, B. de. (2013). Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00810

Kret, M. E., & Sjak-Shie, E. E. (2018). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*(3), 1336–1342. https://doi.org/10.3758/s13428-018-1075-y

Kulke, L., Bayer, M., Grimm, A.-M., & Schacht, A. (2019). Differential effects of learned associations with words and pseudowords on event-related brain potentials. *Neuropsychologia*, *124*, 182–191. https://doi.org/10.1016/j.neuropsychologia.2018.12.012

Kulke, L., Bruemmer, L., Pooresmaeili, A., & Schacht, A. (2021). Overt and covert attention shifts to emotional faces: Combining EEG, eye tracking, and a go/no-go paradigm. *Psychophysiology*, *58*(8). https://doi.org/10.1111/psyp.13838

Kulke, L., Janßen, L., Demel, R., & Schacht, A. (2017). Validating the Goettingen Faces Database. In *OSF*. Open Science Framework. https://doi.org/10.17605/OSF.IO/4KNPF

Kyllingsbæk, S., Van Lommel, S., Sørensen, T. A., & Bundesen, C. (2014). Automatic attraction of visual attention by supraletter features of former target strings. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01383

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

Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other

doi: 10.53846/goediss-9841

probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 35–45. https://doaj.org/article/6c7d5e9fa29a4cb7aac3c97bb4094ad0

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2010). Pupillary stroop effects. *Cognitive Processing*, *12*(1), 13–21. https://doi.org/10.1007/s10339-010-0370-z

Lang, P. J., Bradley, M. M., Cuthbert, B. N., et al. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*.

Lange, K., Kühn, S., & Filevich, E. (2015). ”Just another tool for online studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & Knippenberg, A. van. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Latinus, M., & Taylor, M. J. (2006). Face processing stages: Impact of difficulty and the separation of effects. *Brain Research*, *1123*(1), 179–187. https://doi.org/10.1016/j.brainres.2006.09.031

Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BMC Neuroscience*, *11*(1). https://doi.org/10.1186/1471-2202-11-36

Lausen, A., & Hammerschmidt, K. (2020). Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications*, *7*(1). https://doi.org/10.1057/s41599-020-0499-z

Lausen, A., & Schacht, A. (2018). Gender Differences in the Recognition of Vocal Emotions. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00882

Leber, A. B., Kawahara, J.-I., & Gabari, Y. (2009). Long-term abstract learning of attentional set. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1385–1397. https://doi.org/10.1037/a0016470

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*(1), 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

Leeuw, J. R. de. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

LeMoult, J., Yoon, K. L., & Joormann, J. (2012). Affective priming in major depressive disorder. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00076

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans

doi: 10.53846/goediss-9841

Leo, F., Romei, V., Freeman, E., Ladavas, E., & Driver, J. (2011). Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. *Experimental Brain Research*, *213*(2-3), 193–201. https://doi.org/10.1007/s00221-011-2742-8

Lerche, V., Neubauer, A. B., & Voss, A. (2018). Effects of implicit fear of failure on cognitive processing: A diffusion model analysis. *Motivation and Emotion*, *42*(3), 386–402. https://doi.org/10.1007/s11031-018-9691-5

Lerche, V., Voss, A., & Nagler, M. (2016). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2

Lim, S.-L., Padmala, S., & Pessoa, L. (2008). Affective learning modulates spatial competition during low-load attentional conditions. *Neuropsychologia*, *46*(5), 1267–1278. https://doi.org/10.1016/j.neuropsychologia.2007.12.003

Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters*, *253*(3), 147–150. https://doi.org/10.1016/s0304-3940(98)00586-2

Liu, C. C., & Watanabe, T. (2012). Accounting for speed–accuracy tradeoff in perceptual learning. *Vision Research*, *61*, 107–114. https://doi.org/10.1016/j.visres.2011.09.007

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, *23*(2), 108–112. https://doi.org/10.1097/wnr.0b013e32834ea757

Liu, T., Pinheiro, A., Zhao, Z., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Emotional Cues during Simultaneous Face and Voice Processing: Electrophysiological Insights. *PLoS ONE*, *7*(2), e31001. https://doi.org/10.1371/journal.pone.0031001

Liu, Y., Keil, A., & Ding, M. (2011). Effects of emotional conditioning on early visual processing: Temporal dynamics revealed by ERP single-trial analysis. *Human Brain Mapping*, *33*(4), 909–919. https://doi.org/10.1002/hbm.21259

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01171

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, *12*(1), 43–65.

Lomoriello, A. S., Meconi, F., Rinaldi, I., & Sessa, P. (2018). Out of Sight Out of Mind: Perceived Physical Distance Between the Observer and Someone in Pain Shapes Observer's Neural Em-

doi: 10.53846/goediss-9841

pathic Reactions. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01824

Luce, R. D. et al. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press on Demand.

Luck, C. C., & Lipp, O. V. (2015). A potential pathway to the relapse of fear? Conditioned negative stimulus evaluation (but not physiological responding) resists instructed extinction. *Behaviour Research and Therapy*, *66*, 18–31. https://doi.org/10.1016/j.brat.2015.01.001

Luck, C. C., & Lipp, O. V. (2016). The influence of contingency reversal instructions on electrodermal responding and conditional stimulus valence evaluations during differential fear conditioning. *Learning and Motivation*, *54*, 1–11. https://doi.org/10.1016/j.lmot.2016.05.001

Luck, S. J. (2005). *An introduction to the event-related potential technique, second edition*. Cambridge, MA: MIT Press.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440. https://doi.org/10.1016/s1364-6613(00)01545-x

Lundqvist, D., & Öhman, A. (2005). Emotion regulates attention: The relation between facial configurations, facial emotion, and visual attention. *Visual Cognition*, *12*(1), 51–84. https://doi.org/10.1080/13506280444000085

Lunghi, C., & Pooresmaeili, A. (2023). Learned value modulates the access to visual awareness during continuous flash suppression. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-28004-5

Luo, Q. L., Wang, H. L., Dzhelyova, M., Huang, P., & Mo, L. (2016). Effect of Affective Personality Information on Face Processing: Evidence from ERPs. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00810

Luzardo, A., Alonso, E., & Mondragón, E. (2017). A rescorla-wagner drift-diffusion model of conditioning and timing. *PLOS Computational Biology*, *13*(11), e1005796. https://doi.org/10.1371/journal.pcbi.1005796

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E. L. T., & Anna di Palma, M. (2021). *robustbase: Basic Robust Statistics*. http://robustbase.r-forge.r-project.org/

Maiworm, M., Bellantoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception, & Psychophysics*, *74*(6), 1302–1311. https://doi.org/10.3758/s13414-012-0310-3

Mara, C. A., & Cribbie, R. A. (2012). Paired-samples tests of equivalence. *Communications in Statistics - Simulation and Computation*, *41*(10), 1928–1943. https:

doi: 10.53846/goediss-9841

//doi.org/10.1080/03610918.2011.626545

Marcar, V. L., & Wolf, M. (2021). An investigation into the relationship between stimulus property, neural response and its manifestation in the visual evoked potential involving retinal resolution. *European Journal of Neuroscience*, *53*(8), 2612–2628. https://doi.org/10.1111/ejn.15112

Marcos, J. L., & Marcos, A. (2021). Concurrent Electrodermal and Eyeblink Conditioning With Masked and Unmasked Stimuli. *Journal of Psychophysiology*, *35*(1), 35–42. https://doi.org/10.1027/0269-8803/a000259

Marletta, A., & Sciandra, M. (2020). GAMLSS for high-variability data: An application to liver fibrosis case. *The International Journal of Biostatistics*, *16*(2). https://doi.org/10.1515/ijb-2019-0113

Marmolejo-Ramos, F., Barrera-Causil, C., Kuang, S., Fazlali, Z., Wegener, D., Kneib, T., Bastiani, F. D., & Martinez-Flórez, G. (2022). Generalised exponential-gaussian distribution: A method for neural reaction time analysis. *Cognitive Neurodynamics*. https://doi.org/10.1007/s11571-022-09813-2

Marmolejo-Ramos, F., Tejo, M., Brabec, M., Kuzilek, J., Joksimovic, S., Kovanovic, V., González, J., Kneib, T., Bühlmann, P., Kook, L., Briseño-Sánchez, G., & Ospina, R. (2022). Distributional regression modeling via generalized additive models for location, scale, and shape: An overview through a data set from learning analytics. *WIREs Data Mining and Knowledge Discovery*. https://doi.org/10.1002/widm.1479

Martin, A., Becker, S. I., & Pegna, A. J. (2021). Attention is prioritised for proximate and approaching fearful faces. *Cortex*, *134*, 52–64. https://doi.org/10.1016/j.cortex.2020.10.011

Masuda, S., Tsujii, T., & Watanabe, S. (2005). An interference effect of voice presentation on face gender discrimination task: Evidence from event-related potentials. *International Congress Series*, *1278*, 156–159. https://doi.org/10.1016/j.ics.2004.11.193

MATLAB. (2018). *version 9.4.0.949201(R2018a)*. The MathWorks Inc.

Matzel, L. D., Schachtman, T. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, *16*(4), 398–412. https://doi.org/10.1016/0023-9690(85)90023-2

McConnell, B. L., Urushihara, K., & Miller, R. R. (2010). Contrasting predictions of extended comparator hypothesis and acquisition-focused models of learning concerning retrospective revaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(1), 137–147. https://doi.org/10.1037/a0015774

McCrackin, S. D., & Itier, R. J. (2018). Is it about me? Time-course of self-relevance and valence effects on the perception of neutral faces with direct and averted gaze. *Biological Psychology*,

doi: 10.53846/goediss-9841

135, 47–64. https://doi.org/10.1016/j.biopsycho.2018.03.003

McGill, W. J., & Gibbon, J. (1965). The general-gamma distribution and reaction times. *Journal of Mathematical Psychology*, *2*(1), 1–18. https://doi.org/10.1016/0022-2496(65)90014-3

Meah, L. F. S., & Moore, R. K. (2014). The uncanny valley: A focus on misaligned cues. *Social Robotics*, 256–265. https://doi.org/10.1007/978-3-319-11973-1_26

Melinscak, F., & Bach, D. R. (2020). Computational optimization of associative learning experiments. *PLOS Computational Biology*, *16*(1), e1007593. https://doi.org/10.1371/journal.pcbi.1007593

Menon, V., & Levitin, D. J. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *NeuroImage*, *28*(1), 175–184. https://doi.org/10.1016/j.neuroimage.2005.05.053

Menz, M. M., Rihm, J. S., & Büchel, C. (2016). REM Sleep Is Causal to Successful Consolidation of Dangerous and Safety Stimuli and Reduces Return of Fear after Extinction. *The Journal of Neuroscience*, *36*(7), 2148–2160. https://doi.org/10.1523/jneurosci.3083-15.2016

Mertens, G., & Engelhard, I. M. (2020). A systematic review and meta-analysis of the evidence for unaware fear conditioning. *Neuroscience and Biobehavioral Reviews*, *108*, 254–268. https://doi.org/10.1016/j.neubiorev.2019.11.012

Metin, B., Roeyers, H., Wiersema, J. R., Meere, J. J. van der, Thompson, M., & Sonuga-Barke, E. (2013). ADHD performance reflects inefficient but not impulsive information processing: A diffusion model analysis. *Neuropsychology*, *27*(2), 193–200. https://doi.org/10.1037/a0031533

Minagawa-Kawai, Y., Lely, H. van der, Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2010). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex*, *21*(2), 254–261. https://doi.org/10.1093/cercor/bhq082

Miskovic, V., & Keil, A. (2012). Acquired fears reflected in cortical sensory processing: A review of electrophysiological studies of human classical conditioning. *Psychophysiology*, *49*(9), 1230–1241. https://doi.org/10.1111/j.1469-8986.2012.01398.x

Mitchell, D. G. V., Nakic, M., Fridberg, D., Kamel, N., Pine, D. S., & Blair, R. J. R. (2007). The impact of processing load on emotion. *NeuroImage*, *34*(3), 1299–1309. https://doi.org/10.1016/j.neuroimage.2006.10.012

Morel, S., Beaucousin, V., Perrin, M., & George, N. (2012). Very early modulation of brain responses to neutral faces by a single prior association with an emotional context: Evidence from MEG. *NeuroImage*, *61*(4), 1461–1470. https://doi.org/10.1016/j.neuroimage.2012.04.016

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning

doi: 10.53846/goediss-9841

in the human amygdala. *Nature*, *393*(6684), 467–470. https://doi.org/10.1038/30976

Mueller, C. J., & Kuchinke, L. (2016). Individual differences in emotion word processing: A diffusion model analysis. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(3), 489–501. https://doi.org/10.3758/s13415-016-0408-5

Mueller, C. J., & Kuchinke, L. (2015). Processing of face identity in the affective flanker task: A diffusion model analysis. *Psychological Research*, *80*(6), 963–973. https://doi.org/10.1007/s00426-015-0696-3

Mueller, E. M., & Pizzagalli, D. A. (2015). One-year-old fear memories rapidly activate human fusiform gyrus. *Social Cognitive and Affective Neuroscience*, *11*(2), 308–316. https://doi.org/10.1093/scan/nsv122

Muench, H. M., Westermann, S., Pizzagalli, D. A., Hofmann, S. G., & Mueller, E. M. (2016). Self-relevant threat contexts enhance early processing of fear-conditioned faces. *Biological Psychology*, *121*, 194–202. https://doi.org/10.1016/j.biopsycho.2016.07.017

Mullen, T. (2012). *NITRC: CleanLine: Tool/Resource Info*. http://www.nitrc.org/projects/cleanlin

Müller, M. M., Andersen, S. K., & Keil, A. (2007). Time course of competition for visual processing resources between emotional pictures and foreground task. *Cerebral Cortex*, *18*(8), 1892–1899. https://doi.org/10.1093/cercor/bhm215

Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I., & Eickhoff, S. B. (2011). Incongruence effects in crossmodal emotional integration. *NeuroImage*, *54*(3), 2257–2266. https://doi.org/10.1016/j.neuroimage.2010.10.047

Müller-Bardorff, M., Bruchmann, M., Mothes-Lasch, M., Zwitserlood, P., Schlossmacher, I., Hofmann, D., Miltner, W., & Straube, T. (2018). Early brain responses to affective faces: A simultaneous EEG-fMRI study. *NeuroImage*, *178*, 660–667. https://doi.org/10.1016/j.neuroimage.2018.05.081

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

Navarro-Pardo, E., Navarro-Prados, A. B., Gamermann, D., & Moret-Tatay, C. (2013). Differences Between Young and Old University Students on a Lexical Decision Task: Evidence Through an Ex-Gaussian Approach. *The Journal of General Psychology*, *140*(4), 251–268. https://doi.org/10.1080/00221309.2013.817964

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

doi: 10.53846/goediss-9841

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

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, *135*(2), 151–166. https://doi.org/10.3200/genp.135.2.151-166

Odling-Smee, F. J. (1975). The role of background stimuli during pavlovian conditioning. *Quarterly Journal of Experimental Psychology*, *27*(2), 201–209. https://doi.org/10.1080/14640747508400480

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, *30*(10), 953–958. https://doi.org/10.1016/j.psyneuen.2005.03.019

Öhman, A., Eriksson, A., & Olofsson, C. (1975). One-trial learning and superior resistance to extinction of autonomic responses conditioned to potentially phobic stimuli. *Journal of Comparative and Physiological Psychology*, *88*(2), 619–627. https://doi.org/10.1037/h0078388

Öhman, A., Erixon, G., & Löfberg, I. (1975). Phobias and preparedness: Phobic versus neutral pictures as conditioned stimuli for human autonomic responses. *Journal of Abnormal Psychology*, *84*(1), 41–45. https://doi.org/10.1037/h0076255

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

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037/0033-295x.108.3.483

Ojala, K. E., & Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations. *Neuroscience and Biobehavioral Reviews*, *114*, 96–112. https://doi.org/10.1016/j.neubiorev.2020.04.019

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oliva, M., & Anikin, A. (2018). Pupil dilation reflects the time course of emotion recognition in human vocalizations. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-23265-x

Olofsson, J. K., & Polich, J. (2008). Affective visual event-related potentials: Arousal, repetition, and time-on-task. *Biological Psychology*, *75*(1), 101–108. https://doi.org/10.1016/j.biopsycho.2006.12.006

Pace-Schott, E. F., Germain, A., & Milad, M. R. (2015). Effects of sleep on memory for condi-

doi: 10.53846/goediss-9841

tioned fear and fear extinction. *Psychological Bulletin*, *141*(4), 835–857. https://doi.org/10.1037/bul0000014

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, *45*(1), 75–92. https://doi.org/10.1016/j.neuropsychologia.2006.04.025

Palmer, C., Cheng, S.-Y., & Seidemann, E. (2007). Linking neuronal and behavioral performance in a reaction-time visual detection task. *Journal of Neuroscience*, *27*(30), 8122–8137. https://doi.org/10.1523/jneurosci.1940-07.2007

Pan, Z., Liu, X., Luo, Y., & Chen, X. (2017). Emotional Intensity Modulates the Integration of Bimodal Angry Expressions: ERP Evidence. *Frontiers in Neuroscience*, *11*(JUN), 1–9. https://doi.org/10.3389/fnins.2017.00349

Panitz, C., Hermann, C., & Mueller, E. M. (2015). Conditioned and extinguished fear modulate functional corticocardiac coupling in humans. *Psychophysiology*, *52*(10), 1351–1360. https://doi.org/10.1111/psyp.12498

Paquette, S., Rigoulot, S., Grunewald, K., & Lehmann, A. (2020). Temporal decoding of vocal and musical emotions: Same code, different timecourse? *Brain Research*, *1741*, 146887. https://doi.org/10.1016/j.brainres.2020.146887

Pastor, M. C., Rehbein, M. A., Junghöfer, M., Poy, R., López, R., & Moltó, J. (2015). Facing Challenges in Differential Classical Conditioning Research: Benefits of a Hybrid Design for Simultaneous Electrodermal and Electroencephalographic Recording. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00336

Patterson, R. D., Allerhand, M. H., & Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*, *98*(4), 1890–1894. https://doi.org/10.1121/1.414456

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00345

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, *52*(1), 111–139. https://doi.org/10.1146/annurev.psych.52.1.111

Pedersen, M. L., & Frank, M. J. (2020). Simultaneous Hierarchical Bayesian Parameter Estimation for Reinforcement Learning and Drift Diffusion Models: a Tutorial and Links to Neural Data. *Computational Brain and Behavior*, *3*(4), 458–471. https://doi.org/10.1007/s42113-020-00084-w

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10. https://doi.org/10.3389/neuro.11.010.2008

doi: 10.53846/goediss-9841

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Pell, M. D., Sethi, S., Rigoulot, S., Rothermich, K., Liu, P., & Jiang, X. (2022). Emotional voices modulate perception and predictions about an upcoming face. *Cortex*, *149*, 148–164. https://doi.org/10.1016/j.cortex.2021.12.017

Pelley, M. E. L., Haselgrove, M., & Esber, G. R. (2012). Modeling attention in associative learning: Two processes or one? *Learning & Behavior*, *40*(3), 292–304. https://doi.org/10.3758/s13420-012-0084-4

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neutral and emotional stimuli. *Cognitive Brain Research*, *15*(1), 31–45. https://doi.org/10.1016/s0926-6410(02)00214-8

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

Petrov, A. A., Horn, N. M. V., & Ratcliff, R. (2011). Dissociable perceptual-learning mechanisms revealed by diffusion-model analysis. *Psychonomic Bulletin & Review*, *18*(3), 490–497. https://doi.org/10.3758/s13423-011-0079-8

Pfabigan, D. M., Sailer, U., & Lamm, C. (2015). Size does matter! Perceptual stimulus properties affect event-related potentials during feedback processing. *Psychophysiology*, *52*(9), 1238–1247. https://doi.org/10.1111/psyp.12458

Philiastides, M. G. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience*, *26*(35), 8965–8975. https://doi.org/10.1523/jneurosci.1655-06.2006

Pion-Tonachini, L., Makeig, S., & Kreutz-Delgado, K. (2017). Crowd labeling latent Dirichlet allocation. *Knowledge and Information Systems*, *53*(3), 749–765. https://doi.org/10.1007/s10115-017-1053-1

Pischek-Simpson, L. K., Boschen, M. J., Neumann, D. L., & Waters, A. M. (2009). The development of an attentional bias for angry faces following Pavlovian fear conditioning. *Behaviour Research and Therapy*, *47*(4), 322–330. https://doi.org/10.1016/j.brat.2009.01.007

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*(4), 2356–2363. https://doi.org/10.1016/j.neuroimage.2011.03.067

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993).

doi: 10.53846/goediss-9841

Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, *30*(6), 547–558. https://doi.org/10.1111/j.1469-8986.1993.tb02081.x

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79–106. https://doi.org/10.1037/bul0000026

Pooresmaeili, A., FitzGerald, T. H. B., Bach, D. R., Toelch, U., Ostendorf, F., & Dolan, R. J. (2014). Cross-modal effects of value on perceptual acuity and stimulus encoding. *Proceedings of the National Academy of Sciences*, *111*(42), 15244–15249. https://doi.org/10.1073/pnas.1408873111

Pourtois, G. (2004). Electrophysiological Correlates of Rapid Spatial Orienting Towards Fearful Faces. *Cerebral Cortex*, *14*(6), 619–633. https://doi.org/10.1093/cercor/bhh023

Pourtois, G., & Dhar, M. (2012). Integration of Face and Voice During Emotion Perception: Is There Anything Gained for the Perceptual System Beyond Stimulus Modality Redundancy? In *Integrating face and voice in person perception* (pp. 181–206). Springer New York. https://doi.org/10.1007/978-1-4614-3585-3_10

Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological Psychology*, *92*(3), 492–512. https://doi.org/10.1016/j.biopsycho.2012.02.007

Pratt, N., Willoughby, A., & Swick, D. (2011). Effects of Working Memory Load on Visual Selective Attention: Behavioral and Electrophysiological Evidence. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00057

Proverbio, A. M., Vanutelli, M. E., & Viganò, S. (2019). Remembering faces: The effects of emotional valence and temporal recency. *Brain and Cognition*, *135*, 103584. https://doi.org/10.1016/j.bandc.2019.103584

Puls, S., & Rothermund, K. (2017). Attending to emotional expressions: No evidence for automatic capture in the dot-probe task. *Cognition and Emotion*, *32*(3), 450–463. https://doi.org/10.1080/02699931.2017.1314932

Pumarola, A., Agudo, A., Martinez, A. M., Sanfeliu, A., & Moreno-Noguer, F. (2018). *GANimation: Anatomically-aware facial animation from a single image*. arXiv. https://doi.org/10.48550/ARXIV.1807.09251

Purcell, D. G., & Stewart, A. L. (2016). Reacting to emotion: Anger arrests and happiness helps. *The American Journal of Psychology*, *129*(4), 363–380. https://doi.org/10.5406/amerjpsyc.129.4.0363

doi: 10.53846/goediss-9841

Quadflieg, S., Mohr, A., Mentzel, H.-J., Miltner, W. H. R., & Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: The role of task-relevance and social phobia. *Biological Psychology*, *78*(2), 129–137. https://doi.org/10.1016/j.biopsycho.2008.01.014

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ranger, J., & Kuhn, J.-T. (2013). Analyzing response times in tests with rank correlation approaches. *Journal of Educational and Behavioral Statistics*, *38*(1), 61–80. https://doi.org/10.3102/1076998611431086

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*(3), 190–214. https://doi.org/10.1037/0033-295x.83.3.190

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295x.111.2.333

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon Bull Rev.*, *9*(3), 438–481. https://doi.org/10.3758/bf03196302

Recio, G., Schacht, A., & Sommer, W. (2014). Recognizing dynamic facial expressions of emotion: Specificity and intensity effects in event-related brain potentials. *Biological Psychology*, *96*, 111–125. https://doi.org/10.1016/j.biopsycho.2013.12.003

Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Research*, *1376*, 66–75. https://doi.org/10.1016/j.brainres.2010.12.041

Reeves, B., Lang, A., Kim, E. Y., & Tatar, D. (1999). The effects of screen size and message content on attention and arousal. *Media Psychology*, *1*(1), 49–67. https://doi.org/10.1207/s1532785xmep0101_4

Rehbein, M. A., Pastor, M. C., Moltó, J., Poy, R., López-Penadés, R., & Junghöfer, M. (2018). Identity and expression processing during classical conditioning with faces. *Psychophysiology*,

doi: 10.53846/goediss-9841

55(10), e13203. https://doi.org/10.1111/psyp.13203

Rehbein, M. A., Steinberg, C., Wessing, I., Pastor, M. C., Zwitserlood, P., Keuper, K., & Junghöfer, M. (2014). Rapid Plasticity in the Prefrontal Cortex during Affective Associative Learning. *PLOS ONE*, *9*(10), e110720. https://doi.org/10.1371/journal.pone.0110720

Rehbein, M. A., Wessing, I., Zwitserlood, P., Steinberg, C., Eden, A. S., Dobel, C., & Junghoefer, M. (2015). Rapid prefrontal cortex activation towards aversively paired faces and enhanced contingency detection are observed in highly trait-anxious women under challenging conditions. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00155

Rellecke, J. (2012). *Automaticity in Affective Face Processing* [PhD thesis]. Humboldt University of Berlin.

Rellecke, J., Palazova, M., Sommer, W., & Schacht, A. (2011). On the automaticity of emotion processing in words and faces: Event-related brain potentials evidence from a superficial task. *Brain and Cognition*, *77*(1), 23–32. https://doi.org/10.1016/j.bandc.2011.07.001

Rellecke, J., Sommer, W., & Schacht, A. (2012a). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biological Psychology*, *90*(1), 23–32. https://doi.org/10.1016/j.biopsycho.2012.02.002

Rellecke, J., Sommer, W., & Schacht, A. (2012b). Emotion Effects on the N170: A Question of Reference? *Brain Topography*, *26*(1), 62–71. https://doi.org/10.1007/s10548-012-0261-y

Rescorla, R. A. (2008). Evaluating conditioning of related and unrelated stimuli using a compound test. *Learning & Behavior*, *36*(2), 67–74. https://doi.org/10.3758/lb.36.2.67

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*(3), 151–160. https://doi.org/10.1037/0003-066x.43.3.151

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, *54*, 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Röer, J. P., & Cowan, N. (2021). A preregistered replication and extension of the cocktail party phenomenon: One's name captures attention, unexpected words do not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(2), 234–242. https://doi.org/10.1037/xlm0000874

Roesch, E. B., Sander, D., Mumenthaler, C., Kerzel, D., & Scherer, K. R. (2010). Psychophysics of emotion: The QUEST for Emotional Attention. *Journal of Vision*, *10*(3), 1–9. https://doi.org/10.1167/10.3.4

Roesmann, K., Wiens, N., Winker, C., Rehbein, M. A., Wessing, I., & Junghoefer, M. (2020). Fear generalization of implicit conditioned facial features – Behavioral and magnetoen-

doi: 10.53846/goediss-9841

cephalographic correlates. *NeuroImage*, *205*, 116302. https://doi.org/10.1016/j.neuroimage.2019.116302

Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research*, *65*(1). https://doi.org/10.1007/bf00243828

Rooijen, R. van, Ploeger, A., & Kret, M. E. (2017). The dot-probe task to measure emotional attention: A suitable measure in comparative studies? *Psychonomic Bulletin & Review*, *24*(6), 1686–1717. https://doi.org/10.3758/s13423-016-1224-1

Rossi, V., Vanlessen, N., Bayer, M., Grass, A., Pourtois, G., & Schacht, A. (2017). Motivational Salience Modulates Early Visual Cortex Responses across Task Sets. *Journal of Cognitive Neuroscience*, *29*(6), 968–979. https://doi.org/10.1162/jocn_a_01093

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*(12), 1297–1311. https://doi.org/10.1016/j.visres.2011.04.003

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objectsan electrophysiological account of face-specific processes in the human brain. *NeuroReport*, *11*(1), 69–72. https://doi.org/10.1097/00001756-200001170-00014

Roux, P., Christophe, A., & Passerieux, C. (2010). The emotional paradox: Dissociation between explicit and implicit processing of emotional prosody in schizophrenia. *Neuropsychologia*, *48*(12), 3642–3649. https://doi.org/10.1016/j.neuropsychologia.2010.08.021

Rozenkrants, B., Olofsson, J., & Polich, J. (2007). Affective visual event-related potentials: Arousal, valence, and repetition effects for normal and distorted pictures. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2007.10.010

Ruiz-Soler, M., & Beltran, F. S. (2005). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research Psychologische Forschung*, *70*(4), 273–292. https://doi.org/10.1007/s00426-005-0215-z

Russo, F. D. (2003). Source Analysis of Event-related Cortical Activity during Visuo-spatial Attention. *Cerebral Cortex*, *13*(5), 486–499. https://doi.org/10.1093/cercor/13.5.486

Sá, D. S. F. de, Michael, T., Wilhelm, F. H., & Peyk, P. (2018). Learning to see the threat: Temporal dynamics of ERPs of motivated attention in fear conditioning. *Social Cognitive and Affective Neuroscience*, *14*(2), 189–203. https://doi.org/10.1093/scan/nsy103

Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, *28*(2), 259–287. https://doi.org/10.1207/s15516709cog2802_7

doi: 10.53846/goediss-9841

Samelli, A. G., & Schochat, E. (2008). The gaps-in-noise test: Gap detection thresholds in normal-hearing young adults. *International Journal of Audiology*, *47*(5), 238–245. https://doi.org/10.1080/14992020801908244

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *NeuroImage*, *28*(4), 848–858. https://doi.org/10.1016/j.neuroimage.2005.06.023

Sauter, D. A., & Eimer, M. (2010). Rapid Detection of Emotion from Human Vocalizations. *Journal of Cognitive Neuroscience*, *22*(3), 474–481. https://doi.org/10.1162/jocn.2009.21215

Schacht, A., Adler, N., Chen, P., Guo, T., & Sommer, W. (2012). Association with positive outcome induces early effects in event-related brain potentials. *Biological Psychology*, *89*(1), 130–136. https://doi.org/10.1016/j.biopsycho.2011.10.001

Schacht, A., & Sommer, W. (2009). Emotions in word and face processing: Early and late cortical responses. *Brain and Cognition*, *69*(3), 538–550. https://doi.org/10.1016/j.bandc.2008.11.005

Schacht, A., Werheid, K., & Sommer, W. (2008). The appraisal of facial beauty is rapid but not mandatory. *Cognitive, Affective and Behavioral Neuroscience*, *8*(2), 132–142. https://doi.org/10.3758/cabn.8.2.132

Schaerlaeken, S., & Grandjean, D. (2018). Unfolding and dynamics of affect bursts decoding in humans. *PLOS ONE*, *13*(10), e0206216. https://doi.org/10.1371/journal.pone.0206216

Schall, J. D. (2019). Accumulators, neurons, and response time. *Trends in Neurosciences*, *42*(12), 848–860. https://doi.org/10.1016/j.tins.2019.10.001

Schellhaas, S., Arnold, N., Schmahl, C., & Bublatzky, F. (2020). Contextual source information modulates neural face processing in the absence of conscious recognition: A threat-of-shock study. *Neurobiology of Learning and Memory*, *174*, 107280. https://doi.org/10.1016/j.nlm.2020.107280

Scherer, K. R. (2013). Affect bursts as evolutionary precursors of speech and music. In *Stephen j. Gould: The scientific legacy* (pp. 147–167). Springer Milan. https://doi.org/10.1007/978-88-470-5424-0_10

Schimmack, U. (2005). Attentional Interference Effects of Emotional Pictures: Threat, Negativity, or Arousal? *Emotion*, *5*(1), 55–66. https://doi.org/10.1037/1528-3542.5.1.55

Schindler, S., Bruchmann, M., Bublatzky, F., & Straube, T. (2019). Modulation of face- and emotion-selective ERPs by the three most common types of face image manipulations. *Social Cognitive and Affective Neuroscience*, *14*(5), 493–503. https://doi.org/10.1093/scan/nsz027

Schindler, S., Bruchmann, M., Gathmann, B., Moeck, R., & Straube, T. (2021). Effects of low-

doi: 10.53846/goediss-9841

level visual information and perceptual load on P1 and N170 responses to emotional expressions. *Cortex*, *136*, 14–27. https://doi.org/10.1016/j.cortex.2020.12.011

Schindler, S., Bruchmann, M., Krasowski, C., Moeck, R., & Straube, T. (2021). Charged With a Crime: The Neuronal Signature of Processing Negatively Evaluated Faces Under Different Attentional Conditions. *Psychological Science*, *32*(8), 1311–1324. https://doi.org/10.1177/0956797621996667

Schindler, S., Bruchmann, M., Steinweg, A.-L., Moeck, R., & Straube, T. (2020). Attentional conditions differentially affect early, intermediate and late neural responses to fearful and neutral faces. *Social Cognitive and Affective Neuroscience*, *15*(7), 765–774. https://doi.org/10.1093/scan/nsaa098

Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex*, *130*, 362–386. https://doi.org/10.1016/j.cortex.2020.06.010

Schindler, S., Heinemann, J., Bruchmann, M., Moeck, R., & Straube, T. (2022). No trait anxiety influences on early and late differential neuronal responses to aversively conditioned faces across three different tasks. *Cognitive, Affective and Behavioral Neuroscience*. https://doi.org/10.3758/s13415-022-00998-x

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

Schirmer, A., & Adolphs, R. (2017). Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence. *Trends in Cognitive Sciences*, *21*(3), 216–228. https://doi.org/10.1016/j.tics.2017.01.001

Schirmer, A., Chen, C.-B., Ching, A., Tan, L., & Hong, R. Y. (2012). Vocal emotions influence verbal memory: Neural correlates and interindividual differences. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(1), 80–93. https://doi.org/10.3758/s13415-012-0132-8

Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

Schultz, D. H., & Helmstetter, F. J. (2010). Classical conditioning of autonomic fear responses is independent of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(4), 495–500. https://doi.org/10.1037/a0020263

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). *Emotion and attention: Event-related brain potential studies*. 31–51.

doi: 10.53846/goediss-9841

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The Facilitated Processing of Threatening Faces: An ERP Analysis. *Emotion*, *4*(2), 189–200. https://doi.org/10.1037/1528-3542.4.2.189

Schupp, H. T., Stockburger, J., Bublatzky, F., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2008). The selective processing of emotional visual stimuli while detecting auditory targets: An ERP analysis. *Brain Research*, *1230*, 168–176. https://doi.org/10.1016/j.brainres.2008.07.024

Schupp, H. T., Stockburger, J., Codispoti, M., Junghofer, M., Weike, A. I., & Hamm, A. O. (2007). Selective visual attention to emotion. *Journal of Neuroscience*, *27*(5), 1082–1089. https://doi.org/10.1523/JNEUROSCI.3223-06.2007

Schwarz, K. A., Wieser, M. J., Gerdes, A. B. M., Muehlberger, A., & Pauli, P. (2012). Why are you looking like that? How the context influences evaluation and processing of human faces. *Social Cognitive and Affective Neuroscience*, *8*(4), 438–445. https://doi.org/10.1093/scan/nss013

Seitz, A., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in Cognitive Sciences*, *9*(7), 329–334. https://doi.org/10.1016/j.tics.2005.05.010

Seligman, M. E. (1970). On the generality of the laws of learning. *Psychological Review*, *77*(5), 406–418. https://doi.org/10.1037/h0029790

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(6814), 788.

Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2020). Context-dependent memory effects in two immersive virtual reality environments: On mars and underwater. *Psychonomic Bulletin & Review*, *28*(2), 574–582. https://doi.org/10.3758/s13423-020-01835-3

Shinners, P. (2011). *PyGame*. http://pygame.org/.

Siegman, A. W., Anderson, R. A., & Berger, T. (1990). The angry voice: Its effects on the experience of anger and cardiovascular reactivity. *Psychosomatic Medicine*, *52*(6), 631–643. https://doi.org/10.1097/00006842-199011000-00005

Sim, E.-J., Harpaintner, M., & Kiefer, M. (2020). Is subliminal face processing modulated by attentional task sets? Evidence from masked priming effects in a gender decision task. *Open Psychology*, *2*(1), 76–89. https://doi.org/10.1515/psych-2020-0006

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). *afex: Analysis of Factorial Experiments*. https://CRAN.R-project.org/package=afex

Skuk, V. G., & Schweinberger, S. R. (2013). Adaptation Aftereffects in Vocal Emotion Perception Elicited by Expressive Faces and Voices. *PLOS ONE*, *8*(11), e81691. https://doi.org/10.1371/journal.pone.0081691

doi: 10.53846/goediss-9841

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology*, *17*(19), 1680–1685. https://doi.org/10.1016/j.cub.2007.08.043

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Sommer, W., Stürmer, B., Shmuilovich, O., Martin-Loeches, M., & Schacht, A. (2013). How about lunch? Consequences of the meal context on cognition and emotion. *PLoS ONE*, *8*(7), e70314. https://doi.org/10.1371/journal.pone.0070314

Sopp, M. R., Michael, T., Weeß, H.-G., & Mecklinger, A. (2017). Remembering specific features of emotional events across time: The role of REM sleep and prefrontal theta oscillations. *Cognitive, Affective and Behavioral Neuroscience*, *17*(6), 1186–1209. https://doi.org/10.3758/s13415-017-0542-8

Soto-Faraco, S., & Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits. *The Quarterly Journal of Experimental Psychology Section A*, *55*(1), 23–40. https://doi.org/10.1080/02724980143000136

Sperl, M. F. J., Panitz, C., Hermann, C., & Mueller, E. M. (2016). A pragmatic comparison of noise burst and electric shock unconditioned stimuli for fear conditioning research with many trials. *Psychophysiology*, *53*(9), 1352–1365. https://doi.org/10.1111/psyp.12677

Sperl, M. F. J., Wroblewski, A., Mueller, M., Straube, B., & Mueller, E. M. (2021). Learning dynamics of electrophysiological brain signals during human fear conditioning. *NeuroImage*, *226*, 117569. https://doi.org/10.1016/j.neuroimage.2020.117569

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmüller, E., & Münte, T. F. (2009). Neural processing of vocal emotion and identity. *Brain and Cognition*, *69*(1), 121–126. https://doi.org/10.1016/j.bandc.2008.06.003

Stahl, J., Wiese, H., & Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: An event-related potential study. *NeuroReport*, *19*(5), 583–587. https://doi.org/10.1097/wnr.0b013e3282f97b4d

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., & Hoyer, J. (1999). Die Erfassung sozialer Phobie durch die Social Interaction Anxiety Scale (SIAS) und die Social Phobia Scale (SPS). *Zeitschrift für Klinische Psychologie Und Psychotherapie*, *28*(1), 28–36. https://doi.org/10.1026//0084-5345.28.1.28

Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, *23*(7). https://doi.org/10.18637/jss.v023.i07

doi: 10.53846/goediss-9841

Steinberg, C., Bröckelmann, A.-K., Dobel, C., Elling, L., Zwanzger, P., Pantev, C., & Junghöfer, M. (2013). Preferential responses to extinguished face stimuli are preserved in frontal and occipito-temporal cortex at initial but not later stages of processing. *Psychophysiology*, *50*(3), 230–239. https://doi.org/10.1111/psyp.12005

Steinberg, C., Bröckelmann, A.-K., Rehbein, M., Dobel, C., & Junghöfer, M. (2013). Rapid and highly resolving associative affective learning: Convergent electro- and magnetoencephalographic evidence from vision and audition. *Biological Psychology*, *92*(3), 526–540. https://doi.org/10.1016/j.biopsycho.2012.02.009

Steinberg, C., Dobel, C., Schupp, H. T., Kissler, J., Elling, L., Pantev, C., & Junghöfer, M. (2012). Rapid and Highly Resolving: Affective Evaluation of Olfactorily Conditioned Faces. *Journal of Cognitive Neuroscience*, *24*(1), 17–27. https://doi.org/10.1162/jocn_a_00067

Stekelenburg, J. J., & Vroomen, J. (2007). Neural Correlates of Multisensory Integration of Ecologically Valid Audiovisual Events. *Journal of Cognitive Neuroscience*, *19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, *14*(12), 1–16. https://doi.org/10.1167/14.12.6

Straube, T., Mothes-Lasch, M., & Miltner, W. H. R. (2011). Neural mechanisms of the automatic processing of emotional information from faces and voices. *British Journal of Psychology*, *102*(4), 830–848. https://doi.org/10.1111/j.2044-8295.2011.02056.x

Suess, F., Rabovsky, M., & Rahman, R. A. (2014). Perceiving emotions in neutral faces: Expression processing is biased by affective person knowledge. *Social Cognitive and Affective Neuroscience*, *10*(4), 531–536. https://doi.org/10.1093/scan/nsu088

Summerfield, C., & Egner, T. (2016). Feature-Based Attention and Feature-Based Expectation. *Trends in Cognitive Sciences*, *20*(6), 401–404. https://doi.org/10.1016/j.tics.2016.03.008

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

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep32239

Tauscher, J.-P., Castillo, S., Bosse, S., & Magnor, M. (2021, September). EEG-based analysis of the impact of familiarity in the perception of deepfake videos. *2021 IEEE International Conference on Image Processing (ICIP)*. https://doi.org/10.1109/icip42928.2021.9506082

Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Trans-*

doi: 10.53846/goediss-9841

actions of the Royal Society B: Biological Sciences, *368*(1628), 20130055–20130055. https://doi.org/10.1098/rstb.2013.0055

Theodoridou, A., Penton-Voak, I. S., & Rowe, A. C. (2013). A direct examination of the effect of intranasal administration of oxytocin on approach-avoidance motor responses to emotional stimuli. *PLoS ONE*, *8*(2), e58113. https://doi.org/10.1371/journal.pone.0058113

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, *66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorović, D. (2010). Context effects in visual perception and their explanations. *Review of Psychology*, *17*(1), 17–32.

Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, *20*(1), 12–16.

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences*, *105*(49), 19514–19519. https://doi.org/10.1073/pnas.0809662105

Tucciarelli, R., Vehar, N., Chandaria, S., & Tsakiris, M. (2022). On the realness of people who do not exist: The social processing of artificial faces. *iScience*, 105441. https://doi.org/10.1016/j.isci.2022.105441

Unkelbach, C., Stahl, C., & Förderer, S. (2012). Changing CS features alters evaluative responses in evaluative conditioning. *Learning and Motivation*, *43*(3), 127–134. https://doi.org/10.1016/j.lmot.2012.04.003

Valdés-Conroy, B., Aguado, L., Fernández-Cahill, M., Romero-Ferreiro, V., & Diéguez-Risco, T. (2014). Following the time course of face gender and expression processing: A task-dependent ERP study. *International Journal of Psychophysiology*, *92*(2), 59–66. https://doi.org/10.1016/j.ijpsycho.2014.02.005

Valk, J. M. de, Wijnen, J. G., & Kret, M. E. (2015). Anger fosters action. Fast responses in a motor task involving approach movements toward angry faces and bodies. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01240

Ventura-Bort, C., Löw, A., Wendt, J., Dolcos, F., Hamm, A. O., & Weymar, M. (2016). When neutral turns significant: Brain dynamics of rapidly formed associations between neutral stimuli and emotional contexts. *European Journal of Neuroscience*, *44*(5), 2176–2183. https://doi.org/10.1111/ejn.13319

Verosky, S. C., Porter, J., Martinez, J. E., & Todorov, A. (2018). Robust effects of affective person learning on evaluation of faces. *Journal of Personality and Social Psychology*, *114*(4), 516–528. https://doi.org/10.1037/pspa0000109

doi: 10.53846/goediss-9841

Victeur, Q., Huguet, P., & Silvert, L. (2019). Attentional allocation to task-irrelevant fearful faces is not automatic: Experimental evidence for the conditional hypothesis of emotional selection. *Cognition and Emotion*, *34*(2), 288–301. https://doi.org/10.1080/02699931.2019.1622512

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory and Cognition*, *32*(7), 1206–1220. https://doi.org/10.3758/bf03196893

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00336

Voyer, D., & Myles, D. (2017). The effect of unimodal affective priming on dichotic emotion recognition. *Laterality: Asymmetries of Body, Brain and Cognition*, 1–21. https://doi.org/10.1080/1357650x.2017.1404095

Vroomen, J., & Stekelenburg, J. J. (2010). Visual Anticipatory Information Modulates Multisensory Interactions of Artificial Audiovisual Stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. https://doi.org/10.1016/j.tics.2005.10.011

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*(11), 1271–1278. https://doi.org/10.1038/nn1341

Wagenbreth, C., Wattenberg, L., Heinze, H.-J., & Zaehle, T. (2016). Implicit and explicit processing of emotional facial expressions in Parkinson's disease. *Behavioural Brain Research*, *303*, 182–190. https://doi.org/10.1016/j.bbr.2016.01.059

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, *21*(5), 641–671. https://doi.org/10.1080/09541440802205067

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830–841. https://doi.org/10.1037/0033-295x.114.3.830

Walla, P., Rosser, L., Scharfenberger, J., Duregger, C., & Bosshard, S. (2013). Emotion ownership: Different effects on explicit ratings and implicit responses. *Psychology*, *04*(03), 213–216. https://doi.org/10.4236/psych.2013.43a032

doi: 10.53846/goediss-9841

Wang, X., Guo, X., Chen, L., Liu, Y., Goldberg, M. E., & Xu, H. (2016). Auditory to Visual Cross-Modal Adaptation for Emotion: Psychophysical and Neural Correlates. *Cerebral Cortex*, bhv321. https://doi.org/10.1093/cercor/bhv321

Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards real-world blind face restoration with generative facial prior. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wassenhove, V. van, Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 505–517. https://doi.org/10.1016/0022-5371(76)90045-1

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal Adaptation in Right Posterior Superior Temporal Sulcus during Face-Voice Emotional Integration. *Journal of Neuroscience*, *34*(20), 6813–6821. https://doi.org/10.1523/jneurosci.4478-13.2014

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2013). Dissociating task difficulty from incongruence in face-voice emotion integration. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00744

Watters, A. J., Rupert, P. E., Wolf, D. H., Calkins, M. E., Gur, R. C., Gur, R. E., & Turetsky, B. I. (2018). Social aversive conditioning in youth at clinical high risk for psychosis and with psychosis: An ERP study. *Schizophrenia Research*, *202*, 291–296. https://doi.org/10.1016/j.schres.2018.06.027

Weidner, E. M., Schindler, S., Grewe, P., Moratti, S., Bien, C. G., & Kissler, J. (2022). Emotion and attention in face processing: Complementary evidence from surface event-related potentials and intracranial amygdala recordings. *Biological Psychology*, *173*, 108399. https://doi.org/10.1016/j.biopsycho.2022.108399

Weindel, G., Anders, R., Alario, F. X., & Burle, B. (2021). Assessing model-based inferences in decision making with single-trial response time decomposition. *Journal of Experimental Psychology: General*, *150*(8), 1528–1555.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *The Journal of Neuroscience*, *18*(1), 411–418. https://doi.org/10.1523/jneurosci.18-01-00411.1998

doi: 10.53846/goediss-9841

Whalley, K. (2015). Face-to-face with fear generalization. *Nature Reviews Neuroscience*, *17*(1), 3–3. https://doi.org/10.1038/nrn.2015.9

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*. https://doi.org/10.3389/fninf.2013.00014

Wiemer, J., Leimeister, F., & Pauli, P. (2021). Subsequent memory effects on event-related potentials in associative fear learning. *Social Cognitive and Affective Neuroscience*, *16*(5), 525–536. https://doi.org/10.1093/scan/nsab015

Wieser, M. J., & Brosch, T. (2012). Faces in Context: A Review and Systematization of Contextual Influences on Affective Face Processing. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00471

Wieser, M. J., Gerdes, A. B. M., Büngel, I., Schwarz, K. A., Mühlberger, A., & Pauli, P. (2014). Not so harmless anymore: How context impacts the perception and electrocortical processing of neutral faces. *NeuroImage*, *92*, 74–82. https://doi.org/10.1016/j.neuroimage.2014.01.022

Wieser, M. J., Miskovic, V., Rausch, S., & Keil, A. (2014). Different time course of visuocortical signal changes to fear-conditioned faces with direct or averted gaze: A ssVEP study with single-trial analysis. *Neuropsychologia*, *62*, 101–110. https://doi.org/10.1016/j.neuropsychologia.2014.07.009

Wieser, M., Mühlberger, a, Alpers, G., Macht, M., Ellgring, H., & Pauli, P. (2006). Emotion processing in parkinson's disease: Dissociation between early neuronal processing and explicit ratings. *Clinical Neurophysiology*, *117*(1), 94–102. https://doi.org/10.1016/j.clinph.2005.09.009

Wilf, M., Ramot, M., Furman-Haran, E., Arzi, A., Levkovitz, Y., & Malach, R. (2016). Diminished auditory responses during NREM sleep correlate with the hierarchy of language processing. *PLOS ONE*, *11*(6), e0157143. https://doi.org/10.1371/journal.pone.0157143

Willenbockel, V., Lepore, F., Nguyen, D. K., Bouthillier, A., & Gosselin, F. (2012). Spatial frequency tuning during the conscious and non-conscious perception of emotional facial expressions – an intracranial ERP study. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00237

Wronka, E., & Walentowska, W. (2011). Attention modulates emotional expression processing. *Psychophysiology*, *48*(8), 1047–1056. https://doi.org/10.1111/j.1469-8986.2011.01180.x

Xu, M., Li, Z., Diao, L., Fan, L., & Yang, D. (2016). Contextual Valence and Sociality Jointly Influence the Early and Later Stages of Neutral Face Processing. *Frontiers in Psychology*, *07*. https://doi.org/10.3389/fpsyg.2016.01258

doi: 10.53846/goediss-9841

Yan, X., Liu-Shuang, J., & Rossion, B. (2019). Effect of face-related task on rapid individual face discrimination. *Neuropsychologia*, *129*, 236–245. https://doi.org/10.1016/j.neuropsychologia.2019.04.002

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 121–134. https://doi.org/10.1037/0096-1523.16.1.121

Yiannikas, C., & Walsh, J. C. (1983). The variation of the pattern shift visual evoked response with the size of the stimulus field. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 427–435. https://doi.org/10.1016/0013-4694(83)90131-1

Yoon, K. L., & Zinbarg, R. E. (2008). Interpreting neutral faces as threatening is a default mode for socially anxious individuals. *Journal of Abnormal Psychology*, *117*(3), 680–685. https://doi.org/10.1037/0021-843x.117.3.680

Young, A. W. (2018). Faces, people and the brain: The 45th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *71*(3), 569–594. https://doi.org/10.1177/1747021817740275

Zandt, T. V. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/bf03214357

Zandt, T. V. (2002). Analysis of Response Time Distributions. In *Stevens' handbook of experimental psychology* (Vol. 4, pp. 461–516). https://doi.org/10.1002/0471214426.pas0412

Zani, A., & Proverbio, A. M. (1995). ERP signs of early selective attention effects to check size. *Electroencephalography and Clinical Neurophysiology*, *95*(4), 277–292. https://doi.org/10.1016/0013-4694(95)00078-d

Zeelenberg, R., & Bocanegra, B. R. (2010). Auditory emotional cues enhance visual perception. *Cognition*, *115*(1), 202–206. https://doi.org/10.1016/j.cognition.2009.12.004

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*, 233121651877717. https://doi.org/10.1177/2331216518777174

Zhang, J., & Rowe, J. B. (2014). Dissociable mechanisms of speed-accuracy tradeoff during visual perceptual learning are revealed by a hierarchical drift-diffusion model. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00069

Zhang, X., Li, Q., Sun, S., & Zuo, B. (2019). Facial expressions can inhibit the activation of gender stereotypes. *Cognition and Emotion*, *33*(7), 1424–1435. https://doi.org/10.1080/02699931.2019.1586648

Zhang, Y., Ding, Y., Huang, J., Zhou, W., Ling, Z., Hong, B., & Wang, X. (2021). Hierarchical

doi: 10.53846/goediss-9841

cortical networks of "voice patches" for processing voices in human brain. *Proceedings of the National Academy of Sciences*, *118*(52). https://doi.org/10.1073/pnas.2113887118

Zhao, Y., Sun, Q., Chen, G., & Yang, J. (2016). Hearing emotional sounds: Category representation in the human amygdala. *Social Neuroscience*, *13*(1), 117–128. https://doi.org/10.1080/17470919.2016.1267040

Zion-Golumbic, E., & Bentin, S. (2006). Dissociated neural mechanisms for face detection and configural encoding: Evidence from N170 and induced gamma-band oscillation effects. *Cerebral Cortex*, *17*(8), 1741–1749. https://doi.org/10.1093/cercor/bhl100

doi: 10.53846/goediss-9841

doi: 10.53846/goediss-9841

A Appendix of Study 1

Momentary affect (PANAS)

Besides the expected main effect of the negative vs. positive affect subscales (β_{neg} = -8.57, *SE* = 0.3, *t* = -28.99, *p* <.001), we found a trend between learning and test session (β_{test} = -0.57, *SE* = 0.3, *t* = -1.92, *p* = .053)[1]. Descriptively, there was a small difference of positive affect between pre- and post-assessment for the learning session ($M_{\text{pre-post}}$ = 2.46). However, none of the interaction terms reached significance (all *p* > .05). Experimenters inquired about participants' mood after each session, and while there were no noticeable changes for most participants, a few indicated that concentrating on both, the visual and auditory stimuli during the learning session was more challenging compared to the purely visual task in the test session. Participants were aware at all times that they could discontinue the study at any time without receiving any disadvantage.

Table A1
Momentary affect (PANAS) of the learning and test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p
(Intercept)	20.16	0.67	29.89	18.86	21.43	19.92	20.40	(Intercept)	-	-	-
prepostr_post-g.m	-0.43	0.30	-1.47	-1.04	0.12	-0.52	-0.37	prepostr	2.21	1	.137
affect_negative-g.m	-8.57	0.30	-28.99	-9.17	-8.00	-8.77	-8.29	affect	354.75	1	<.001
session_test-g.m	-0.57	0.30	-1.92	-1.16	0.03	-0.61	-0.48	session	3.75	1	.053
prepostr_post-g.m:affect_negative-g.m	0.13	0.30	0.44	-0.43	0.71	0.06	0.22	prepostr:affect	0.2	1	.658
prepostr_post-g.m:session_test-g.m	0.31	0.30	1.04	-0.26	0.85	0.24	0.38	prepostr:session	1.12	1	.290
affect_negative-g.m:session_test-g.m	0.00	0.30	-0.01	-0.56	0.57	-0.07	0.17	affect:session	0	1	.989
prepostr_post-g.m:affect_negative-g.m:session_test-g.m	-0.24	0.30	-0.81	-0.78	0.33	-0.30	-0.17	prepostr:affect:session	0.67	1	.413

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

[1] Please note, that for sum.contrast coding the estimates β reflect the difference to the group mean (e.g. for affect with two levels (pos. and neg.) it resembles half of the difference between positive and negative affect)

doi: 10.53846/goediss-9841

Learning session.

Table A2

Statistical results for the response times of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	758.50	28.24	26.86	702.65	812.53	750.05	766.49	(Intercept)	-	-	-	-
emotion_happy-g.m	25.63	6.79	3.78	11.65	38.68	23.52	27.45	emotion	15.49	2	<.001	0.10
emotion_angry-g.m	-5.66	6.79	-0.83	-18.29	8.05	-8.32	-3.92					
congruence_mismatch-g.m	44.97	4.80	9.37	36.06	54.29	42.18	46.63	congruence	71.84	1	<.001	0.57
emotion_happy-g.m:congruence_mismatch-g.m	4.22	6.79	0.62	-9.06	17.61	1.77	6.05	emotion:congruence	0.81	2	.667	0.01
emotion_angry-g.m:congruence_mismatch-g.m	-5.83	6.79	-0.86	-19.03	7.63	-8.56	-3.23					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A3

Statistical results for the P1 mean amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	3.77	0.53	7.05	2.72	4.85	3.57	3.96	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.05	0.07	-0.76	-0.20	0.08	-0.08	-0.01	emotion	3.91	2	.141	0.02
emotion_angry-g.m	-0.09	0.07	-1.19	-0.23	0.05	-0.12	-0.04					
congruence_mismatch-g.m	0.09	0.05	1.83	-0.01	0.19	0.08	0.11	congruence	3.41	1	.065	0.02
emotion_happy-g.m:congruence_mismatch-g.m	0.04	0.07	0.54	-0.10	0.18	0.02	0.06	emotion:congruence	3.6	2	.165	0.02
emotion_angry-g.m:congruence_mismatch-g.m	0.09	0.07	1.29	-0.05	0.24	0.06	0.12					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A4

Statistical results for the P1 peak amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	6.04	0.59	10.30	4.89	7.18	5.82	6.25	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.08	0.09	-0.90	-0.25	0.09	-0.11	-0.04	emotion	2.86	2	.239	0.02
emotion_angry-g.m	-0.07	0.09	-0.77	-0.25	0.11	-0.10	-0.02					
congruence_mismatch-g.m	0.05	0.06	0.81	-0.07	0.18	0.03	0.08	congruence	0.68	1	.409	0.00
emotion_happy-g.m:congruence_mismatch-g.m	0.00	0.09	0.05	-0.17	0.20	-0.01	0.04	emotion:congruence	4.05	2	.132	0.03
emotion_angry-g.m:congruence_mismatch-g.m	0.15	0.09	1.70	-0.02	0.33	0.11	0.20					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table A5

Statistical results for the N170 mean amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-5.48	0.57	-9.65	-6.59	-4.41	-5.63	-5.26	(Intercept)	-	-	-	-
emotion_happy-g.m	0.09	0.07	1.26	-0.05	0.23	0.07	0.11	emotion	5.75	2	.057	0.04
emotion_angry-g.m	0.08	0.07	1.12	-0.06	0.23	0.04	0.12					
congruence_mismatch-g.m	0.08	0.05	1.55	-0.02	0.18	0.06	0.10	congruence	2.47	1	.116	0.02
emotion_happy-g.m:congruence_mismatch-g.m	0.05	0.07	0.66	-0.09	0.19	0.02	0.08	emotion:congruence	2.6	2	.273	0.02
emotion_angry-g.m:congruence_mismatch-g.m	-0.12	0.07	-1.58	-0.26	0.03	-0.13	-0.10					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A6

Statistical results for the N170 peak amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-10.03	0.80	-12.54	-11.55	-8.44	-10.26	-9.69	(Intercept)	-	-	-	-
emotion_happy-g.m	0.11	0.09	1.17	-0.07	0.29	0.08	0.14	emotion	2.24	2	.327	0.01
emotion_angry-g.m	0.02	0.09	0.20	-0.16	0.21	-0.02	0.07					
congruence_mismatch-g.m	0.09	0.07	1.41	-0.03	0.23	0.06	0.13	congruence	2.03	1	.155	0.01
emotion_happy-g.m:congruence_mismatch-g.m	0.06	0.09	0.64	-0.13	0.24	0.02	0.09	emotion:congruence	0.82	2	.664	0.01
emotion_angry-g.m:congruence_mismatch-g.m	-0.08	0.09	-0.86	-0.27	0.11	-0.12	-0.06					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A7

Statistical results for the EPN mean amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	0.48	0.63	0.76	-0.79	1.75	0.20	0.76	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.06	0.10	-0.64	-0.25	0.12	-0.10	-0.04	emotion	0.73	2	.693	0.00
emotion_angry-g.m	-0.01	0.10	-0.15	-0.22	0.18	-0.05	0.02					
congruence_mismatch-g.m	0.19	0.07	2.72	0.06	0.32	0.15	0.21	congruence	7.49	1	.006	0.05
emotion_happy-g.m:congruence_mismatch-g.m	0.00	0.10	0.02	-0.20	0.19	-0.04	0.03	emotion:congruence	1.47	2	.480	0.01
emotion_angry-g.m:congruence_mismatch-g.m	-0.10	0.10	-1.05	-0.30	0.08	-0.13	-0.07					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A8

Statistical results for the LPC mean amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	3.03	0.39	7.86	2.29	3.81	2.90	3.16	(Intercept)	-	-	-	-
emotion_happy-g.m	0.06	0.09	0.73	-0.10	0.24	0.04	0.10	emotion	0.84	2	.658	0.01
emotion_angry-g.m	-0.07	0.09	-0.82	-0.24	0.10	-0.11	-0.04					
congruence_mismatch-g.m	-0.12	0.06	-2.03	-0.25	-0.01	-0.14	-0.10	congruence	4.21	1	.040	0.03
emotion_happy-g.m:congruence_mismatch-g.m	-0.01	0.09	-0.17	-0.19	0.16	-0.04	0.01	emotion:congruence	0.84	2	.657	0.01
emotion_angry-g.m:congruence_mismatch-g.m	-0.06	0.09	-0.68	-0.22	0.11	-0.09	-0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table A9

Statistical results for the auditory N1 mean amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.71	0.12	-5.86	-0.92	-0.46	-0.74	-0.66	(Intercept)	-	-	-	-
emotion_happy-g.m	0.02	0.04	0.45	-0.06	0.10	0.00	0.03	emotion	3.35	2	.187	0.02
emotion_angry-g.m	0.05	0.04	1.29	-0.03	0.13	0.04	0.07					
congruence_mismatch-g.m	0.00	0.03	-0.10	-0.06	0.05	-0.02	0.02	congruence	0.01	1	.918	0.00
emotion_happy-g.m:congruence_mismatch-g.m	0.00	0.04	0.04	-0.08	0.09	-0.01	0.02	emotion:congruence	2.98	2	.226	0.02
emotion_angry-g.m:congruence_mismatch-g.m	-0.06	0.04	-1.50	-0.15	0.02	-0.08	-0.04					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A10

Statistical results for the auditory P2 mean amplitudes of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	1.07	0.21	5.14	0.68	1.47	1.00	1.16	(Intercept)	-	-	-	-
emotion_happy-g.m	0.06	0.05	1.27	-0.04	0.15	0.04	0.08	emotion	6.3	2	.043	0.04
emotion_angry-g.m	0.06	0.05	1.22	-0.03	0.15	0.04	0.08					
congruence_mismatch-g.m	-0.19	0.03	-5.60	-0.26	-0.12	-0.21	-0.18	congruence	29.51	1	<.001	0.20
emotion_happy-g.m:congruence_mismatch-g.m	-0.06	0.05	-1.18	-0.15	0.03	-0.08	-0.04	emotion:congruence	1.44	2	.486	0.01
emotion_angry-g.m:congruence_mismatch-g.m	0.03	0.05	0.63	-0.06	0.12	0.01	0.05					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A11

Statistical results for the pupil size (constriction) of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-22.38	7.12	-3.14	-36.70	-9.47	-25.09	-18.39	(Intercept)	-	-	-	-
emotion_happy-g.m	0.41	1.46	0.28	-2.50	3.44	-0.56	1.00	emotion	3.69	2	.158	0.02
emotion_angry-g.m	-2.58	1.46	-1.77	-5.35	0.17	-3.24	-1.02					
congruence_mismatch-g.m	-2.45	1.03	-2.37	-4.46	-0.44	-2.96	-1.99	congruence	5.7	1	.017	0.04
emotion_happy-g.m:congruence_mismatch-g.m	-1.27	1.46	-0.87	-4.16	1.80	-1.86	-0.94	emotion:congruence	2.86	2	.240	0.02
emotion_angry-g.m:congruence_mismatch-g.m	-1.17	1.46	-0.80	-4.04	1.53	-1.57	-0.43					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A12

Statistical results for the pupil size (dilation) of the learning session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	87.66	14.03	6.25	60.85	113.12	81.52	92.62	(Intercept)	-	-	-	-
emotion_happy-g.m	3.02	2.65	1.14	-1.79	8.35	1.50	3.96	emotion	8.07	2	.018	0.05
emotion_angry-g.m	-7.47	2.65	-2.81	-12.64	-2.51	-8.11	-4.79					
congruence_mismatch-g.m	2.30	1.88	1.23	-1.47	5.94	1.42	3.00	congruence	1.55	1	.214	0.01
emotion_happy-g.m:congruence_mismatch-g.m	-2.65	2.65	-1.00	-7.63	2.63	-3.55	-1.74	emotion:congruence	5.16	2	.076	0.03
emotion_angry-g.m:congruence_mismatch-g.m	-3.32	2.65	-1.25	-8.48	2.07	-4.10	-2.22					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Test session.

Table A13

Statistical results for the response times of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	598.74	9.99	59.93	579.44	620.03	594.99	602.42	(Intercept)	-	-	-	-
emotion_happy-g.m	-2.58	1.59	-1.62	-5.56	0.63	-3.18	-1.81	emotion	2.7	2	.259	0.02
emotion_angry-g.m	1.23	1.59	0.77	-1.75	4.43	0.67	1.81					
congruence_mismatch-g.m	2.51	1.12	2.23	0.40	4.60	1.97	2.98	congruence	5.06	1	.024	0.03
emotion_happy-g.m:congruence_mismatch-g.m	-0.01	1.59	-0.01	-3.35	3.03	-0.82	0.44	emotion:congruence	0.73	2	.694	0.00
emotion_angry-g.m:congruence_mismatch-g.m	-1.15	1.59	-0.73	-4.24	2.22	-1.70	-0.69					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A14

Statistical results for the P1 mean amplitudes of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	5.14	0.58	8.78	3.93	6.33	4.91	5.36	(Intercept)	-	-	-	-
emotion_happy-g.m	0.08	0.08	1.05	-0.06	0.24	0.05	0.12	emotion	2.73	2	.255	0.02
emotion_angry-g.m	-0.13	0.08	-1.61	-0.29	0.02	-0.16	-0.09					
congruence_mismatch-g.m	0.14	0.06	2.49	0.03	0.24	0.11	0.15	congruence	6.3	1	.012	0.04
emotion_happy-g.m:congruence_mismatch-g.m	-0.01	0.08	-0.09	-0.16	0.15	-0.03	0.01	emotion:congruence	0.08	2	.963	0.00
emotion_angry-g.m:congruence_mismatch-g.m	0.02	0.08	0.27	-0.13	0.17	0.00	0.05					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A15

Statistical results for the P1 peak amplitudes of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	7.45	0.63	11.84	6.10	8.57	7.23	7.66	(Intercept)	-	-	-	-
emotion_happy-g.m	0.15	0.09	1.59	-0.03	0.32	0.12	0.18	emotion	3.52	2	.172	0.02
emotion_angry-g.m	-0.15	0.09	-1.62	-0.33	0.03	-0.18	-0.11					
congruence_mismatch-g.m	0.15	0.06	2.33	0.02	0.26	0.12	0.17	congruence	5.52	1	.019	0.04
emotion_happy-g.m:congruence_mismatch-g.m	-0.02	0.09	-0.16	-0.18	0.18	-0.04	0.01	emotion:congruence	0.19	2	.910	0.00
emotion_angry-g.m:congruence_mismatch-g.m	-0.02	0.09	-0.26	-0.21	0.15	-0.05	0.01					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table A16

Statistical results for the N170 mean amplitudes of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-5.74	0.68	-8.43	-7.03	-4.44	-5.94	-5.44	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.02	0.06	-0.34	-0.13	0.09	-0.05	0.01	emotion	1.7	2	.428	0.01
emotion_angry-g.m	-0.05	0.06	-0.90	-0.16	0.06	-0.08	-0.02					
congruence_mismatch-g.m	0.07	0.04	1.61	-0.02	0.15	0.05	0.08	congruence	2.66	1	.103	0.02
emotion_happy-g.m:congruence_mismatch-g.m	0.00	0.06	-0.02	-0.11	0.12	-0.02	0.02	emotion:congruence	6.16	2	.046	0.04
emotion_angry-g.m:congruence_mismatch-g.m	-0.12	0.06	-2.13	-0.23	-0.01	-0.14	-0.11					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A17

Statistical results for the N170 peak amplitudes of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-10.51	0.91	-11.49	-12.47	-8.64	-10.76	-10.04	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.01	0.07	-0.17	-0.16	0.13	-0.05	0.01	emotion	0.18	2	.914	0.00
emotion_angry-g.m	-0.02	0.07	-0.25	-0.16	0.12	-0.05	0.01					
congruence_mismatch-g.m	0.05	0.05	0.98	-0.05	0.16	0.04	0.07	congruence	0.98	1	.323	0.01
emotion_happy-g.m:congruence_mismatch-g.m	-0.02	0.07	-0.20	-0.16	0.13	-0.03	0.01	emotion:congruence	3.25	2	.197	0.02
emotion_angry-g.m:congruence_mismatch-g.m	-0.11	0.07	-1.43	-0.26	0.05	-0.13	-0.08					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A18

Statistical results for the EPN mean amplitudes of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.02	0.72	-0.03	-1.43	1.32	-0.45	0.28	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.04	0.08	-0.55	-0.19	0.11	-0.07	-0.01	emotion	1.59	2	.451	0.01
emotion_angry-g.m	-0.05	0.08	-0.70	-0.20	0.10	-0.08	-0.03					
congruence_mismatch-g.m	0.16	0.05	2.98	0.06	0.26	0.14	0.18	congruence	8.91	1	.003	0.06
emotion_happy-g.m:congruence_mismatch-g.m	0.03	0.08	0.45	-0.12	0.19	0.01	0.06	emotion:congruence	2.8	2	.247	0.02
emotion_angry-g.m:congruence_mismatch-g.m	-0.12	0.08	-1.60	-0.27	0.03	-0.15	-0.08					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A19

Statistical results for the LPC mean amplitudes of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	3.94	0.47	8.33	3.02	4.88	3.76	4.06	(Intercept)	-	-	-	-
emotion_happy-g.m	0.01	0.07	0.15	-0.14	0.14	-0.01	0.04	emotion	0.8	2	.669	0.01
emotion_angry-g.m	-0.06	0.07	-0.83	-0.20	0.08	-0.09	-0.03					
congruence_mismatch-g.m	-0.01	0.05	-0.20	-0.11	0.09	-0.05	0.01	congruence	0.04	1	.843	0.00
emotion_happy-g.m:congruence_mismatch-g.m	-0.04	0.07	-0.49	-0.18	0.11	-0.07	-0.01	emotion:congruence	3.33	2	.189	0.02
emotion_angry-g.m:congruence_mismatch-g.m	0.13	0.07	1.75	-0.02	0.28	0.10	0.16					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table A20

Statistical results for the pupil size (constriction) of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.84	9.68	0.50	-13.75	23.40	1.54	9.80	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.85	1.18	-0.72	-3.22	1.53	-1.21	-0.57	emotion	1.05	2	.591	0.01
emotion_angry-g.m	-0.29	1.18	-0.25	-2.58	1.98	-0.70	0.15					
congruence_mismatch-g.m	-0.09	0.83	-0.11	-1.70	1.47	-0.37	0.30	congruence	0.01	1	.911	0.00
emotion_happy-g.m:congruence_mismatch-g.m	0.14	1.18	0.12	-2.16	2.41	-0.28	0.66	emotion:congruence	1.23	2	.540	0.01
emotion_angry-g.m:congruence_mismatch-g.m	-1.18	1.18	-1.00	-3.53	1.13	-1.84	-0.66					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A21

Statistical results for the pupil size (dilation) of the test session

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	34.72	11.75	2.95	10.41	57.85	30.93	39.40	(Intercept)	-	-	-	-
emotion_happy-g.m	1.00	2.11	0.47	-3.03	5.20	0.20	1.67	emotion	0.37	2	.832	0.00
emotion_angry-g.m	-1.16	2.11	-0.55	-4.83	3.26	-2.36	-0.39					
congruence_mismatch-g.m	-0.39	1.49	-0.26	-3.12	2.38	-1.11	0.18	congruence	0.07	1	.791	0.00
emotion_happy-g.m:congruence_mismatch-g.m	0.93	2.11	0.44	-3.13	4.92	0.32	2.24	emotion:congruence	3.37	2	.185	0.02
emotion_angry-g.m:congruence_mismatch-g.m	-3.69	2.11	-1.75	-7.72	0.65	-4.23	-2.91					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table A22

Statistical results of the ordinal mixed model for the likability rating

	OR Full model	OR Old-New model
emotionangry	1.53 [0.84;2.79]	
emotionhappy	0.99 [0.54;1.81]	
congruencematch	1.41 [0.75;2.63]	
emotionangry:congruencematch	0.91 [0.38;2.17]	
emotionhappy:congruencematch	1.21 [0.51;2.91]	
oldnewnovel		0.45 [0.31;0.65]
1|2	0.01 [0.01;0.04]	0.01 [0;0.02]
2|3	0.08 [0.04;0.18]	0.06 [0.03;0.12]
3|4	0.42 [0.19;0.9]	0.3 [0.16;0.56]
4|5	1.41 [0.66;3.01]	1.03 [0.55;1.91]
5|6	7.29 [3.33;15.98]	5.6 [2.95;10.61]
6|7	49.75 [20.38;121.47]	35.31 [16.84;74.03]

Note: Model estimates and threshold coefficients in Odds Ratios (OR) of the ordinal models (left: full emotion $\times$ congruence model, right: familiar vs. novel faces).Brackets indicate the 95% confidence intervals of the OR.

doi: 10.53846/goediss-9841

Exploratory models for ERPs and pupil size over the course of the experiment

To model the possible non-linear learning function over the learning and extinction sessions, we used P-splines with different bin sizes of averaged trials across the session times (50, 25, 10). A backward stepwise generalized additive regression of location shape and scale (GAMLSS) was used to identify possible predictors of the ERP and pupil size outcomes. Each model started with full interaction model with the three-way-interaction of the fixed effects Emotion × Congruence × pb(Block) + random(participant). At each step, variables were dropped based on the generalized Akaike information criterion.

Table A23

EPN learning: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			8829.76	53664.82	53761.29
- emotion:congruence:pb(block)	2.00	1.94	8831.76	53666.77	53759.24
- congruence:pb(block)	1.00	0.35	8832.77	53667.11	53757.58
- emotion:congruence	2.01	3.92	8834.77	53671.04	53757.49

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

Table A24

EPN learning: GAMLSS model results for different block bin sizes

	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	-0.03 (0.26)	0.21 (0.19)	-0.07 (0.28)	-0.001 (0.17)	-0.12 (0.31)	0.04 (0.24)
emotionangry	-0.01 (0.37)	-0.48 (0.26)	0.02 (0.40)	0.13 (0.19)	-0.09 (0.44)	-0.32 (0.34)
emotionhappy	0.30 (0.37)	-0.09 (0.26)	0.29 (0.40)	-0.10 (0.19)	0.38 (0.44)	0.03 (0.34)
congruencemismatch	0.85 (0.37)	0.38 (0.11)	0.89 (0.40)	0.60 (0.19)	0.90 (0.44)	0.59 (0.20)
pb(block)	0.01 (0.009)	0.007 (0.006)	0.03 (0.02)	0.02 (0.008)	0.06 (0.05)	0.04 (0.04)
emotionangry × congruencemismatch	-0.94 (0.53)		-1.03 (0.56)	-0.55 (0.27)	-1.00 (0.62)	-0.54 (0.29)
emotionhappy × congruencemismatch	-0.80 (0.52)		-0.64 (0.56)	-0.14 (0.27)	-0.78 (0.62)	-0.08 (0.29)
emotionangry × pb(block)	0.007 (0.01)	0.02 (0.009)	0.008 (0.03)		0.05 (0.07)	0.09 (0.05)
emotionhappy × pb(block)	-0.01 (0.01)	-0.002 (0.009)	-0.03 (0.03)		-0.08 (0.07)	-0.02 (0.05)
congruencemismatch × pb(block)	-0.01 (0.01)		-0.02 (0.03)		-0.06 (0.07)	
emotionangry × congruencemismatch × pb(block)	0.02 (0.02)		0.04 (0.04)		0.08 (0.10)	
emotionhappy × congruencemismatch × pb(block)	0.02 (0.02)		0.04 (0.04)		0.13 (0.10)	
(Intercept):sigma	1.60 (0.008)	1.60 (0.008)	1.34 (0.01)	1.34 (0.01)	0.94 (0.02)	0.94 (0.02)
Num.Obs.	8878	8878	4706	4706	1914	1914
AIC	53761.3	53757.5	26027.0	26022.4	9115.5	9111.3
BIC	54103.3	54064.0	26337.9	26301.0	9382.1	9361.2
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Note: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

doi: 10.53846/goediss-9841

Table A25

LPC learning: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			8833.67	54478.08	54566.74
- emotion:congruence:pb(block)	2	1.28	8835.67	54479.36	54564.02
- emotion:congruence	2	1.52	8837.67	54480.87	54561.54
- congruence:pb(block)	1	0.55	8838.67	54481.43	54560.09
- emotion:pb(block)	2	3.82	8840.67	54485.25	54559.91
- emotion	2	0.74	8842.67	54485.99	54556.65

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

Table A26

LPC learning: GAMLSS model results for different block bin sizes

	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	2.49 (0.27)	2.54 (0.13)	2.48 (0.30)	2.49 (0.14)	2.51 (0.33)	2.75 (0.24)
emotionangry	-0.38 (0.39)		-0.54 (0.43)		-0.68 (0.46)	-0.73 (0.33)
emotionhappy	0.28 (0.39)		0.30 (0.42)		0.21 (0.46)	-0.15 (0.33)
congruencemismatch	0.21 (0.39)	-0.26 (0.11)	0.19 (0.42)	-0.27 (0.12)	0.29 (0.46)	-0.20 (0.12)
pb(block)	0.02 (0.009)	0.02 (0.004)	0.05 (0.02)	0.05 (0.008)	0.11 (0.05)	0.08 (0.04)
emotionangry × congruencemismatch	-0.31 (0.55)		-0.12 (0.60)		-0.11 (0.65)	
emotionhappy × congruencemismatch	-0.66 (0.55)		-0.72 (0.60)		-0.73 (0.65)	
emotionangry × pb(block)	0.02 (0.01)		0.04 (0.03)		0.12 (0.07)	0.11 (0.05)
emotionhappy × pb(block)	-0.007 (0.01)		-0.01 (0.03)		-0.03 (0.07)	0.03 (0.05)
congruencemismatch × pb(block)	-0.01 (0.01)		-0.02 (0.03)		-0.06 (0.07)	
emotionangry × congruencemismatch × pb(block)	-0.0009 (0.02)		-0.02 (0.04)		-0.03 (0.10)	
emotionhappy × congruencemismatch × pb(block)	0.02 (0.02)		0.04 (0.04)		0.10 (0.10)	
(Intercept):sigma	1.65 (0.008)	1.65 (0.008)	1.41 (0.01)	1.41 (0.01)	0.99 (0.02)	0.99 (0.02)
Num.Obs.	8878	8878	4706	4706	1914	1914
AIC	54566.7	54556.7	26681.9	26673.0	9300.5	9293.7
BIC	54881.1	54807.2	26967.5	26900.4	9546.0	9511.4
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

doi: 10.53846/goediss-9841

Table A27

P2 learning: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			9119.83	41558.46	41650.81
- emotion:congruence:pb(block)	1.99	0.95	9121.82	41559.42	41647.78
- congruence:pb(block)	1.00	0.27	9122.82	41559.69	41646.04
- emotion:pb(block)	2.01	2.41	9124.83	41562.10	41644.44
- emotion:congruence	1.99	3.30	9126.81	41565.40	41643.77

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

Table A28

P2 learning: GAMLSS model results for different block bin sizes

	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	1.06 (0.12)	1.06 (0.07)	1.04 (0.13)	1.02 (0.07)	1.03 (0.15)	1.03 (0.08)
emotionangry	0.19 (0.17)	0.17 (0.06)	0.21 (0.19)	0.20 (0.06)	0.19 (0.21)	0.18 (0.07)
emotionhappy	0.22 (0.17)	0.19 (0.06)	0.23 (0.19)	0.20 (0.06)	0.26 (0.21)	0.17 (0.07)
congruencemismatch	-0.26 (0.17)	-0.40 (0.05)	-0.27 (0.19)	-0.39 (0.05)	-0.26 (0.21)	-0.40 (0.06)
pb(block)	0.002 (0.004)	0.004 (0.002)	0.005 (0.009)	0.009 (0.004)	0.02 (0.02)	0.02 (0.01)
emotionangry × congruencemismatch	-0.16 (0.24)		-0.16 (0.27)		-0.15 (0.30)	
emotionhappy × congruencemismatch	-0.39 (0.24)		-0.37 (0.27)		-0.45 (0.30)	
emotionangry × pb(block)	-0.0007 (0.006)		-0.001 (0.01)		-0.0001 (0.03)	
emotionhappy × pb(block)	0.003 (0.006)		0.004 (0.01)		0.004 (0.03)	
congruencemismatch × pb(block)	-0.003 (0.006)		-0.004 (0.01)		-0.01 (0.03)	
emotionangry × congruencemismatch × pb(block)	0.006 (0.008)		0.01 (0.02)		0.02 (0.05)	
emotionhappy × congruencemismatch × pb(block)	0.008 (0.008)		0.02 (0.02)		0.04 (0.05)	
(Intercept):sigma	0.85 (0.007)	0.85 (0.007)	0.59 (0.01)	0.59 (0.01)	0.21 (0.02)	0.21 (0.02)
Num.Obs.	9166	9166	4740	4740	1916	1916
AIC	41650.8	41643.8	19135.2	19126.8	6329.1	6320.5
BIC	41979.7	41922.9	19429.8	19376.1	6575.7	6528.3
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

doi: 10.53846/goediss-9841

Table A29

Pupil size (T1) learning: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			9359.83	109977.8	110084.2
- emotion:congruence:pb(block)	2.00	2.17	9361.83	109980.0	110082.3
- emotion:pb(block)	2.00	1.79	9363.83	109981.8	110080.1
- congruence:pb(block)	1.01	0.81	9364.84	109982.6	110078.9
- emotion:congruence	2.00	3.18	9366.84	109985.8	110078.1

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

Table A30

Pupil T1 (learning) GAMLSS model results for different block bin sizes

	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	-17.30 (4.28)	-16.37 (2.30)	-17.00 (4.51)	-16.66 (2.40)	-16.58 (5.00)	-16.24 (2.62)
emotionangry	-5.38 (6.08)	-4.69 (2.11)	-5.51 (6.38)	-4.29 (2.18)	-5.87 (7.06)	-4.67 (2.31)
emotionhappy	-1.92 (6.05)	-1.37 (2.10)	-3.06 (6.37)	-1.18 (2.18)	-3.93 (7.06)	-1.56 (2.31)
congruencemismatch	-0.91 (6.07)	-3.81 (1.72)	-2.36 (6.38)	-4.13 (1.78)	-1.48 (7.08)	-3.93 (1.89)
pb(block)	-0.11 (0.15)	-0.07 (0.06)	-0.27 (0.30)	-0.14 (0.12)	-0.70 (0.81)	-0.39 (0.33)
emotionangry × congruencemismatch	4.69 (8.59)		5.91 (9.02)		6.09 (10.00)	
emotionhappy × congruencemismatch	-5.36 (8.58)		-2.98 (9.02)		-4.24 (10.00)	
emotionangry × pb(block)	0.13 (0.21)		0.30 (0.43)		0.72 (1.14)	
emotionhappy × pb(block)	0.16 (0.21)		0.41 (0.43)		1.06 (1.14)	
congruencemismatch × pb(block)	0.05 (0.21)		0.18 (0.43)		0.31 (1.14)	
emotionangry × congruencemismatch × pb(block)	-0.40 (0.29)		-0.87 (0.61)		-2.11 (1.61)	
emotionhappy × congruencemismatch × pb(block)	-0.07 (0.29)		-0.31 (0.61)		-0.49 (1.61)	
(Intercept):sigma	4.42 (0.007)	4.42 (0.007)	4.12 (0.01)	4.12 (0.01)	3.72 (0.02)	3.72 (0.02)
Num.Obs.	9413	9413	4773	4773	1916	1916
AIC	110084.2	110078.1	52962.0	52955.6	19788.3	19781.5
BIC	110464.4	110408.1	53298.7	53246.7	20051.9	20006.2
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

doi: 10.53846/goediss-9841

Table A31

Pupil size (T2) learning: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			9390.52	120677.5	120770.5
- emotion:congruence:pb(block)	2	2.71	9392.51	120680.2	120769.2
- emotion:pb(block)	2	3.95	9394.52	120684.2	120769.2

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

Table A32

Pupil T2 (learning) GAMLSS model results for different block bin sizes

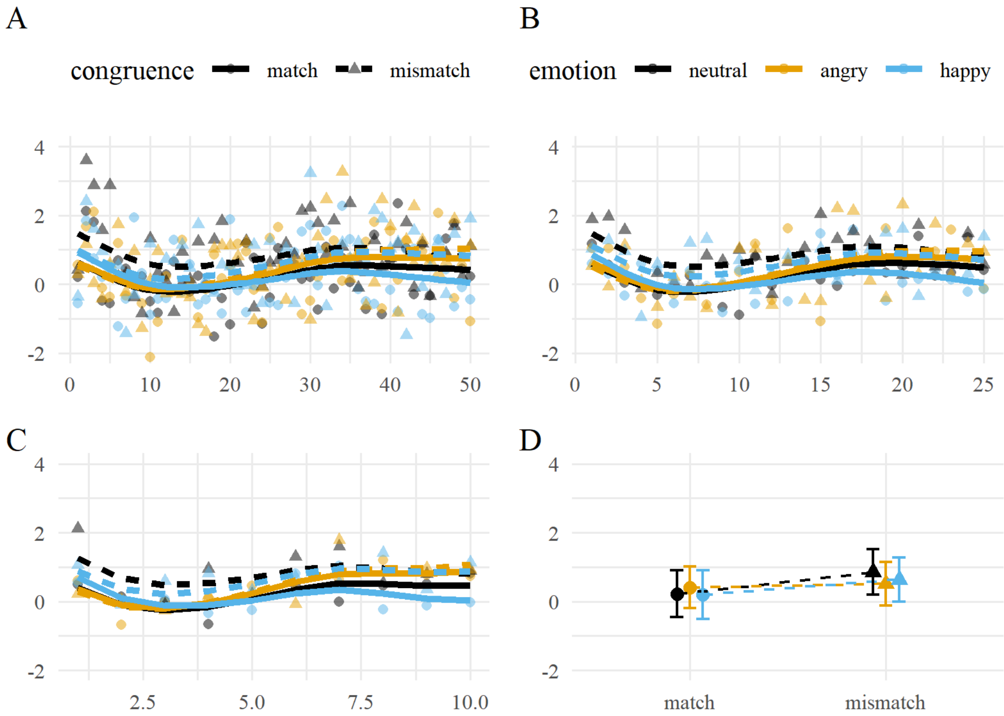
	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	137.11 (7.45)	135.92 (5.23)	138.03 (8.02)	136.05 (5.61)	140.56 (9.15)	136.50 (7.86)
emotionangry	-5.24 (10.55)	-5.11 (5.16)	-5.18 (11.33)	-3.43 (5.48)	-4.48 (12.92)	6.93 (10.07)
emotionhappy	5.37 (10.51)	8.81 (5.16)	5.72 (11.32)	9.88 (5.47)	3.17 (12.92)	3.89 (10.07)
congruencemismatch	17.99 (10.53)	24.53 (7.40)	17.20 (11.34)	25.11 (7.93)	18.79 (12.94)	26.89 (8.93)
pb(block)	-1.82 (0.25)	-1.77 (0.15)	-3.70 (0.54)	-3.55 (0.31)	-9.16 (1.47)	-8.42 (1.20)
emotionangry × congruencemismatch	2.49 (14.91)	-16.48 (7.31)	3.64 (16.03)	-17.60 (7.75)	5.17 (18.28)	-17.70 (8.46)
emotionhappy × congruencemismatch	-18.19 (14.87)	-18.99 (7.29)	-17.49 (16.02)	-19.94 (7.75)	-18.80 (18.28)	-20.23 (8.46)
emotionangry × pb(block)	0.005 (0.36)		0.13 (0.76)		0.05 (2.08)	-2.02 (1.47)
emotionhappy × pb(block)	0.13 (0.36)		0.32 (0.76)		1.13 (2.08)	1.00 (1.47)
congruencemismatch × pb(block)	-0.11 (0.36)	-0.36 (0.21)	-0.11 (0.76)	-0.72 (0.44)	-0.53 (2.08)	-2.00 (1.20)
emotionangry × congruencemismatch × pb(block)	-0.74 (0.51)		-1.63 (1.08)		-4.15 (2.94)	
emotionhappy × congruencemismatch × pb(block)	-0.03 (0.51)		-0.19 (1.08)		-0.26 (2.94)	
(Intercept):sigma	4.97 (0.007)	4.98 (0.007)	4.69 (0.01)	4.69 (0.01)	4.32 (0.02)	4.32 (0.02)
Num.Obs.	9437	9437	4775	4775	1916	1916
AIC	120770.5	120769.2	58465.1	58462.9	22098.5	22097.0
BIC	121103.0	121073.0	58764.1	58736.0	22353.5	22340.8
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

doi: 10.53846/goediss-9841

Figure A1

EPN amplitudes by block bins: Learning.

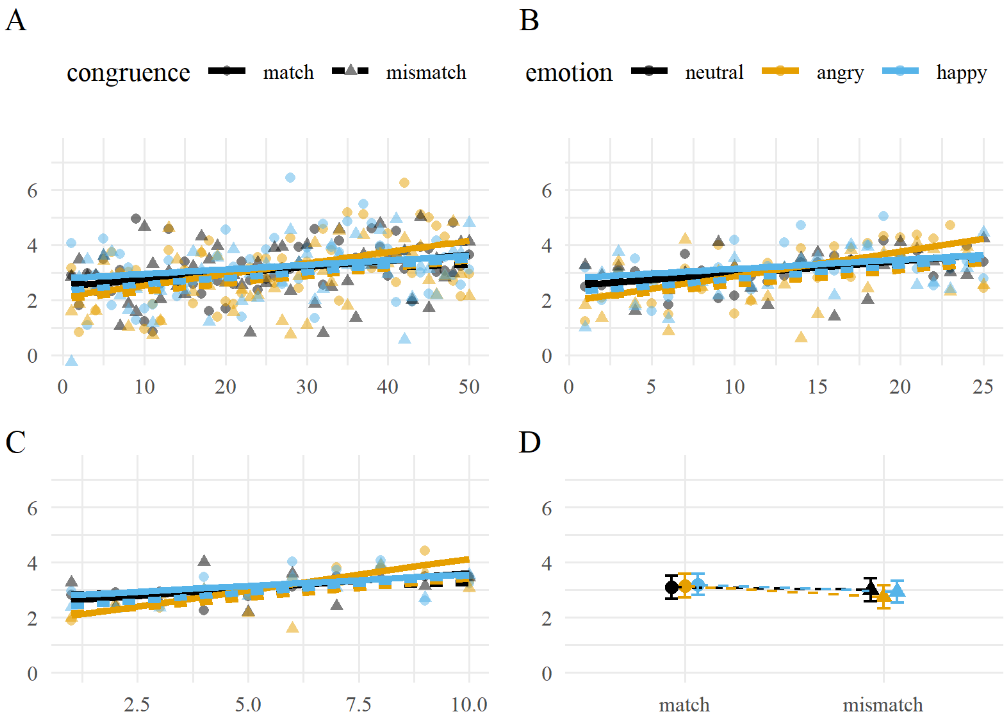


Notes: Face-locked EPN over the course of the learning session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in μV per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in μV of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages in μV of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole learning session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

Figure A2

LPC amplitudes by block bins: Learning.

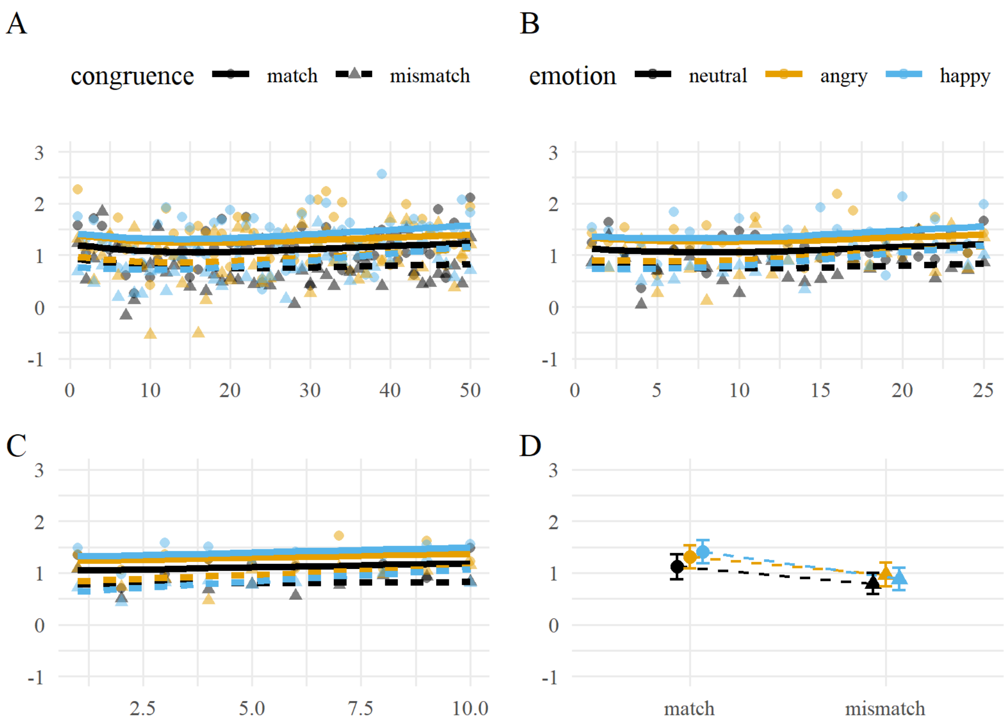


Notes: Face-locked LPC over the course of the learning session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in μV per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in μV of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages in μV of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole learning session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

Figure A3

P2 amplitudes by block bins: Learning.

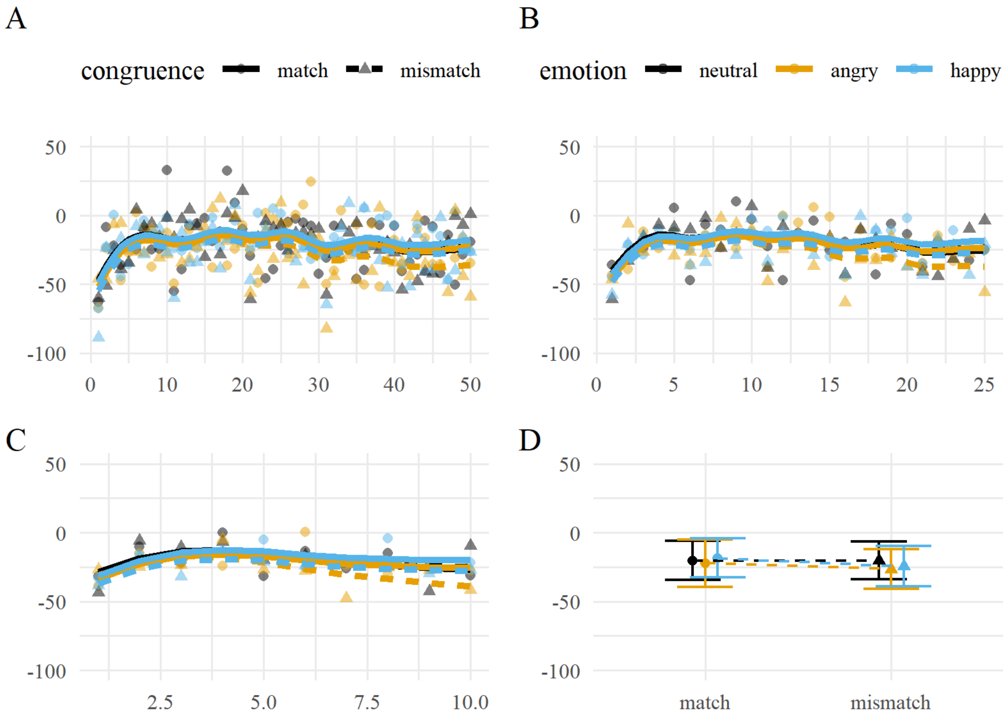


Notes: Voice-locked P2 over the course of the learning session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in μV per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in μV of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages in μV of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole learning session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

Figure A4

Pupil size (constriction) by block bins: Learning.

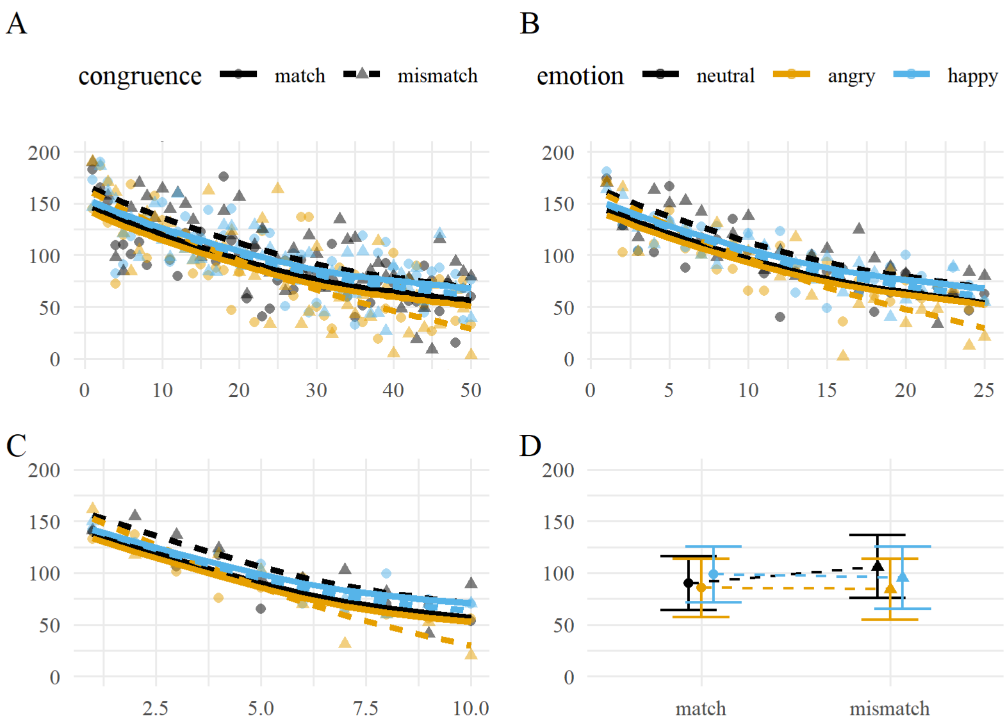


Notes: Pupilsize T1 (Constriction)over the course of the learning session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in pixel per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in pixel of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages inpixel of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole learning session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

Figure A5

Pupil size (dilation) by block bins: Learning.



Notes: Pupilsize T2 (Dilation)over the course of the learning session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in pixel per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in pixel of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages inpixel of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole learning session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

Table A33
P1 mean amplitudes test: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			8952.28	54865.56	54955

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

Table A34
P1 mean amplitudes test: GAMLSS model results for different block bin sizes

	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	4.64 (0.27)	4.64 (0.27)	4.64 (0.28)	4.64 (0.28)	4.70 (0.29)	4.70 (0.29)
emotionangry	-0.07 (0.38)	-0.07 (0.38)	-0.03 (0.40)	-0.03 (0.40)	-0.11 (0.41)	-0.11 (0.41)
emotionhappy	0.51 (0.38)	0.51 (0.38)	0.59 (0.40)	0.59 (0.40)	0.59 (0.41)	0.59 (0.41)
congruencemismatch	0.84 (0.38)	0.84 (0.38)	0.97 (0.40)	0.97 (0.40)	0.80 (0.41)	0.80 (0.41)
pb(block)	0.02 (0.009)	0.02 (0.009)	0.04 (0.02)	0.04 (0.02)	0.07 (0.05)	0.07 (0.05)
emotionangry × congruencemismatch	-0.33 (0.54)	-0.33 (0.54)	-0.50 (0.56)	-0.50 (0.56)	-0.38 (0.58)	-0.38 (0.58)
emotionhappy × congruencemismatch	-1.04 (0.54)	-1.04 (0.54)	-1.28 (0.56)	-1.28 (0.56)	-1.17 (0.58)	-1.17 (0.58)
emotionangry × pb(block)	-0.004 (0.01)	-0.004 (0.01)	-0.009 (0.03)	-0.009 (0.03)	-0.02 (0.07)	-0.02 (0.07)
emotionhappy × pb(block)	-0.02 (0.01)	-0.02 (0.01)	-0.04 (0.03)	-0.04 (0.03)	-0.09 (0.07)	-0.09 (0.07)
congruencemismatch × pb(block)	-0.02 (0.01)	-0.02 (0.01)	-0.06 (0.03)	-0.06 (0.03)	-0.11 (0.07)	-0.11 (0.07)
emotionangry × congruencemismatch × pb(block)	0.01 (0.02)	0.01 (0.02)	0.03 (0.04)	0.03 (0.04)	0.07 (0.09)	0.07 (0.09)
emotionhappy × congruencemismatch × pb(block)	0.04 (0.02)	0.04 (0.02)	0.09 (0.04)	0.09 (0.04)	0.20 (0.09)	0.20 (0.09)
(Intercept):sigma	1.63 (0.007)	1.63 (0.007)	1.34 (0.01)	1.34 (0.01)	0.88 (0.02)	0.88 (0.02)
Num.Obs.	8997	8997	4733	4733	1915	1915
AIC	54955.0	54955.0	26210.3	26210.3	8879.1	8879.1
BIC	55272.7	55272.7	26499.1	26499.1	9127.5	9127.5
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

Table A35
N170 mean amplitudes test: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			8948.38	51881.01	51978.24

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

doi: 10.53846/goediss-9841

Table A36

N170 mean amplitudes test: GAMLSS model results for different block bin sizes

	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	-6.14 (0.23)	-6.14 (0.23)	-6.19 (0.24)	-6.19 (0.24)	-6.13 (0.26)	-6.13 (0.26)
emotionangry	0.66 (0.32)	0.66 (0.32)	0.71 (0.34)	0.71 (0.34)	0.73 (0.37)	0.73 (0.37)
emotionhappy	0.41 (0.32)	0.41 (0.32)	0.48 (0.34)	0.48 (0.34)	0.53 (0.37)	0.53 (0.37)
congruencemismatch	0.85 (0.32)	0.85 (0.32)	0.96 (0.34)	0.96 (0.34)	0.88 (0.37)	0.88 (0.37)
pb(block)	0.009 (0.008)	0.009 (0.008)	0.02 (0.02)	0.02 (0.02)	0.04 (0.04)	0.04 (0.04)
emotionangry × congruencemismatch	-1.25 (0.46)	-1.25 (0.46)	-1.30 (0.48)	-1.30 (0.48)	-1.15 (0.52)	-1.15 (0.52)
emotionhappy × congruencemismatch	-1.37 (0.46)	-1.37 (0.46)	-1.56 (0.48)	-1.56 (0.48)	-1.53 (0.52)	-1.53 (0.52)
emotionangry × pb(block)	-0.02 (0.01)	-0.02 (0.01)	-0.04 (0.02)	-0.04 (0.02)	-0.09 (0.06)	-0.09 (0.06)
emotionhappy × pb(block)	-0.01 (0.01)	-0.01 (0.01)	-0.03 (0.02)	-0.03 (0.02)	-0.07 (0.06)	-0.07 (0.06)
congruencemismatch × pb(block)	-0.01 (0.01)	-0.01 (0.01)	-0.04 (0.02)	-0.04 (0.02)	-0.07 (0.06)	-0.07 (0.06)
emotionangry × congruencemismatch × pb(block)	0.02 (0.02)	0.02 (0.02)	0.04 (0.03)	0.04 (0.03)	0.08 (0.08)	0.08 (0.08)
emotionhappy × congruencemismatch × pb(block)	0.04 (0.02)	0.04 (0.02)	0.09 (0.03)	0.09 (0.03)	0.21 (0.08)	0.21 (0.08)
(Intercept):sigma	1.46 (0.007)	1.46 (0.007)	1.17 (0.01)	1.17 (0.01)	0.77 (0.02)	0.77 (0.02)
Num.Obs.	8997	8997	4733	4733	1915	1915
AIC	51978.2	51978.2	24604.9	24604.9	8467.7	8467.7
BIC	52323.7	52323.7	24917.5	24917.5	8735.9	8735.9
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

Table A37

EPN mean amplitudes test: Stepwise model selection of the GAMLSS model including 50 blocks

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
			8948.86	54050.90	54147.19
- emotion:congruence:pb(block)	2	1.73	8950.86	54052.63	54144.92
- emotion:pb(block)	2	0.63	8952.85	54053.25	54141.55
- congruence:pb(block)	1	0.36	8953.86	54053.62	54139.91

Note: The model selection started with the full model and dropped each term unitl the smalles GAIC value was reached. Each line shows model properties if this predictor would be removed.

doi: 10.53846/goediss-9841

Table A38
EPN mean amplitudes test: GAMLSS model results for different block bin sizes

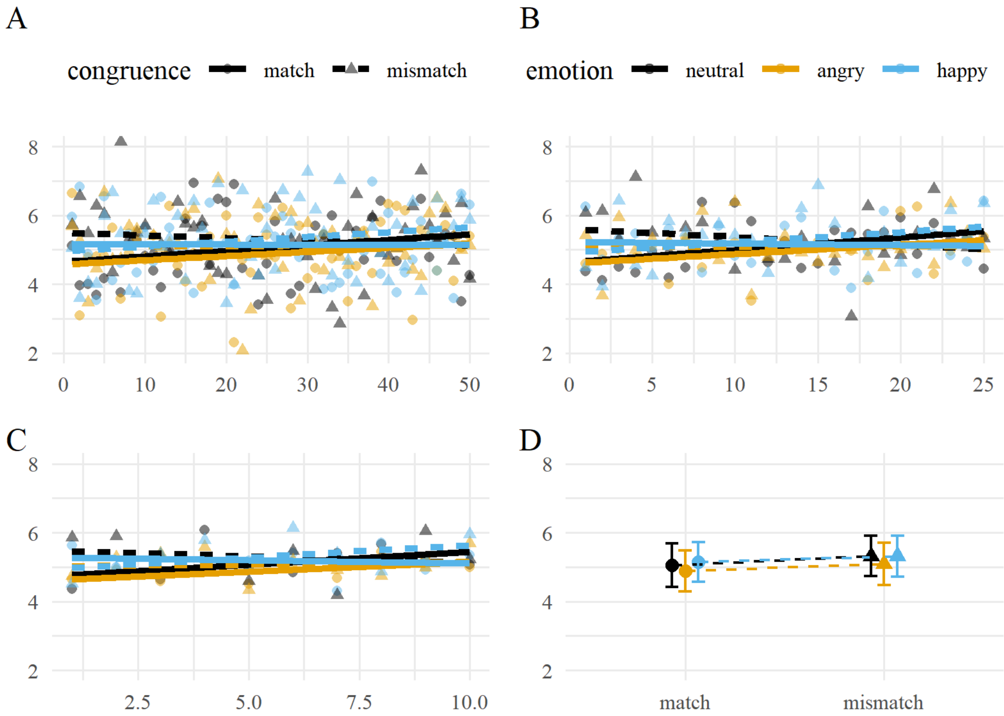
	Full mod 50	Red mod 50	Full mod 25	Red mod 25	Full mod 10	Red mod 10
(Intercept)	0.05 (0.26)	0.11 (0.16)	-0.07 (0.27)	0.11 (0.16)	0.11 (0.30)	0.18 (0.17)
emotionangry	0.34 (0.36)	0.07 (0.18)	0.55 (0.38)	0.11 (0.18)	0.43 (0.42)	0.13 (0.19)
emotionhappy	-0.01 (0.37)	-0.11 (0.18)	0.17 (0.38)	-0.09 (0.18)	0.13 (0.42)	-0.05 (0.19)
congruencemismatch	0.82 (0.36)	0.59 (0.18)	0.96 (0.38)	0.62 (0.18)	0.79 (0.42)	0.60 (0.19)
pb(block)	-0.009 (0.009)	-0.01 (0.004)	-0.01 (0.02)	-0.03 (0.007)	-0.06 (0.05)	-0.07 (0.02)
emotionangry × congruencemismatch	-1.14 (0.52)	-0.61 (0.25)	-1.35 (0.54)	-0.67 (0.26)	-1.24 (0.59)	-0.68 (0.27)
emotionhappy × congruencemismatch	-0.66 (0.52)	-0.17 (0.25)	-0.85 (0.54)	-0.21 (0.26)	-0.73 (0.59)	-0.21 (0.27)
emotionangry × pb(block)	-0.01 (0.01)		-0.03 (0.03)		-0.05 (0.07)	
emotionhappy × pb(block)	-0.004 (0.01)		-0.02 (0.03)		-0.03 (0.07)	
congruencemismatch × pb(block)	-0.009 (0.01)		-0.03 (0.03)		-0.03 (0.07)	
emotionangry × congruencemismatch × pb(block)	0.02 (0.02)		0.05 (0.04)		0.10 (0.10)	
emotionhappy × congruencemismatch × pb(block)	0.02 (0.02)		0.05 (0.04)		0.09 (0.10)	
(Intercept):sigma	1.58 (0.007)	1.59 (0.007)	1.30 (0.01)	1.30 (0.01)	0.89 (0.02)	0.89 (0.02)
Num.Obs.	8997	8997	4733	4733	1915	1915
AIC	54147.2	54139.9	25812.8	25806.1	8943.1	8935.4
BIC	54489.2	54446.4	26123.3	26084.3	9209.1	9173.6
RMSE	1.00	1.00	1.00	1.00	1.00	1.00

Notes: Model estimates (standard errors) of the gamlss models for different bin sizes of blocks of the experimental session. The full model refers to a model including all main effects and interactions and the reduced model to the model with the lowest GAIC.

doi: 10.53846/goediss-9841

Figure A6

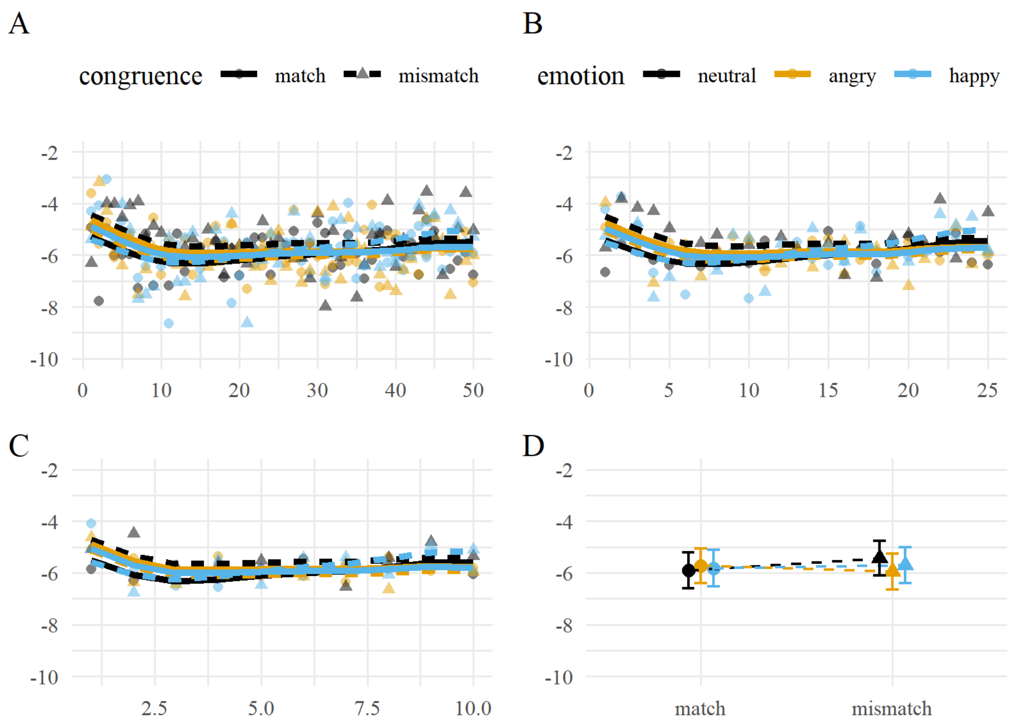
P1 amplitudes by block bins: Test.



Notes: Face-locked P1 over the course of the test session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in μV per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in μV of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages in μV of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole test session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

Figure A7
N170 amplitudes by block bins: Test.

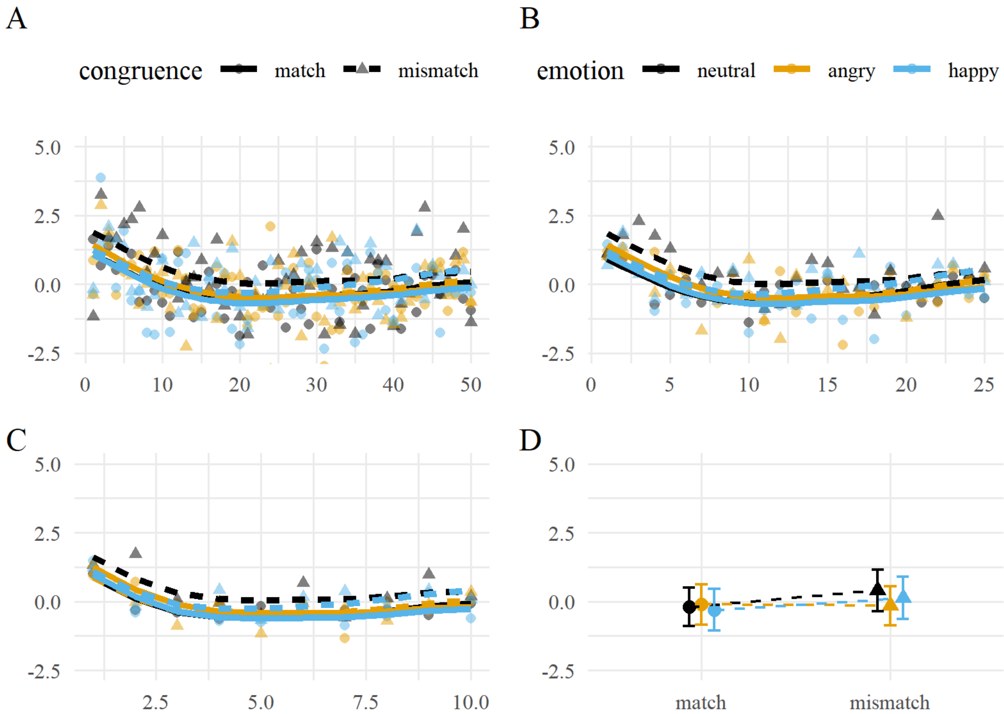


Notes: Face-locked N170 over the course of the test session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in μV per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in μV of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages in μV of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole test session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

Figure A8

EPN amplitudes by block bins: Test.



Notes: Face-locked EPN over the course of the test session. The curves in the panels **A-C** represent the fitted values of the full- GAMLSS model. **A** Dots represent the grand averages in μV per repetition block (50 bins) of each stimulus and condition. **B** Dots represent the grand averages in μV of two repetition blocks each (25 bins) and condition. **C** Dots represent the grand averages in μV of five repetition blocks each (10 bins) and condition. **D** Mean amplitudes averaged over the whole test session. Dots represent the grand averages per condition.

doi: 10.53846/goediss-9841

B Appendix of Study 2

Table B1

Identification codes of stimuli used in Experiment 1 and Experiment 2.

faces_Exp1	voices_Exp1	faces_Exp2	voices_Exp
010ff32y	42_neutral	010mf24n	6_neutral
107mf23n	45_neutral	037mf25n	53_neutral
030ff20n	60_neutral	042mf24y	58_neutral
039mf26n	61_neutral	068mf19y	59_neutral
039ff20y	6_happiness	071mf20n	55_happiness
113mf23y	42_hapiness	073mf22y	58_happiness
076mf24n	46_happiness	107mf23n	59_happiness
162ff20n	58_happiness	152mf23y	60_happiness
030mf20n	6_anger	001ff20y	45_anger
086mf28y	45_anger	042ff21n	58_anger
113ff24y	59_anger	044ff22n	59_anger
026ff28n	60_anger	072ff26y	61_anger
		074ff23y	
		128ff26n	
		135ff23n	
		122ff19y	

Notes: Face stimuli were selected from the Goettingen Faces Database (Kulke et al., 2017) and vocal stimuli selected from the Montreal Affective Voices database (Belin et al., 2008). Please note that in Exp 2, face-voice pairs were pseudorandomized (keeping the gender structure). 12 out of 16 face voice pairs were sampled as conditioned stimuli, whereas the remaining 4 faces were used in a different task as novel stimuli (not used in the current study).

doi: 10.53846/goediss-9841

Table B2

Statistical results of the full models of the learning sessions for Experiment 1 and Experiment 2.

		Exp 1							Exp 2						
		$\beta\,(SE)$	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	t	p	$\beta\,(SE)$	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	t	p
mu	(Intercept)	550.79 (3.14)	544.63	556.95	543.50	582.41	140.17	<.001	462.27 (2.51)	457.35	467.18	457.42	466.32	168.39	<.001
	angry	39.10 (4.60)	30.09	48.11	36.39	42.58	8.51	<.001	11.78 (3.53)	4.86	18.69	9.48	13.92	3.34	<.001
	happy	63.55 (4.71)	54.33	72.77	59.40	65.64	13.51	<.001	13.37 (3.48)	6.54	20.19	11.43	15.75	3.84	<.001
	mismatch	32.56 (4.50)	23.74	41.37	28.68	35.85	7.24	<.001	45.97 (3.75)	38.62	53.32	43.09	48.15	12.27	<.001
	pb(block.z)	-63.23 (3.00)	-69.12	-57.35	-65.79	-56.61	-21.06	<.001	-47.81 (1.93)	-51.59	-44.03	-49.25	-44.66	-24.79	<.001
	angry:mismatch	-27.95 (6.44)	-40.56	-15.33	-31.54	-24.77	-4.34	<.001	1.97 (5.32)	-8.46	12.40	-2.05	6.21	0.37	.711
	happy:mismatch	-23.48 (6.41)	-36.04	-10.92	-26.49	-16.21	-3.67	<.001	0.13 (5.31)	-10.27	10.54	-3.31	2.15	0.03	.980
	angry:pb(block.z)	-16.19 (4.71)	-25.42	-6.95	-18.54	-13.72	-3.44	<.001	-6.35 (2.83)	-11.90	-0.80	-7.69	-5.64	-2.24	.025
	happy:pb(block.z)	-6.01 (4.53)	-14.89	2.87	-9.53	-2.88	-1.33	.184	-8.46 (2.78)	-13.91	-3.00	-9.82	-7.46	-3.04	.002
	mismatch:pb(block.z)	11.31 (4.28)	2.93	19.70	8.06	13.49	2.65	.008	-12.06 (2.88)	-17.72	-6.41	-13.87	-9.71	-4.18	<.001
	angry:mismatch:pb(block.z)	8.35 (6.36)	-4.11	20.82	4.46	11.75	1.31	.189	6.34 (4.22)	-1.93	14.62	4.36	9.62	1.50	.133
	happy:mismatch:pb(block.z)	-1.45 (6.19)	-13.58	10.68	-5.93	1.71	-0.23	.815	2.56 (4.14)	-5.56	10.67	-0.06	5.61	0.62	.537
sigma	(Intercept)	4.16 (0.04)	4.08	4.24	4.09	4.20	97.50	<.001	3.92 (0.04)	3.85	3.99	3.86	3.96	105.73	<.001
	angry	0.13 (0.06)	0.02	0.24	0.03	0.18	2.23	.026	0.00 (0.06)	-0.11	0.11	-0.08	0.05	0.04	.965
	happy	0.29 (0.06)	0.16	0.41	0.24	0.34	4.62	<.001	-0.07 (0.06)	-0.18	0.04	-0.10	-0.04	-1.23	.221
	mismatch	0.07 (0.06)	-0.04	0.19	-0.04	0.12	1.26	.209	0.23 (0.05)	0.13	0.34	0.21	0.27	4.27	<.001
	pb(block.z)	-0.37 (0.04)	-0.45	-0.28	-0.41	-0.34	-8.54	<.001	-0.55 (0.03)	-0.62	-0.48	-0.59	-0.52	-15.87	<.001
	angry:mismatch	-0.22 (0.08)	-0.39	-0.06	-0.30	-0.10	-2.72	.007	0.03 (0.08)	-0.12	0.19	-0.01	0.14	0.44	.662
	happy:mismatch	-0.40 (0.08)	-0.56	-0.24	-0.48	-0.27	-4.83	<.001	0.08 (0.08)	-0.07	0.24	0.04	0.12	1.05	.296
	angry:pb(block.z)	0.03 (0.06)	-0.09	0.15	-0.04	0.06	0.48	.629	-0.07 (0.06)	-0.20	0.05	-0.14	-0.01	-1.19	.235
	happy:pb(block.z)	0.02 (0.06)	-0.10	0.14	-0.03	0.05	0.34	.732	-0.08 (0.06)	-0.20	0.05	-0.12	-0.04	-1.20	.229
	mismatch:pb(block.z)	0.03 (0.06)	-0.08	0.14	-0.07	0.07	0.51	.609	0.07 (0.05)	-0.04	0.17	0.04	0.12	1.23	.219
	angry:mismatch:pb(block.z)	-0.02 (0.08)	-0.18	0.14	-0.07	0.10	-0.23	.818	0.13 (0.08)	-0.03	0.30	0.07	0.21	1.58	.115
	happy:mismatch:pb(block.z)	0.01 (0.08)	-0.15	0.18	-0.03	0.11	0.18	.858	0.05 (0.09)	-0.12	0.22	-0.02	0.10	0.57	.572
nu	(Intercept)	5.57 (0.02)	5.53	5.61	5.52	5.58	283.67	<.001	5.63 (0.02)	5.60	5.67	5.60	5.64	308.86	<.001
	angry	0.03 (0.03)	-0.03	0.08	-0.01	0.04	0.92	.359	0.04 (0.03)	-0.01	0.09	0.02	0.08	1.54	.123
	happy	-0.07 (0.03)	-0.13	-0.02	-0.10	-0.05	-2.55	.011	0.04 (0.03)	-0.01	0.09	0.03	0.07	1.67	.095
	mismatch	0.01 (0.03)	-0.05	0.07	-0.01	0.06	0.36	.721	0.07 (0.03)	0.02	0.12	0.06	0.10	2.87	.004
	pb(block.z)	-0.05 (0.02)	-0.08	-0.01	-0.07	-0.03	-2.35	.019	-0.04 (0.02)	-0.08	-0.01	-0.06	-0.03	-2.54	.011
	angry:mismatch	0.06 (0.04)	-0.02	0.14	0.03	0.08	1.52	.128	-0.04 (0.04)	-0.11	0.03	-0.08	-0.01	-1.07	.285
	happy:mismatch	0.03 (0.04)	-0.05	0.11	-0.03	0.05	0.79	.428	0.01 (0.04)	-0.06	0.08	-0.02	0.04	0.28	.777
	angry:pb(block.z)	0.03 (0.03)	-0.03	0.08	0.01	0.06	1.01	.313	-0.08 (0.02)	-0.13	-0.03	-0.09	-0.07	-3.38	<.001
	happy:pb(block.z)	0.05 (0.03)	-0.01	0.11	0.03	0.07	1.57	.116	0.01 (0.02)	-0.04	0.06	0.00	0.02	0.46	.645
	mismatch:pb(block.z)	0.01 (0.03)	-0.04	0.07	-0.02	0.03	0.41	.685	0.04 (0.02)	-0.01	0.09	0.02	0.05	1.53	.126
	angry:mismatch:pb(block.z)	0.00 (0.04)	-0.08	0.08	-0.03	0.02	-0.02	.981	0.08 (0.03)	0.01	0.14	0.05	0.10	2.22	.026
	happy:mismatch:pb(block.z)	-0.08 (0.04)	-0.16	0.00	-0.11	-0.04	-1.88	.060	0.00 (0.03)	-0.07	0.07	-0.01	0.02	0.06	.949
	aic	-2564.77							-2273.02						
	bic	-1608.48							-1283.20						
	rmse	1.01							1.00						
	nobs	13873.00							21563.00						

Notes: β = model estimate, SE = standard error of the estimate, CI = lower and upper asymptotic confidence intervals, Stab = estimate ranges leaving out one participant at a time; Note that SE and CI might not be reliable due to the involvement of smoothing terms. The estimates of σ and ν are on the log scale.

doi: 10.53846/goediss-9841

Table B3

Statistical results of the full models of the test sessions for Experiment 1 and Experiment 2.

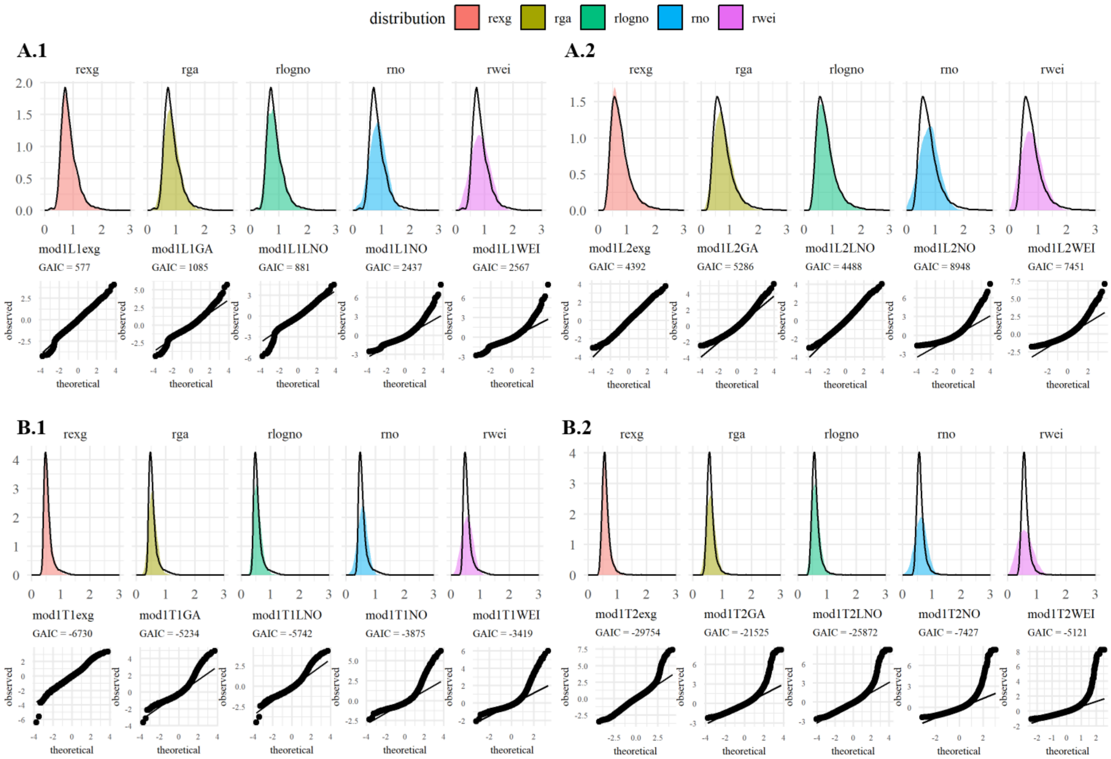
		Exp 1							Exp 2						
		$\beta\,(SE)$	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	t	p	$\beta\,(SE)$	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	t	p
mu	(Intercept)	450.18 (1.96)	446.33	454.03	445.16	472.09	181.04	<.001	497.74 (1.24)	495.31	500.17	488.51	700.61	225.11	<.001
	angry	16.87 (2.87)	11.23	22.50	15.79	18.37	5.87	<.001	-1.62 (1.75)	-5.06	1.82	-4.73	-0.87	-0.92	.356
	happy	1.95 (2.66)	-3.25	7.16	0.64	3.30	0.74	.462	-4.90 (1.75)	-8.33	-1.47	-7.57	-3.81	-2.80	.005
	mismatch	11.66 (3.03)	5.72	17.61	9.96	13.41	3.85	<.001	-4.41 (1.81)	-7.95	-0.87	-5.38	-3.30	-2.44	.015
	pb(block.z)	-4.19 (1.87)	-7.87	-0.52	-5.96	-2.62	-2.24	.025	1.49 (1.16)	-0.79	3.77	0.61	3.44	1.28	.200
	angry:mismatch	-34.88 (4.07)	-42.86	-26.90	-37.75	-32.54	-8.57	<.001	-3.51 (2.57)	-8.54	1.52	-4.75	2.73	-1.37	.171
	happy:mismatch	-22.90 (3.87)	-30.50	-15.31	-24.51	-20.53	-5.91	<.001	7.65 (2.55)	2.65	12.65	5.91	9.54	3.00	.003
	angry:pb(block.z)	1.56 (2.84)	-4.01	7.13	-0.15	4.31	0.55	.583	-5.87 (1.60)	-9.01	-2.72	-6.38	-3.48	-3.66	<.001
	happy:pb(block.z)	-2.68 (2.62)	-7.82	2.46	-4.15	-0.28	-1.02	.306	-3.29 (1.64)	-6.51	-0.08	-4.34	-0.03	-2.01	.044
	mismatch:pb(block.z)	5.07 (3.07)	-0.96	11.09	3.43	7.07	1.65	.100	-2.39 (1.68)	-5.68	0.90	-3.15	0.67	-1.42	.155
	angry:mismatch:pb(block.z)	-4.60 (4.19)	-12.82	3.62	-7.89	-2.47	-1.10	.273	4.96 (2.34)	0.37	9.55	3.48	5.81	2.12	.034
	happy:mismatch:pb(block.z)	-4.46 (3.92)	-12.13	3.22	-6.83	-2.39	-1.14	.255	4.23 (2.38)	-0.44	8.90	-0.42	5.40	1.77	.076
sigma	(Intercept)	3.65 (0.04)	3.56	3.73	3.59	3.69	84.16	<.001	3.63 (0.03)	3.56	3.69	3.59	3.75	109.71	<.001
	angry	0.17 (0.05)	0.07	0.27	0.12	0.21	3.39	<.001	-0.01 (0.05)	-0.10	0.09	-0.05	0.01	-0.13	.893
	happy	0.03 (0.06)	-0.08	0.15	0.02	0.07	0.58	.565	-0.02 (0.05)	-0.11	0.08	-0.05	0.00	-0.33	.745
	mismatch	0.25 (0.05)	0.14	0.36	0.18	0.27	4.54	<.001	-0.02 (0.05)	-0.12	0.07	-0.04	-0.01	-0.49	.626
	pb(block.z)	-0.13 (0.04)	-0.21	-0.06	-0.15	-0.09	-3.54	<.001	0.05 (0.03)	-0.01	0.10	0.03	0.06	1.65	.098
	angry:mismatch	-0.50 (0.07)	-0.64	-0.36	-0.52	-0.38	-6.93	<.001	0.00 (0.07)	-0.14	0.13	-0.03	0.08	-0.04	.968
	happy:mismatch	-0.44 (0.08)	-0.59	-0.28	-0.51	-0.37	-5.59	<.001	0.09 (0.07)	-0.05	0.23	0.06	0.11	1.27	.204
	angry:pb(block.z)	0.09 (0.05)	-0.01	0.18	0.04	0.16	1.75	.079	-0.11 (0.04)	-0.19	-0.03	-0.13	-0.06	-2.66	.008
	happy:pb(block.z)	0.02 (0.05)	-0.08	0.12	-0.01	0.11	0.42	.677	-0.06 (0.04)	-0.15	0.02	-0.08	0.00	-1.48	.139
	mismatch:pb(block.z)	0.13 (0.06)	0.02	0.24	0.03	0.20	2.38	.017	-0.04 (0.04)	-0.13	0.04	-0.06	0.01	-0.98	.327
	angry:mismatch:pb(block.z)	-0.22 (0.08)	-0.38	-0.05	-0.31	-0.12	-2.62	.009	0.06 (0.06)	-0.07	0.19	0.02	0.08	0.96	.339
	happy:mismatch:pb(block.z)	-0.20 (0.07)	-0.34	-0.06	-0.35	-0.09	-2.85	.004	0.10 (0.07)	-0.03	0.23	0.01	0.12	1.45	.146
nu	(Intercept)	4.80 (0.03)	4.75	4.85	4.76	4.82	190.96	<.001	4.76 (0.01)	4.73	4.78	4.57	4.77	408.87	<.001
	angry	-0.05 (0.03)	-0.11	0.02	-0.06	-0.02	-1.33	.182	0.00 (0.02)	-0.04	0.03	-0.01	0.07	-0.13	.896
	happy	-0.25 (0.03)	-0.32	-0.18	-0.28	-0.22	-7.06	<.001	0.01 (0.02)	-0.03	0.04	-0.01	0.07	0.28	.782
	mismatch	0.07 (0.04)	-0.01	0.14	0.04	0.09	1.80	.071	0.16 (0.02)	0.13	0.19	0.15	0.18	9.38	<.001
	pb(block.z)	0.02 (0.03)	-0.03	0.07	0.00	0.03	0.73	.466	-0.02 (0.01)	-0.04	0.00	-0.03	0.01	-1.54	.123
	angry:mismatch	-0.19 (0.05)	-0.28	-0.09	-0.23	-0.16	-3.86	<.001	0.03 (0.02)	-0.01	0.08	-0.11	0.06	1.43	.153
	happy:mismatch	-0.04 (0.05)	-0.14	0.05	-0.07	-0.01	-0.90	.371	-0.11 (0.03)	-0.16	-0.06	-0.14	-0.07	-4.14	<.001
	angry:pb(block.z)	-0.14 (0.03)	-0.21	-0.08	-0.17	-0.13	-4.16	<.001	0.06 (0.02)	0.03	0.10	0.02	0.08	3.73	<.001
	happy:pb(block.z)	-0.03 (0.04)	-0.10	0.04	-0.08	-0.01	-0.79	.432	0.06 (0.02)	0.03	0.09	0.00	0.08	3.61	<.001
	mismatch:pb(block.z)	-0.06 (0.04)	-0.13	0.01	-0.09	-0.04	-1.81	.071	0.01 (0.02)	-0.02	0.04	-0.03	0.03	0.85	.395
	angry:mismatch:pb(block.z)	0.12 (0.05)	0.03	0.22	0.10	0.15	2.47	.013	-0.03 (0.02)	-0.08	0.01	-0.04	-0.02	-1.38	.167
	happy:mismatch:pb(block.z)	0.06 (0.05)	-0.04	0.16	0.02	0.10	1.13	.258	-0.05 (0.02)	-0.10	-0.01	-0.07	0.03	-2.21	.027
	aic	-16290.64							-38245.57						
	bic	-15447.30							-37316.76						
	rmse	1.01							1.02						
	nobs	8979.00							23316.00						

Notes: β = model estimate, SE = standard error of the estimate, CI = lower and upper asymptotic confidence intervals, Stab = estimate ranges leaving out one participant at a time; Note that SE and CI might not be reliable due to the involvement of smoothing terms. The estimates of σ and ν are on the log scale.

doi: 10.53846/goediss-9841

Figure B1

Estimated densities of an intercept model vs. observed empirical densities.



Notes: Model-based distributional densities (coloured distributions) were compared with the approximated empirical density depicted by the black lines (upper panels). Lower panels show Q-Q plots of the theoretical vs. observed residuals for intercept models with different distributions. Among the selected distributions, the exGaussian showed the best fit for both experiments and sessions. The learning sessions are shown in panels **A.1** for Exp.1 and **A.2** for Exp.2. The test sessions are shown in panels **B.1** for Exp.1, and **B.2** for Exp.2.

doi: 10.53846/goediss-9841

Table B4
Comparison of AICs for different fitted distributions by experiment and session.

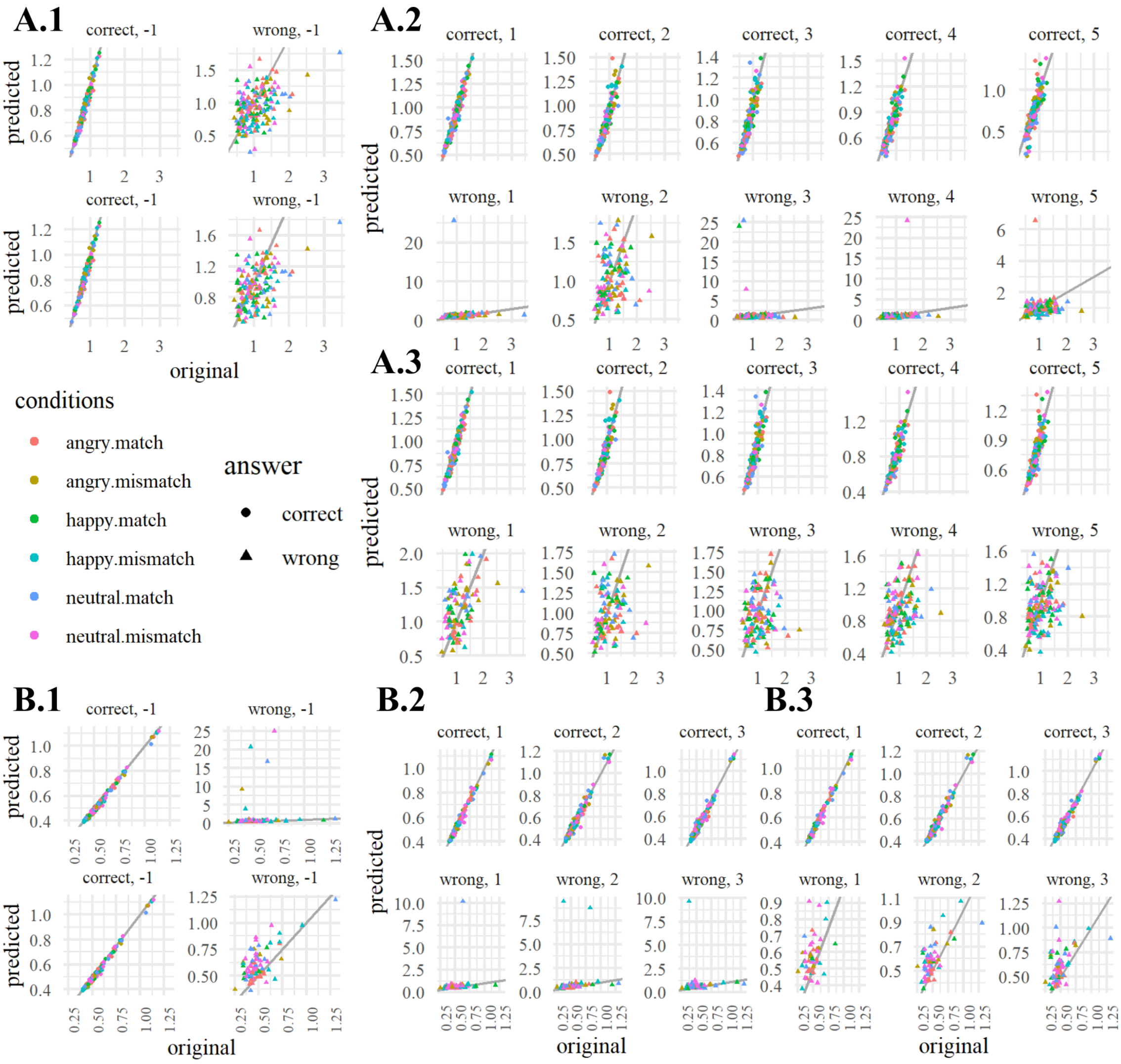
	L1	L2	T1	T2
exGAUS	**-337.63**	2374.66	-6878.12	-29755.08
GB2	-284.93	2022.34	**-6987.09**	-31145.96
BCTo	-274.31	1970.96	-6982.66	**-31437.01**
BCT	-271.08	1970.63	-6982.45	-31436.76
BCPE	28.32	2970.05	-6794.12	-31212.73
BCCGo	28.38	8071.09	-6725.51	-31045.49
GG	29.51	**1966.05**	-6650.2	-30600.29
BCCG	31.38	8072.85	-6725	-31045.58
BCPEo	33.39	3019.95	-6794.61	-31212.94
LNO	36.58	2285.91	-5852.56	-25882.1
LOGNO	36.58	2285.91	-5852.56	-25882.1
LOGNO2	36.58	2285.91	-5852.56	-25882.1
GIG	115.3	1969.45	-6232.54	-28353.37
IGAMMA	147.5	1974.47	-6232.38	-28355.48
IG	272.2	2554.86	-5819.72	-25051.25
GA	292.47	3182.45	-5333.69	-21566.21
WEI	1990.51	5777.58	-3479.35	-5437.35
WEI3	1990.51	5777.59	-3479.35	-5437.35
WEI2	1990.51	5777.59	-3479.35	-5437.28
EXP	12143.39	18110.86	4063.68	24202.74
PARETO2o	12211.14	18234.31	4107.97	24415.37
GP	12215.54	18231.67	4112.41	24431.7
PARETO2	12215.54	18231.67	4112.41	24431.7

Notes: Values show the AICs produced by the function choose.Dist() of the GAMLSS R package. Bold values represent the lowest AIC, i.e., the best model fit. All models estimated μ only, including the predictors emotion, congruence and the penalized spline function of block.z and their interactions. No random terms were included. Included were all GAMLSS continuous distributions in the positive real line, i.e., the exGaus: ex-Gaussian, GB2: generalized Beta type 2 and generalized Pareto, BCTo/BCT: Box-Cox t, BCPE/BCPEo: Box-Cox Power Exponential, BCCG/BCCGo: Box-Cox Cole and Green, GG: Generalized Gamma, LNO/LOGNO/LOGNO2: Log Normal, GIG: Generalized Inverse Gaussian, IGAMMA: Inverse Gamma, GA: Gamma, WEI/WEI2/WEI3: Weibull, EXP: Exponential, PARETO2/PARETO2o: Pareto, GP: Generalized Pareto distribution.

doi: 10.53846/goediss-9841

Figure B2

Empirical RT medians vs. theoretical RT medians of the drift diffusion model for Exp.1.

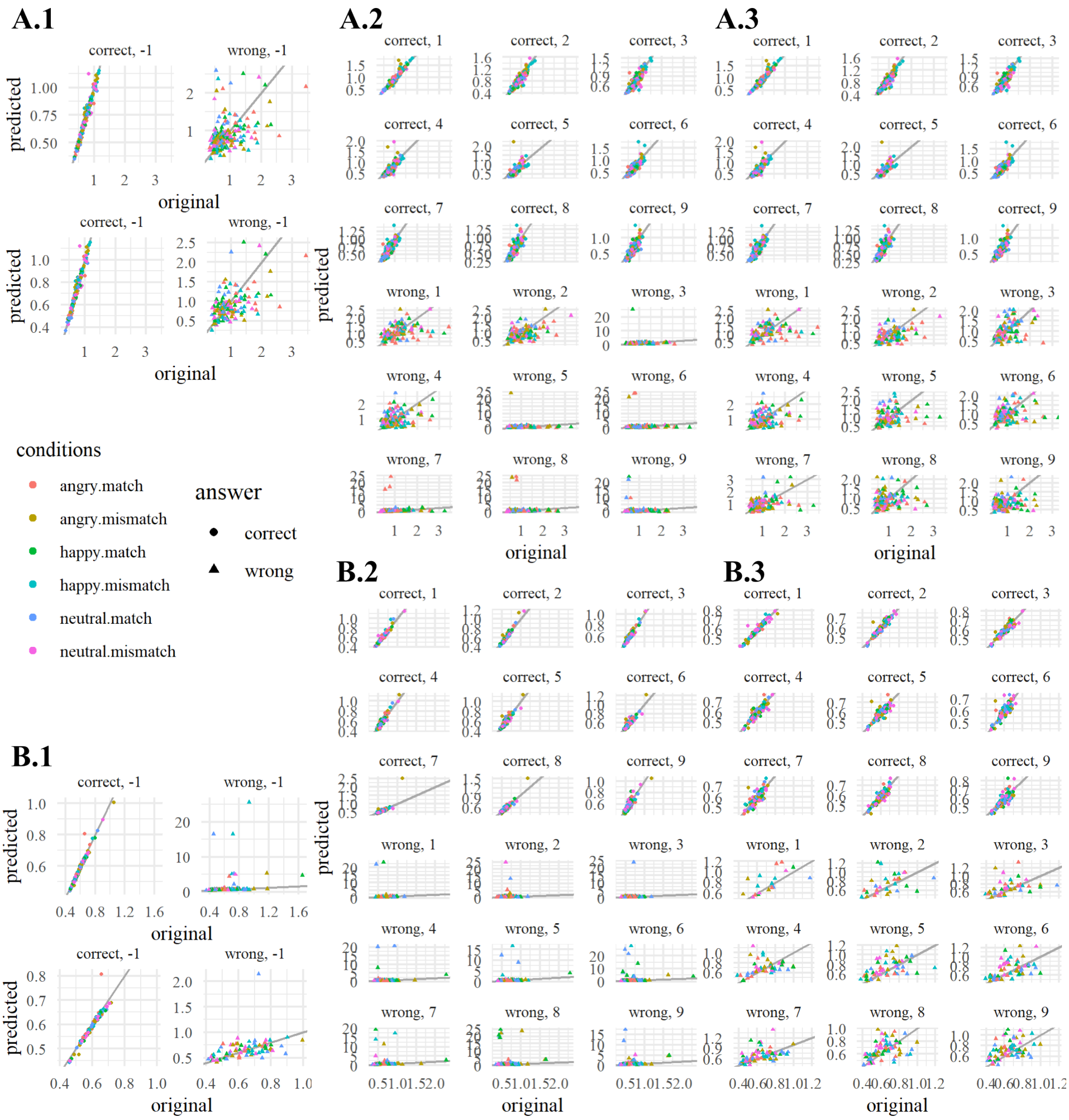


Notes: **A.1** shows the observed vs. predicted RT medians of the global model of the learning session of Exp.1 with the upper panel displaying the original set of participants. The lower panel displays the RT medians for when excluding participants which obtained predicted median RTs larger than 5 seconds. **A.2** shows the observed vs. predicted RT medians for every block window for the original participant set, whereas **A.3** shows the reduced participant set. Analogously, panels **B** display the global and block window-wise values of the test session of Exp 1. Particularly, for wrong decisions, the predicted RT medians deviated from the empirical data, with large prediction errors in the test sessions, probably due to the small number of wrong answers.

doi: 10.53846/goediss-9841

Figure B3

Empirical RT medians vs. theoretical RT medians for each block-bin drift diffusion model for the learning session of Exp.2.

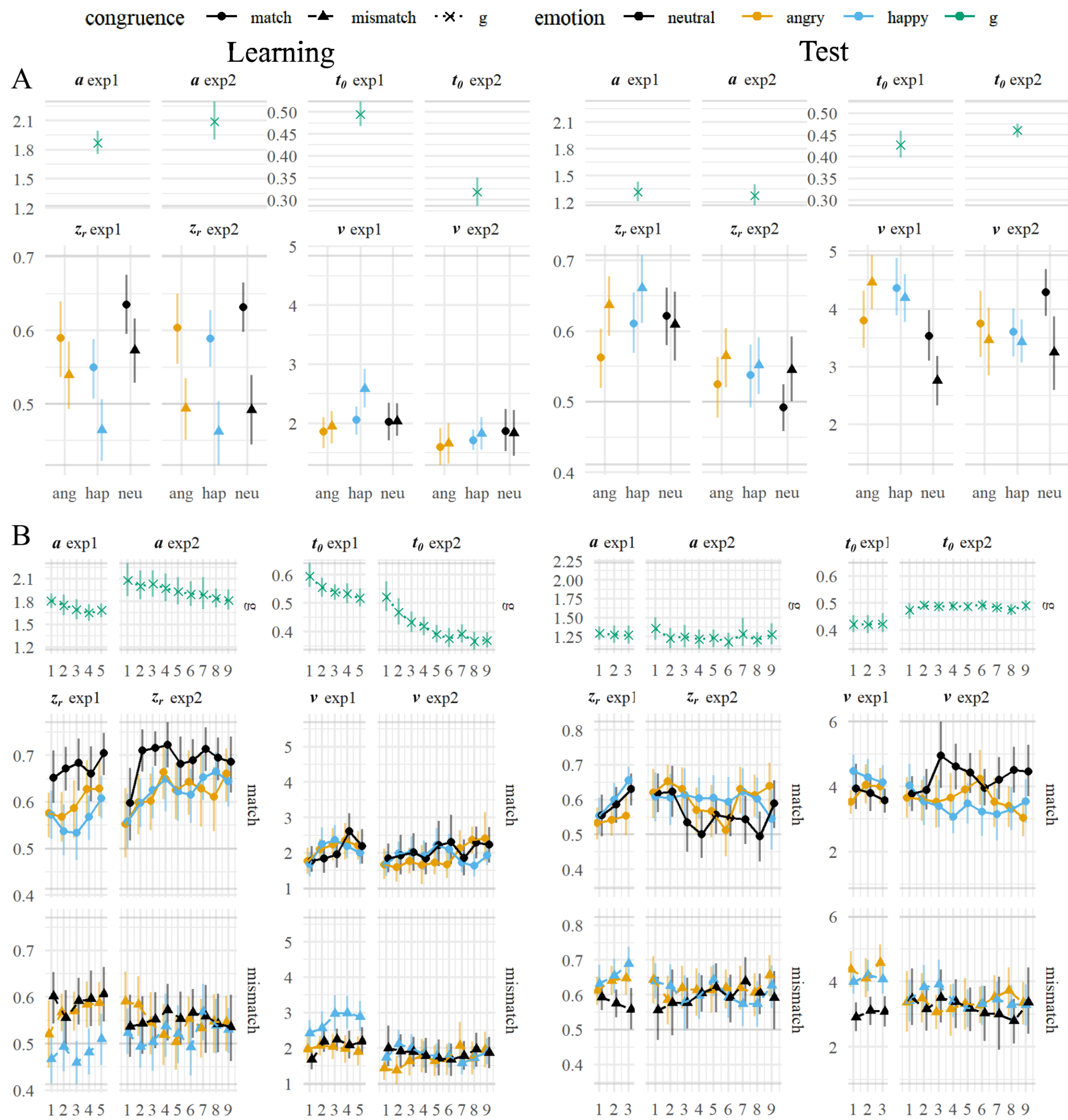


Notes: **A.1** shows the observed vs. predicted RT medians of the global model of the learning session of Exp.2 with the upper panel displaying the original set of participants. The lower panel displays the RT medians for when excluding participants which obtained predicted median RTs larger than 5 seconds. **A.2** shows the observed vs. predicted RT medians for every block window for the original participant set, whereas **A.3** shows the reduced participant set. Analogously, panels **B** display the global and block window-wise values of the test session of Exp 2. Particularly, for wrong decisions, the predicted RT medians deviated from the empirical data, with large prediction errors in the test sessions, probably due to the small number of wrong answers.

doi: 10.53846/goediss-9841

Figure B4

Averaged drift diffusion parameters for the learning and test session based on a subset of participants.



Notes: Drift diffusion parameters were estimated for a reduced participant sample by excluding participants for which predicted median RTs exceeded a 5 seconds threshold. The left panel shows the parameters for the learning sessions, the right for the test sessions. Error bars indicate 95% non-parametric bootstrapped confidence intervals around the sample mean. The boundary separation a and the non-decision time t_0 were estimated independent of the stimulus conditions. The starting point z_r and drift rate v were allowed to vary between conditions, i.e., we obtained separate estimates for each emotion and congruence level. **A** shows the global model across the sessions. **B** shows parameters changes as a function of stimulus repetition. Here, the x-axis refers to the sliding block windows of the experiment, of which one corresponds to blocks 1-10, two to 6-15, etc.

doi: 10.53846/goediss-9841

C Appendix of Study 3

Learning

Table C1

Statistical results for the valence ratings of the affect bursts

	β	SE	t	CI_l	CI_u	LRT:Model	χ^2	df	p
(Intercept)	-0.37	0.08	-4.79	-0.52	-0.22	-	-	-	-
emotionyawning	0.29	0.11	2.68	0.08	0.50				
emotionanger	-1.74	0.15	-11.90	-2.03	-1.45				
emotiondisgust	-1.10	0.13	-8.75	-1.34	-0.85	emotion	6	0.7	.008
emotionamusement	2.17	0.14	15.02	1.89	2.46				
emotionelation	2.22	0.15	14.59	1.93	2.52				
typereaction	0.05	0.11	0.47	-0.16	0.26	type	828.21	23.1	<.001
emotionyawning:typereaction	-0.01	0.15	-0.08	-0.31	0.29				
emotionanger:typereaction	0.67	0.19	3.52	0.30	1.04				
emotiondisgust:typereaction	0.20	0.17	1.17	-0.14	0.54	emotion:type	137.78	6.34	<.001
emotionamusement:typereaction	-0.84	0.19	-4.52	-1.20	-0.48				
emotionelation:typereaction	-1.48	0.19	-7.91	-1.85	-1.12				

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, LRT = Likelihood ratio test

Table C2

Statistical results for accuracy in learning checks during learning

	β	SE	z	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p
(Intercept)	-0.07	0.17	-0.44	-0.41	0.25	-0.14	0.00	(Intercept)	-	-	-
valence_positive-g.m	0.06	0.14	0.40	-0.25	0.39	0.01	0.13	valence	1.9	2	.387
valence_negative-g.m	-0.18	0.13	-1.36	-0.47	0.12	-0.23	-0.14				
checknr	0.45	0.05	8.49	0.36	0.56	0.42	0.47	checknr	51.09	1	<.001
valence_positive-g.m:checknr	0.05	0.02	2.77	-0.02	0.12	0.03	0.07	valence:checknr	7.95	2	.019
valence_negative-g.m:checknr	-0.02	0.02	-1.15	-0.08	0.05	-0.04	0.00				

Notes: beta = model estimate, SE = standard error of the estimate, CI = lower and upper 95 Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test

doi: 10.53846/goediss-9841

Test

Table C3

Statistical results for log response times by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.20	0.02	-9.30	-0.25	-0.16	-0.21	-0.20	(Intercept)	-	-	-	-
valence_positive-g.m	-0.01	0.01	-1.37	-0.02	0.00	0.01	0.01	valence	2.14	2	.343	0.05
valence_negative-g.m	0.00	0.01	-0.08	-0.02	0.01	0.00	0.00					
task_valclass-g.m	0.13	0.01	10.99	0.11	0.15	-0.13	-0.12	task	56.42	1	<.001	3.10
valence_positive-g.m:task_valclass-g.m	0.00	0.00	-0.31	-0.01	0.01	-0.01	-0.01	valence:task	6.55	2	.038	0.09
valence_negative-g.m:task_valclass-g.m	-0.01	0.00	-2.06	-0.02	0.00	0.01	0.01					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C4

Statistical results for log response times by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.20	0.02	-9.37	-0.24	-0.16	-0.21	-0.20	(Intercept)	-	-	-	-
emotion_elation-g.m	0.00	0.01	-0.33	-0.03	0.02	-0.01	0.00					
emotion_amusement-g.m	-0.01	0.01	-1.06	-0.03	0.01	-0.02	-0.01					
emotion_disgust-g.m	0.00	0.01	-0.13	-0.02	0.02	-0.01	0.00	emotion	2.68	5	.749	0.01
emotion_anger-g.m	0.00	0.01	0.07	-0.02	0.02	0.00	0.00					
emotion_yawning-g.m	0.02	0.01	1.37	-0.01	0.04	0.01	0.02					
task_valclass-g.m	0.13	0.01	25.51	0.12	0.14	0.12	0.13	task	406.07	1	<.001	1.52
emotion_elation-g.m:task_valclass-g.m	0.00	0.01	0.24	-0.02	0.02	0.00	0.01					
emotion_amusement-g.m:task_valclass-g.m	-0.01	0.01	-0.51	-0.03	0.02	-0.01	0.00					
emotion_disgust-g.m:task_valclass-g.m	0.00	0.01	-0.01	-0.02	0.02	0.00	0.00	emotion:task	4.64	5	.461	0.01
emotion_anger-g.m:task_valclass-g.m	-0.02	0.01	-1.60	-0.04	0.00	-0.02	-0.01					
emotion_yawning-g.m:task_valclass-g.m	0.02	0.01	1.58	0.00	0.04	0.02	0.02					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C5

Statistical results for log response times in the old-new task (reference = novel)

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.32	0.02	-13.04	-0.37	-0.27	-0.34	-0.32	(Intercept)	-	-	-	-
valence_positive-g.m	-0.02	0.01	-2.35	-0.03	0.00	-0.02	-0.02					
valence_negative-g.m	0.00	0.01	-0.32	-0.02	0.01	0.00	0.00	valence	19.67	3	<.001	0.18
valence_neutral-g.m	-0.01	0.01	-1.71	-0.03	0.00	-0.02	-0.01					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table C6
Statistical results for accuracy by valence

	β	SE	z	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p
(Intercept)	4.73	0.19	25.33	4.38	5.12	4.66	4.78	(Intercept)	-	-	-
valence_positive-g.m	0.05	0.13	0.36	-0.22	0.32	0.00	0.12	valence	0.13	2	.938
valence_negative-g.m	-0.01	0.16	-0.03	-0.32	0.33	-0.06	0.04				
task_valclass-g.m	-0.57	0.15	-3.85	-0.86	-0.30	-0.61	-0.51	task	13.85	1	<.001
valence_positive-g.m:task_valclass-g.m	-0.01	0.09	-0.16	-0.19	0.16	-0.15	0.07	valence:task	0.56	2	.756
valence_negative-g.m:task_valclass-g.m	0.07	0.09	0.73	-0.10	0.26	-0.01	0.16				

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C7
Statistical results for accuracy by emotion

	β	SE	z	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p
(Intercept)	5.00	0.19	25.72	4.66	5.44	4.93	5.07	(Intercept)	-	-	-
emotion_yawning-g.m	-0.26	0.24	-1.05	-0.73	0.26	-0.38	-0.03	emotion	2.74	5	.740
emotion_anger-g.m	0.00	0.23	-0.01	-0.47	0.48	-0.08	0.08				
emotion_disgust-g.m	-0.05	0.24	-0.21	-0.52	0.53	-0.18	0.12				
emotion_amusement-g.m	0.21	0.20	1.05	-0.21	0.72	0.11	0.30				
emotion_elation-g.m	-0.18	0.20	-0.88	-0.62	0.27	-0.29	-0.01				
task_valenceclass-g.m	-0.57	0.14	-4.03	-0.89	-0.30	-0.63	-0.50	task	14.16	1	<.001
emotion_yawning-g.m:task_valenceclass-g.m	-0.15	0.13	-1.18	-0.41	0.09	-0.40	-0.03	emotion:task	12.72	5	.026
emotion_anger-g.m:task_valenceclass-g.m	0.33	0.13	2.60	0.08	0.60	0.17	0.42				
emotion_disgust-g.m:task_valenceclass-g.m	-0.11	0.13	-0.84	-0.40	0.15	-0.30	0.18				
emotion_amusement-g.m:task_valenceclass-g.m	-0.29	0.16	-1.89	-0.69	0.01	-0.43	-0.19				
emotion_elation-g.m:task_valenceclass-g.m	0.21	0.12	1.73	-0.02	0.46	-0.03	0.35				

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C8
Statistical results for the P1 mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.04	0.65	6.22	2.76	5.21	3.78	4.30	(Intercept)	-	-	-	-
valence_positive-g.m	-0.02	0.06	-0.23	-0.14	0.12	-0.04	0.01	valence	0.14	2	.931	0.00
valence_negative-g.m	-0.01	0.06	-0.14	-0.15	0.12	-0.03	0.02					
task_valenceclass-g.m	-0.03	0.05	-0.67	-0.12	0.06	-0.05	-0.01	task	0.46	1	.497	0.00
valence_positive-g.m:task_valenceclass-g.m	-0.03	0.06	-0.45	-0.16	0.10	-0.04	0.00	valence:task	0.54	2	.764	0.00
valence_negative-g.m:task_valenceclass-g.m	-0.02	0.06	-0.26	-0.14	0.11	-0.03	0.00					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table C9

Statistical results for the P1 mean amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.03	0.65	6.20	2.78	5.31	3.77	4.29	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.06	0.10	0.61	-0.13	0.26	0.01	0.10					
emotion_elation-g.m	0.06	0.10	0.58	-0.14	0.28	0.02	0.11					
emotion_disgust-g.m	-0.11	0.10	-1.04	-0.30	0.09	-0.16	-0.07	emotion	2.07	5	.839	0.00
emotion_anger-g.m	0.06	0.10	0.59	-0.14	0.25	0.03	0.09					
emotion_amusement-g.m	-0.05	0.10	-0.45	-0.25	0.15	-0.08	0.00					
task_valenceclass-g.m	-0.02	0.05	-0.51	-0.11	0.07	-0.05	-0.01	task	0.26	1	.607	0.00
emotion_yawning-g.m:task_valenceclass-g.m	0.07	0.10	0.70	-0.15	0.28	0.02	0.11					
emotion_elation-g.m:task_valenceclass-g.m	-0.06	0.10	-0.59	-0.26	0.14	-0.09	0.01					
emotion_disgust-g.m:task_valenceclass-g.m	0.01	0.10	0.13	-0.18	0.23	-0.03	0.04	emotion:task	0.96	5	.965	0.00
emotion_anger-g.m:task_valenceclass-g.m	-0.05	0.10	-0.48	-0.25	0.14	-0.08	-0.01					
emotion_amusement-g.m:task_valenceclass-g.m	0.02	0.10	0.22	-0.16	0.24	-0.01	0.05					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C10

Statistical results for the P1 peak amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	6.91	0.63	11.02	5.68	8.18	6.59	7.12	(Intercept)	-	-	-	-
valence_positive-g.m	0.02	0.08	0.23	-0.13	0.17	-0.03	0.04	valence	0.22	2	.896	0.00
valence_negative-g.m	-0.03	0.08	-0.46	-0.18	0.12	-0.07	-0.01					
task_valenceclass-g.m	-0.04	0.05	-0.81	-0.15	0.06	-0.07	-0.02	task	0.67	1	.413	0.00
valence_positive-g.m:task_valenceclass-g.m	0.04	0.08	0.47	-0.11	0.18	0.01	0.06	valence:task	0.91	2	.635	0.00
valence_negative-g.m:task_valenceclass-g.m	0.04	0.08	0.47	-0.11	0.18	0.01	0.05					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C11

Statistical results for the P1 peak amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	7.21	0.63	11.52	6.08	8.50	6.91	7.44	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.21	0.12	1.81	-0.01	0.44	0.17	0.25					
emotion_elation-g.m	0.09	0.12	0.75	-0.14	0.32	0.04	0.13					
emotion_disgust-g.m	-0.14	0.12	-1.20	-0.35	0.09	-0.21	-0.09	emotion	5.33	5	.377	0.01
emotion_anger-g.m	0.00	0.12	0.04	-0.23	0.22	-0.02	0.03					
emotion_amusement-g.m	-0.08	0.12	-0.73	-0.30	0.13	-0.13	-0.04					
task_valenceclass-g.m	-0.05	0.05	-0.98	-0.15	0.04	-0.07	-0.04	task	0.99	1	.319	0.00
emotion_yawning-g.m:task_valenceclass-g.m	-0.04	0.12	-0.32	-0.27	0.18	-0.10	0.00					
emotion_elation-g.m:task_valenceclass-g.m	-0.05	0.12	-0.43	-0.26	0.18	-0.08	0.03					
emotion_disgust-g.m:task_valenceclass-g.m	0.02	0.12	0.21	-0.21	0.25	-0.03	0.05	emotion:task	1.41	5	.923	0.00
emotion_anger-g.m:task_valenceclass-g.m	-0.05	0.12	-0.43	-0.26	0.19	-0.08	-0.01					
emotion_amusement-g.m:task_valenceclass-g.m	0.12	0.12	1.06	-0.11	0.34	0.08	0.16					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table C12

Statistical results for the N170 mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-6.93	0.59	-11.68	-8.10	-5.74	-7.12	-6.73	(Intercept)	-	-	-	-
valence_positive-g.m	0.04	0.06	0.74	-0.08	0.15	0.02	0.06	valence	2.1	2	.350	0.01
valence_negative-g.m	-0.09	0.06	-1.43	-0.20	0.03	-0.10	-0.06					
task_valenceclass-g.m	-0.23	0.04	-5.49	-0.31	-0.15	-0.25	-0.20	task	28.72	1	<.001	0.15
valence_positive-g.m:task_valenceclass-g.m	-0.04	0.06	-0.66	-0.15	0.07	-0.06	-0.03	valence:task	1.69	2	.430	0.01
valence_negative-g.m:task_valenceclass-g.m	-0.04	0.06	-0.62	-0.15	0.08	-0.05	-0.02					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C13

Statistical results for the N170 mean amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-6.94	0.59	-11.70	-8.09	-5.76	-7.13	-6.74	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.17	0.09	1.85	0.00	0.37	0.14	0.20					
emotion_elation-g.m	0.13	0.09	1.45	-0.04	0.30	0.10	0.16					
emotion_disgust-g.m	-0.28	0.09	-3.04	-0.44	-0.09	-0.30	-0.24	emotion	13.67	5	.018	0.03
emotion_anger-g.m	0.08	0.09	0.88	-0.10	0.26	0.05	0.11					
emotion_amusement-g.m	-0.03	0.09	-0.37	-0.22	0.16	-0.07	0.00					
task_valenceclass-g.m	-0.23	0.04	-5.78	-0.31	-0.16	-0.26	-0.20	task	32.96	1	<.001	0.08
emotion_yawning-g.m:task_valenceclass-g.m	0.02	0.09	0.20	-0.16	0.19	-0.01	0.05					
emotion_elation-g.m:task_valenceclass-g.m	-0.08	0.09	-0.85	-0.24	0.09	-0.11	-0.05					
emotion_disgust-g.m:task_valenceclass-g.m	0.07	0.09	0.74	-0.11	0.24	0.03	0.11	emotion:task	3.58	5	.612	0.01
emotion_anger-g.m:task_valenceclass-g.m	-0.12	0.09	-1.29	-0.30	0.06	-0.14	-0.08					
emotion_amusement-g.m:task_valenceclass-g.m	0.01	0.09	0.09	-0.17	0.18	-0.02	0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C14

Statistical results for the N170 peak amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-10.79	0.71	-15.26	-12.11	-9.43	-11.04	-10.54	(Intercept)	-	-	-	-
valence_positive-g.m	0.01	0.07	0.17	-0.12	0.15	-0.02	0.03	valence	1.65	2	.437	0.01
valence_negative-g.m	-0.08	0.07	-1.18	-0.21	0.05	-0.11	-0.05					
task_valenceclass-g.m	-0.02	0.05	-0.45	-0.11	0.07	-0.05	0.00	task	0.21	1	.648	0.00
valence_positive-g.m:task_valenceclass-g.m	-0.06	0.07	-0.86	-0.19	0.08	-0.08	-0.05	valence:task	0.77	2	.681	0.00
valence_negative-g.m:task_valenceclass-g.m	0.03	0.07	0.46	-0.11	0.16	0.01	0.05					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table C15

Statistical results for the N170 peak amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-11.11	0.71	-15.76	-12.49	-9.69	-11.35	-10.86	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.17	0.10	1.68	-0.03	0.37	0.14	0.21					
emotion_elation-g.m	0.05	0.10	0.52	-0.14	0.27	0.02	0.09					
emotion_disgust-g.m	-0.31	0.10	-3.01	-0.51	-0.12	-0.35	-0.27	emotion	11.58	5	.041	0.03
emotion_anger-g.m	0.11	0.10	1.04	-0.08	0.32	0.08	0.16					
emotion_amusement-g.m	0.04	0.10	0.36	-0.18	0.24	0.00	0.07					
task_valenceclass-g.m	-0.04	0.05	-0.86	-0.12	0.05	-0.06	-0.02	task	0.76	1	.383	0.00
emotion_yawning-g.m:task_valenceclass-g.m	0.07	0.10	0.72	-0.11	0.28	0.03	0.11					
emotion_elation-g.m:task_valenceclass-g.m	-0.06	0.10	-0.57	-0.26	0.13	-0.10	-0.03					
emotion_disgust-g.m:task_valenceclass-g.m	0.08	0.10	0.76	-0.14	0.28	0.04	0.12	emotion:task	1.66	5	.894	0.00
emotion_anger-g.m:task_valenceclass-g.m	-0.04	0.10	-0.37	-0.24	0.17	-0.08	-0.01					
emotion_amusement-g.m:task_valenceclass-g.m	-0.06	0.10	-0.62	-0.26	0.13	-0.10	-0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C16

Statistical results for the EPN mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-2.26	0.49	-4.62	-3.23	-1.33	-2.44	-2.11	(Intercept)	-	-	-	-
valence_positive-g.m	0.09	0.09	1.01	-0.08	0.24	0.06	0.11	valence	10.86	2	.004	0.06
valence_negative-g.m	-0.28	0.09	-3.23	-0.44	-0.12	-0.32	-0.24					
task_valenceclass-g.m	-0.01	0.06	-0.09	-0.12	0.11	-0.04	0.03	task	0.01	1	.925	0.00
valence_positive-g.m:task_valenceclass-g.m	0.02	0.09	0.21	-0.14	0.18	0.00	0.04	valence:task	0.13	2	.936	0.00
valence_negative-g.m:task_valenceclass-g.m	-0.03	0.09	-0.36	-0.19	0.14	-0.05	0.00					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C17

Statistical results for the EPN mean amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-2.27	0.49	-4.62	-3.28	-1.37	-2.45	-2.12	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.30	0.12	2.51	0.07	0.55	0.25	0.36					
emotion_elation-g.m	0.20	0.12	1.64	-0.04	0.43	0.17	0.24					
emotion_disgust-g.m	-0.46	0.12	-3.79	-0.69	-0.23	-0.50	-0.40	emotion	21.61	5	<.001	0.05
emotion_anger-g.m	-0.15	0.12	-1.25	-0.39	0.10	-0.19	-0.11					
emotion_amusement-g.m	-0.01	0.12	-0.12	-0.24	0.22	-0.05	0.02					
task_valenceclass-g.m	-0.01	0.05	-0.20	-0.12	0.09	-0.04	0.03	task	0.04	1	.838	0.00
emotion_yawning-g.m:task_valenceclass-g.m	-0.05	0.12	-0.42	-0.28	0.19	-0.09	-0.01					
emotion_elation-g.m:task_valenceclass-g.m	-0.04	0.12	-0.31	-0.28	0.21	-0.08	0.00					
emotion_disgust-g.m:task_valenceclass-g.m	0.01	0.12	0.05	-0.24	0.23	-0.03	0.03	emotion:task	1.47	5	.917	0.00
emotion_anger-g.m:task_valenceclass-g.m	-0.08	0.12	-0.68	-0.32	0.15	-0.12	-0.04					
emotion_amusement-g.m:task_valenceclass-g.m	0.10	0.12	0.84	-0.12	0.34	0.06	0.14					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table C18

Statistical results for the LPC mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	5.34	0.52	10.30	4.33	6.29	5.00	5.51	(Intercept)	-	-	-	-
valence_positive-g.m	0.09	0.09	0.98	-0.10	0.27	0.05	0.12	valence	2.64	2	.268	0.01
valence_negative-g.m	0.06	0.09	0.61	-0.11	0.24	0.04	0.08					
task_valenceclass-g.m	0.07	0.07	1.03	-0.06	0.19	0.04	0.12	task	1.08	1	.298	0.01
valence_positive-g.m:task_valenceclass-g.m	0.11	0.09	1.16	-0.08	0.30	0.08	0.14	valence:task	4.46	2	.108	0.02
valence_negative-g.m:task_valenceclass-g.m	0.09	0.09	0.93	-0.10	0.27	0.06	0.11					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C19

Statistical results for the LPC mean amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	5.35	0.52	10.31	4.40	6.35	5.02	5.53	(Intercept)	-	-	-	-
emotion_yawning-g.m	-0.09	0.12	-0.72	-0.32	0.17	-0.14	-0.06					
emotion_elation-g.m	0.06	0.12	0.52	-0.18	0.30	0.00	0.11					
emotion_disgust-g.m	-0.05	0.12	-0.37	-0.28	0.21	-0.09	0.01	emotion	5.01	5	.414	0.01
emotion_anger-g.m	0.16	0.12	1.33	-0.07	0.42	0.12	0.20					
emotion_amusement-g.m	0.10	0.12	0.83	-0.13	0.35	0.06	0.14					
task_valenceclass-g.m	0.09	0.05	1.55	-0.01	0.19	0.05	0.13	task	2.46	1	.117	0.01
emotion_yawning-g.m:task_valenceclass-g.m	-0.18	0.12	-1.49	-0.41	0.07	-0.21	-0.16					
emotion_elation-g.m:task_valenceclass-g.m	0.05	0.12	0.39	-0.19	0.31	0.01	0.07					
emotion_disgust-g.m:task_valenceclass-g.m	0.12	0.12	1.00	-0.12	0.35	0.09	0.16	emotion:task	6.2	5	.287	0.01
emotion_anger-g.m:task_valenceclass-g.m	0.03	0.12	0.27	-0.21	0.26	0.00	0.07					
emotion_amusement-g.m:task_valenceclass-g.m	0.16	0.12	1.29	-0.08	0.40	0.13	0.19					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table C20
Statistical results of the ordinal mixed model for the likability rating

	OR_val	CI_val	OR_emo	CI_emo
neutral-g.m	1.64	[1.31;2.16]		
negative-g.m	0.55	[0.4;0.69]		
positive-g.m	3.62	[2.87;5.16]		
clearthroat-g.m			1.47	[0.95;2.32]
yawning-g.m			1.32	[0.94;1.91]
anger-g.m			0.49	[0.31;0.71]
disgust-g.m			0.43	[0.26;0.63]
amusement-g.m			2.78	[1.89;4.47]
elation-g.m			3.37	[2.36;5.44]
1|2	0.02	[0.01;0.02]	0.01	[0.01;0.02]
2|3	0.10	[0.07;0.11]	0.08	[0.06;0.09]
3|4	0.39	[0.3;0.45]	0.33	[0.25;0.38]
4|5	1.30	[1.1;1.56]	1.10	[0.92;1.33]
5|6	6.52	[5.87;9.18]	5.51	[4.9;7.64]
6|7	36.57	[31.43;67.45]	30.99	[25.93;55.15]

Notes: Model estimates and threshold coefficients in Odds Ratios (OR) of the ordinal models (left: valence model, right: emotion model). Brackets indicate the 95% asymptotic confidence intervals of the OR.

doi: 10.53846/goediss-9841

Exploratory analyses

We analyzed a mid-frontal **FN400** (300 - 500 ms, at Fc3, F3, Fc4, F4), which has been related to familiarity of faces (Curran & Hancock, 2007), and a later parietal old/new component **LPON** (500 - 800 ms at CP1, CP2, P3, and P4), which has been related to episodic memory and recollection (Proverbio et al., 2019). Notably, the later old/new component overlaps in time and topography with the LPC component.

FN400. *Associated valence:* FN400 mean amplitudes were only analyzed for the old-new task to compare associated faces with novel faces. Two participants (IDs: 30, 7) were excluded from this analysis due to influential observations and Cook's distance >1 in the model with valence as a predictor. To compare results, we also excluded them in the model with separate emotion categories. There was no an effect of valence ($\chi^2(3) = 2.73$, $p = .436$).

Associated emotion: There was a trend for emotion ($\chi^2(6) = 10.85$, $p = .093$) but none of the post-hoc tests was significant. The largest difference between categories was between yawning and disgust ($\text{diff}_{\text{yaw-dis}} = -0.41$, $p = .137$).

Table C21
Statistical results for the FN400 mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.86	0.28	-3.06	-1.42	-0.31	-0.96	-0.72	(Intercept)	-	-	-	-
valence_novel-g.m	0.00	0.07	0.00	-0.14	0.14	-0.04	0.02					
valence_positive-g.m	0.01	0.07	0.11	-0.13	0.14	-0.03	0.03	valence	2.73	3	.436	0.02
valence_negative-g.m	0.09	0.07	1.28	-0.04	0.22	0.07	0.12					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C22
Statistical results for the FN400 mean amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.87	0.28	-3.14	-1.41	-0.39	-0.97	-0.73	(Intercept)	-	-	-	-
emotion_yawning-g.m	-0.20	0.10	-2.01	-0.40	-0.01	-0.22	-0.17					
emotion_novel-g.m	0.02	0.10	0.16	-0.18	0.21	-0.03	0.04					
emotion_elation-g.m	-0.13	0.10	-1.34	-0.32	0.05	-0.16	-0.09	emotion	10.85	6	.093	0.05
emotion_disgust-g.m	0.21	0.10	2.15	0.03	0.41	0.18	0.26					
emotion_anger-g.m	-0.03	0.10	-0.31	-0.23	0.15	-0.07	0.01					
emotion_amusement-g.m	0.14	0.10	1.42	-0.06	0.34	0.09	0.18					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

LPON. *Associated valence:* We analyzed the LPON component only for the old-new task to compare associated faces with novel faces. There was a main effect of valence ($\chi^2(3) = 84.87$, p <.001) due to the difference between novel faces and all associated faces(diff$_{\text{nov-pos}}$ = 1.36, p <.001; diff$_{\text{nov-neg}}$ = 1.30, p <.001; diff$_{\text{nov-neu}}$ = 1.17, p <.001). No other differences were significant.

Associated emotion: Analogously to the valence model, there was a main effect of emotion ($\chi^2(6) = 88.54$, p <.001). This effect was driven by a difference between novel faces and other emotion categories (all p- values < .01). None of the other post-hoc contrasts were significant.

Table C23

Statistical results for the LPON mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.46	0.28	15.76	3.92	4.98	4.34	4.55	(Intercept)	-	-	-	-
valence_novel-g.m	0.96	0.09	10.99	0.78	1.13	0.91	1.00					
valence_positive-g.m	-0.40	0.09	-4.59	-0.56	-0.22	-0.43	-0.37	valence	84.87	3	<.001	1.11
valence_negative-g.m	-0.34	0.09	-3.96	-0.52	-0.18	-0.39	-0.32					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C24

Statistical results for the LPON mean amplitudes by emotion

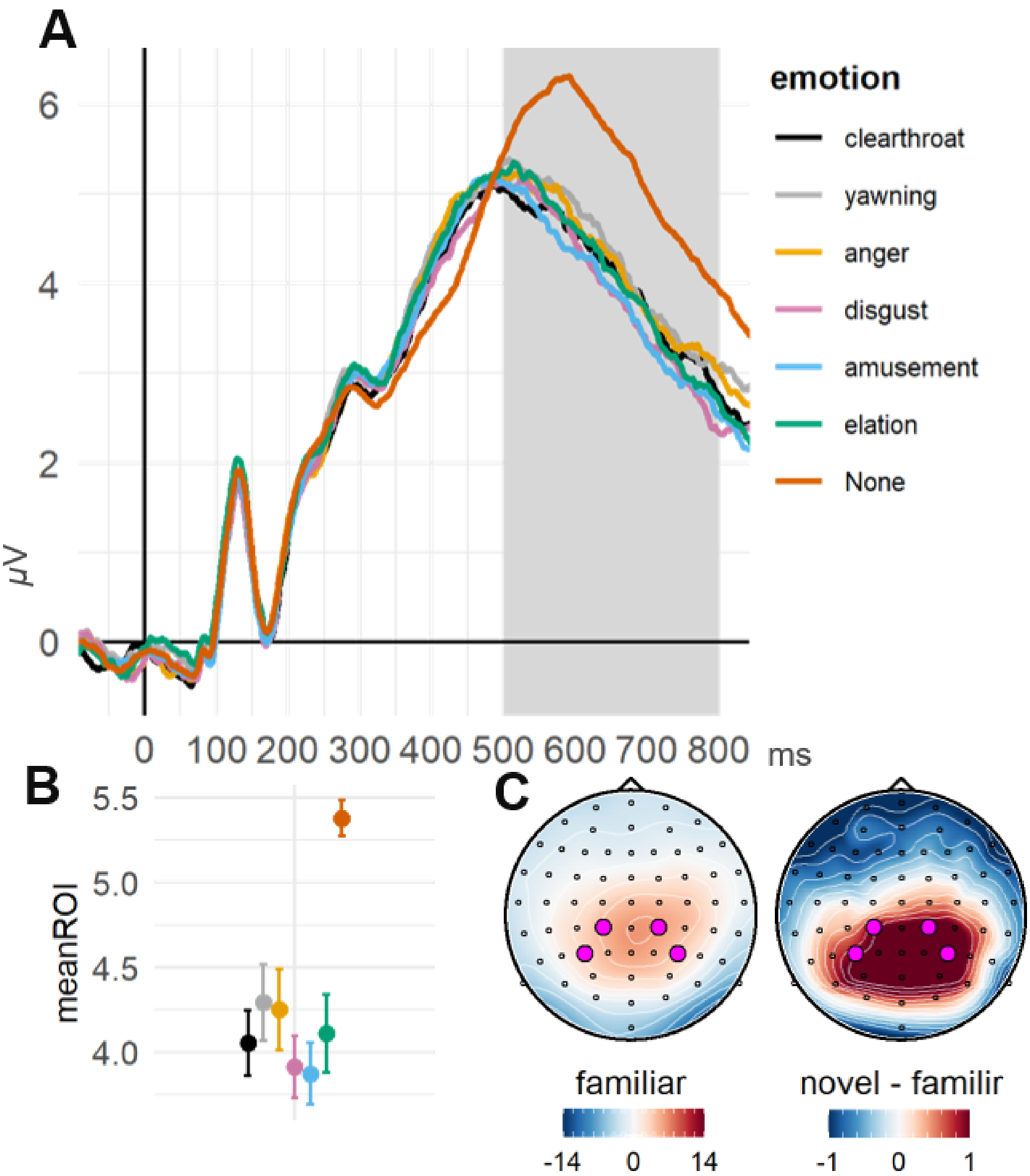
	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.33	0.28	15.44	3.73	4.86	4.21	4.42	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.05	0.11	0.40	-0.17	0.27	0.01	0.10					
emotion_novel-g.m	1.12	0.11	9.80	0.91	1.34	1.07	1.16					
emotion_elation-g.m	-0.14	0.11	-1.20	-0.35	0.08	-0.19	-0.10	emotion	88.54	6	<.001	0.47
emotion_disgust-g.m	-0.35	0.11	-3.05	-0.57	-0.13	-0.38	-0.31					
emotion_anger-g.m	-0.07	0.11	-0.57	-0.27	0.16	-0.12	-0.03					
emotion_amusement-g.m	-0.39	0.11	-3.46	-0.62	-0.16	-0.44	-0.35					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Figure C1

Face-locked later parietal old-new (LPON) by emotion.



Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand-averages of the ROI mean amplitudes, contrasted for the implicit (old-new) task and all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **C** Topographies of the ERP distribution for familiar contrasted with novel faces. ROI channels are highlighted in pink.

doi: 10.53846/goediss-9841

N1 (voice-locked). N1 amplitudes were not significantly modulated by valence ($\chi^2(2) = 0.5$, $p = .779$).

Similarly, N1 amplitudes were not significantly modulated by emotion ($\chi^2(5) = 3.65$, $p = .601$).

Table C25

Statistical results for the (auditory) N1 mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	1.11	0.13	8.64	0.85	1.36	1.07	1.16	(Intercept)	-	-	-	-
valence_positive-g.m	-0.04	0.09	-0.44	-0.21	0.13	-0.06	-0.01	valence	0.5	2	.779	0.01
valence_negative-g.m	0.06	0.09	0.69	-0.13	0.23	0.03	0.09					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C26

Statistical results for the (auditory) N1 mean amplitudes by emotion

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	1.11	0.13	8.62	0.86	1.35	1.08	1.16	(Intercept)	-	-	-	-
emotion_yawning-g.m	-0.19	0.13	-1.40	-0.46	0.08	-0.23	-0.12					
emotion_elation-g.m	-0.06	0.13	-0.44	-0.33	0.20	-0.09	0.00					
emotion_disgust-g.m	0.13	0.13	0.95	-0.15	0.40	0.08	0.17	emotion	3.65	5	.601	0.02
emotion_anger-g.m	-0.02	0.13	-0.17	-0.31	0.23	-0.08	0.03					
emotion_amusement-g.m	-0.01	0.13	-0.06	-0.27	0.23	-0.04	0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

P2 (voice-locked). Due to influential observations (Cook's distance > 1) of one participant in the valence model, this participant was also excluded of the emotion model to allow comparisons of the results. P2 amplitudes were not significantly modulated by valence ($\chi^2(2) = 2.79$, $p = .247$).

However, when including emotion categories separately, there was a main effect of emotion on P2 amplitudes ($\chi^2(5) = 13.78$, $p = .017$). Post-hoc tests showed that there was a significant difference between the two neutral pre-specified emotion categories, yawning and throat-clearing ($diff_{\text{yaw-clt}} = -0.67$, $p = .017$). None of the other pairwise comparisons were significant.

Table C27

Statistical results for the (auditory) P2 mean amplitudes by valence

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	1.84	0.15	12.24	1.55	2.15	1.79	1.89	(Intercept)	-	-	-	-
valence_positive-g.m	-0.12	0.08	-1.45	-0.28	0.04	-0.15	-0.09	valence	2.79	2	.247	0.04
valence_negative-g.m	0.12	0.08	1.43	-0.04	0.28	0.09	0.15					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table C28

Statistical results for the (auditory) P2 mean amplitudes by emotion

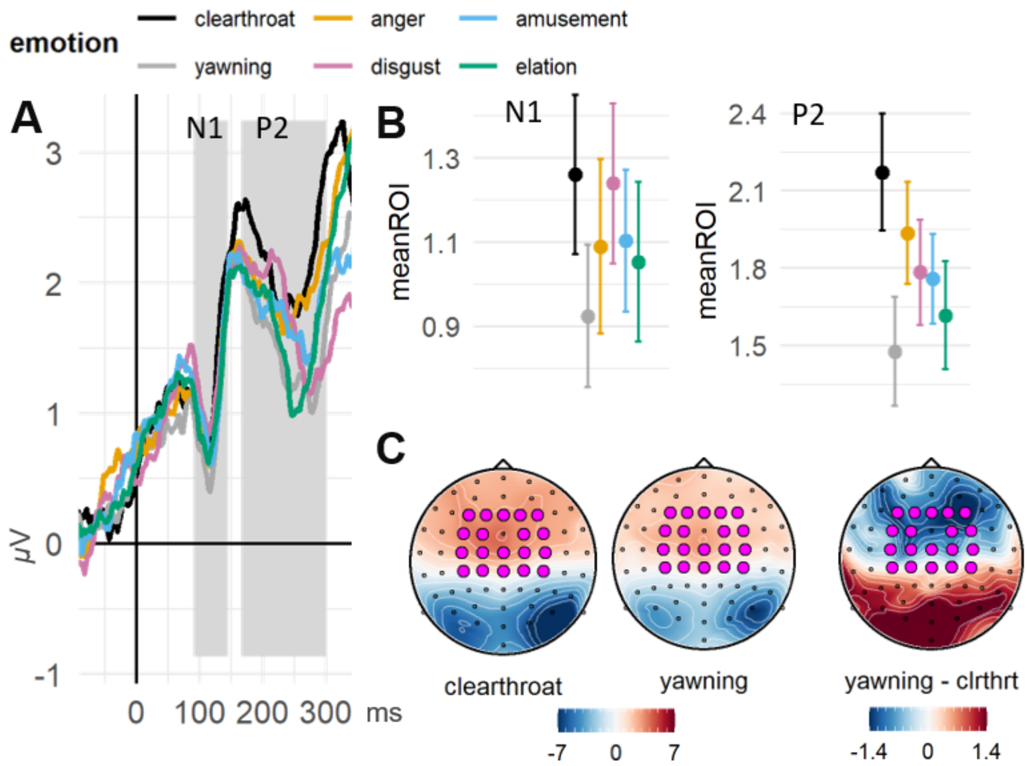
	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	1.85	0.15	12.33	1.58	2.16	1.80	1.90	(Intercept)	-	-	-	-
emotion_yawning-g.m	-0.32	0.13	-2.42	-0.56	-0.06	-0.37	-0.28					
emotion_elation-g.m	-0.16	0.13	-1.24	-0.41	0.09	-0.20	-0.12					
emotion_disgust-g.m	0.03	0.13	0.22	-0.24	0.27	-0.01	0.08	emotion	13.78	5	.017	0.07
emotion_anger-g.m	0.17	0.13	1.27	-0.08	0.43	0.12	0.21					
emotion_amusement-g.m	-0.07	0.13	-0.53	-0.33	0.20	-0.10	-0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Figure C2

Voice-locked N1 and P2 by emotion (refresher trials).



Notes: **A** Grand average ERP time series of the averaged ROI channels. The highlighted area displays the ROI time window. **B** Grand-averages of the ROI mean amplitudes of the N1 (left panel) and P2 (right panel), contrasted for all emotion conditions. Errorbars indicate +/- 1 SE of the mean. **C** Topographies of the ERP distribution of the P2 for the two neutral emotion categories ’throatclearing’ and ’yawning’ and their difference. ROI channels are highlighted in pink.

doi: 10.53846/goediss-9841

Pupil size. We used a desktop-mounted eye-tracker (EyeLink 1000 CL 1 - AAD01, SR Research, software version 4.56) to record the pupil size binocularly in pixels with a sampling rate of 500 Hz. Preprocessing was done in (R Core Team, 2020) and oriented to guidelines proposed by Kret & Sjak-Shie (2018). The continuous data was epoched around the onset of the face stimulus, and the mean of a 200 ms baseline time window was subtracted. Blinks were classified when samples of both eyes were missing. Other invalid, isolated, or implausible samples due to fast changes or deviations from a smoothed trend line were rejected. Additionally, trials were segmented into 60 bins, and outlier samples per bin (>3 *SD* from the bin mean) were excluded. If a trial contained less than 75% of valid samples, the whole trial was rejected. A smoothing 4 Hz filter was applied before trials were averaged by condition and participant. Pupil responses primarily served to identify eye blinks. Additionally, we tested valence differences in the refresher trials as a proxy for arousal without a preregistered hypothesis.

We analyzed pupil size during refresher trials to have an additional indicator for arousal of the face-voice pairs. However, model diagnostics indicated that several observations had a strong impact on model results, which hence might not be robust. Removing influential observations (Cook's distance >1) in several rounds led to other observations being classified as influential down to a remaining sample size of 30 participants[2]. However, there was no significant modulation of pupil size by valence for both, the full sample set ($\chi^2(2)$ = 3.19, p = .203) and the subset of 30 participants ($\chi^2(2)$ = 2.02, p = .365).

Table C29

Statistical results for pupil diameter during refresher trials

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-26.95	6.06	-4.45	-39.11	-15.14	-29.25	-25.49	(Intercept)	-	-	-	-
valence_positive-g.m	-0.87	3.64	-0.24	-8.04	6.19	-2.30	0.40	valence	3.19	2	.203	0.04
valence_negative-g.m	5.99	3.64	1.65	-1.46	13.10	4.22	7.26					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

[2] excluded participant IDs were: 4, 5, 9, 13, 15, 16, 17, 48, 53, 56

doi: 10.53846/goediss-9841

D Appendix of Study 5

Table D1

Statistical results for the P1 mean amplitudes by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.03	0.65	6.20	2.78	5.31	3.77	4.29	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.06	0.10	0.61	-0.13	0.26	0.01	0.10					
emotion_elation-g.m	0.06	0.10	0.58	-0.14	0.28	0.02	0.11					
emotion_disgust-g.m	-0.11	0.10	-1.04	-0.30	0.09	-0.16	-0.07	emotion	2.07	5	.839	0.00
emotion_anger-g.m	0.06	0.10	0.59	-0.14	0.25	0.03	0.09					
emotion_amusement-g.m	-0.05	0.10	-0.45	-0.25	0.15	-0.08	0.00					
task_valenceclass-g.m	-0.02	0.05	-0.51	-0.11	0.07	-0.05	-0.01	task	0.26	1	.607	0.00
emotion_yawning-g.m:task_valenceclass-g.m	0.07	0.10	0.70	-0.15	0.28	0.02	0.11					
emotion_elation-g.m:task_valenceclass-g.m	-0.06	0.10	-0.59	-0.26	0.14	-0.09	0.01					
emotion_disgust-g.m:task_valenceclass-g.m	0.01	0.10	0.13	-0.18	0.23	-0.03	0.04	emotion:task	0.96	5	.965	0.00
emotion_anger-g.m:task_valenceclass-g.m	-0.05	0.10	-0.48	-0.25	0.14	-0.08	-0.01					
emotion_amusement-g.m:task_valenceclass-g.m	0.02	0.10	0.22	-0.16	0.24	-0.01	0.05					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D2

Statistical results for the P1 peak amplitudes by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	7.21	0.63	11.52	6.08	8.50	6.91	7.44	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.21	0.12	1.81	-0.01	0.44	0.17	0.25					
emotion_elation-g.m	0.09	0.12	0.75	-0.14	0.32	0.04	0.13					
emotion_disgust-g.m	-0.14	0.12	-1.20	-0.35	0.09	-0.21	-0.09	emotion	5.33	5	.377	0.01
emotion_anger-g.m	0.00	0.12	0.04	-0.23	0.22	-0.02	0.03					
emotion_amusement-g.m	-0.08	0.12	-0.73	-0.30	0.13	-0.13	-0.04					
task_valenceclass-g.m	-0.05	0.05	-0.98	-0.15	0.04	-0.07	-0.04	task	0.99	1	.319	0.00
emotion_yawning-g.m:task_valenceclass-g.m	-0.04	0.12	-0.32	-0.27	0.18	-0.10	0.00					
emotion_elation-g.m:task_valenceclass-g.m	-0.05	0.12	-0.43	-0.26	0.18	-0.08	0.03					
emotion_disgust-g.m:task_valenceclass-g.m	0.02	0.12	0.21	-0.21	0.25	-0.03	0.05	emotion:task	1.41	5	.923	0.00
emotion_anger-g.m:task_valenceclass-g.m	-0.05	0.12	-0.43	-0.26	0.19	-0.08	-0.01					
emotion_amusement-g.m:task_valenceclass-g.m	0.12	0.12	1.06	-0.11	0.34	0.08	0.16					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D3

Statistical results for the P1 peak latency by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	100.51	1.64	61.24	97.23	103.90	100.07	100.94	(Intercept)	-	-	-	-
emotion_happy-g.m	-1.32	0.51	-2.61	-2.33	-0.38	-1.45	-1.09	emotion	6.94	2	.031	0.02
emotion_angry-g.m	0.76	0.51	1.50	-0.20	1.74	0.62	0.95					
stimsize_medium-g.m	-0.40	0.51	-0.78	-1.39	0.53	-0.60	-0.13	stimsize	40.24	2	<.001	0.13
stimsize_large-g.m	-2.63	0.51	-5.17	-3.58	-1.60	-2.85	-2.35					
emotion_happy-g.m:stimsize_medium-g.m	-0.40	0.72	-0.56	-1.82	0.94	-0.63	-0.19	emotion:stimsize	0.52	4	.971	0.00
emotion_angry-g.m:stimsize_medium-g.m	0.35	0.72	0.48	-1.05	1.73	0.10	0.56					
emotion_happy-g.m:stimsize_large-g.m	0.27	0.72	0.37	-1.05	1.69	-0.13	0.53					
emotion_angry-g.m:stimsize_large-g.m	-0.40	0.72	-0.56	-1.81	1.00	-0.63	-0.23					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect sizeNote that the inspection of residuals indicated a potential misfit of the model, possibly due to the temporal boundary of the ROI time window.

Table D4

Statistical results for the N170 mean amplitudes by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-6.94	0.59	-11.70	-8.09	-5.76	-7.13	-6.74	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.17	0.09	1.85	0.00	0.37	0.14	0.20	emotion	13.67	5	.018	0.03
emotion_elation-g.m	0.13	0.09	1.45	-0.04	0.30	0.10	0.16					
emotion_disgust-g.m	-0.28	0.09	-3.04	-0.44	-0.09	-0.30	-0.24					
emotion_anger-g.m	0.08	0.09	0.88	-0.10	0.26	0.05	0.11					
emotion_amusement-g.m	-0.03	0.09	-0.37	-0.22	0.16	-0.07	0.00					
task_valenceclass-g.m	-0.23	0.04	-5.78	-0.31	-0.16	-0.26	-0.20	task	32.96	1	<.001	0.08
emotion_yawning-g.m:task_valenceclass-g.m	0.02	0.09	0.20	-0.16	0.19	-0.01	0.05	emotion:task	3.58	5	.612	0.01
emotion_elation-g.m:task_valenceclass-g.m	-0.08	0.09	-0.85	-0.24	0.09	-0.11	-0.05					
emotion_disgust-g.m:task_valenceclass-g.m	0.07	0.09	0.74	-0.11	0.24	0.03	0.11					
emotion_anger-g.m:task_valenceclass-g.m	-0.12	0.09	-1.29	-0.30	0.06	-0.14	-0.08					
emotion_amusement-g.m:task_valenceclass-g.m	0.01	0.09	0.09	-0.17	0.18	-0.02	0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D5

Statistical results for the N170 peak amplitudes by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-11.11	0.71	-15.76	-12.49	-9.69	-11.35	-10.86	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.17	0.10	1.68	-0.03	0.37	0.14	0.21	emotion	11.58	5	.041	0.03
emotion_elation-g.m	0.05	0.10	0.52	-0.14	0.27	0.02	0.09					
emotion_disgust-g.m	-0.31	0.10	-3.01	-0.51	-0.12	-0.35	-0.27					
emotion_anger-g.m	0.11	0.10	1.04	-0.08	0.32	0.08	0.16					
emotion_amusement-g.m	0.04	0.10	0.36	-0.18	0.24	0.00	0.07					
task_valenceclass-g.m	-0.04	0.05	-0.86	-0.12	0.05	-0.06	-0.02	task	0.76	1	.383	0.00
emotion_yawning-g.m:task_valenceclass-g.m	0.07	0.10	0.72	-0.11	0.28	0.03	0.11	emotion:task	1.66	5	.894	0.00
emotion_elation-g.m:task_valenceclass-g.m	-0.06	0.10	-0.57	-0.26	0.13	-0.10	-0.03					
emotion_disgust-g.m:task_valenceclass-g.m	0.08	0.10	0.76	-0.14	0.28	0.04	0.12					
emotion_anger-g.m:task_valenceclass-g.m	-0.04	0.10	-0.37	-0.24	0.17	-0.08	-0.01					
emotion_amusement-g.m:task_valenceclass-g.m	-0.06	0.10	-0.62	-0.26	0.13	-0.10	-0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D6

Statistical results for the N170 peak latency by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	152.28	1.76	86.57	148.60	155.69	151.59	152.71	(Intercept)	-	-	-	-
emotion_happy-g.m	0.05	0.26	0.18	-0.45	0.57	-0.06	0.14	emotion	3.66	2	.160	0.01
emotion_angry-g.m	0.41	0.26	1.54	-0.07	0.94	0.32	0.50					
stimsize_medium-g.m	-1.90	0.26	-7.21	-2.42	-1.38	-1.98	-1.76	stimsize	337.03	2	<.001	1.87
stimsize_large-g.m	-4.31	0.26	-16.34	-4.78	-3.76	-4.45	-4.12					
emotion_happy-g.m:stimsize_medium-g.m	0.02	0.37	0.04	-0.78	0.76	-0.08	0.11	emotion:stimsize	2.21	4	.697	0.01
emotion_angry-g.m:stimsize_medium-g.m	0.29	0.37	0.78	-0.45	1.03	0.17	0.39					
emotion_happy-g.m:stimsize_large-g.m	-0.31	0.37	-0.83	-1.01	0.43	-0.37	-0.21					
emotion_angry-g.m:stimsize_large-g.m	-0.18	0.37	-0.48	-0.92	0.56	-0.29	-0.07					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D7

Statistical results for the EPN mean amplitudes by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-2.27	0.49	-4.62	-3.28	-1.37	-2.45	-2.12	(Intercept)	-	-	-	-
emotion_yawning-g.m	0.30	0.12	2.51	0.07	0.55	0.25	0.36	emotion	21.61	5	<.001	0.05
emotion_elation-g.m	0.20	0.12	1.64	-0.04	0.43	0.17	0.24					
emotion_disgust-g.m	-0.46	0.12	-3.79	-0.69	-0.23	-0.50	-0.40					
emotion_anger-g.m	-0.15	0.12	-1.25	-0.39	0.10	-0.19	-0.11					
emotion_amusement-g.m	-0.01	0.12	-0.12	-0.24	0.22	-0.05	0.02					
task_valenceclass-g.m	-0.01	0.05	-0.20	-0.12	0.09	-0.04	0.03	task	0.04	1	.838	0.00
emotion_yawning-g.m:task_valenceclass-g.m	-0.05	0.12	-0.42	-0.28	0.19	-0.09	-0.01	emotion:task	1.47	5	.917	0.00
emotion_elation-g.m:task_valenceclass-g.m	-0.04	0.12	-0.31	-0.28	0.21	-0.08	0.00					
emotion_disgust-g.m:task_valenceclass-g.m	0.01	0.12	0.05	-0.24	0.23	-0.03	0.03					
emotion_anger-g.m:task_valenceclass-g.m	-0.08	0.12	-0.68	-0.32	0.15	-0.12	-0.04					
emotion_amusement-g.m:task_valenceclass-g.m	0.10	0.12	0.84	-0.12	0.34	0.06	0.14					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D8

Statistical results for the LPC mean amplitudes by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	5.35	0.52	10.31	4.40	6.35	5.02	5.53	(Intercept)	-	-	-	-
emotion_yawning-g.m	-0.09	0.12	-0.72	-0.32	0.17	-0.14	-0.06	emotion	5.01	5	.414	0.01
emotion_elation-g.m	0.06	0.12	0.52	-0.18	0.30	0.00	0.11					
emotion_disgust-g.m	-0.05	0.12	-0.37	-0.28	0.21	-0.09	0.01					
emotion_anger-g.m	0.16	0.12	1.33	-0.07	0.42	0.12	0.20					
emotion_amusement-g.m	0.10	0.12	0.83	-0.13	0.35	0.06	0.14					
task_valenceclass-g.m	0.09	0.05	1.55	-0.01	0.19	0.05	0.13	task	2.46	1	.117	0.01
emotion_yawning-g.m:task_valenceclass-g.m	-0.18	0.12	-1.49	-0.41	0.07	-0.21	-0.16	emotion:task	6.2	5	.287	0.01
emotion_elation-g.m:task_valenceclass-g.m	0.05	0.12	0.39	-0.19	0.31	0.01	0.07					
emotion_disgust-g.m:task_valenceclass-g.m	0.12	0.12	1.00	-0.12	0.35	0.09	0.16					
emotion_anger-g.m:task_valenceclass-g.m	0.03	0.12	0.27	-0.21	0.26	0.00	0.07					
emotion_amusement-g.m:task_valenceclass-g.m	0.16	0.12	1.29	-0.08	0.40	0.13	0.19					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D9

Statistical results for accuracy in the naturalness classification task by emotion and stimulus size

	β	SE	z	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p
(Intercept)	5.00	0.19	25.72	4.66	5.44	4.93	5.07	(Intercept)	-	-	-
emotion_yawning-g.m	-0.26	0.24	-1.05	-0.73	0.26	-0.38	-0.03				
emotion_anger-g.m	0.00	0.23	-0.01	-0.47	0.48	-0.08	0.08				
emotion_disgust-g.m	-0.05	0.24	-0.21	-0.52	0.53	-0.18	0.12	emotion	2.74	5	.740
emotion_amusement-g.m	0.21	0.20	1.05	-0.21	0.72	0.11	0.30				
emotion_elation-g.m	-0.18	0.20	-0.88	-0.62	0.27	-0.29	-0.01				
task_valenceclass-g.m	-0.57	0.14	-4.03	-0.89	-0.30	-0.63	-0.50	task	14.16	1	<.001
emotion_yawning-g.m:task_valenceclass-g.m	-0.15	0.13	-1.18	-0.41	0.09	-0.40	-0.03				
emotion_anger-g.m:task_valenceclass-g.m	0.33	0.13	2.60	0.08	0.60	0.17	0.42				
emotion_disgust-g.m:task_valenceclass-g.m	-0.11	0.13	-0.84	-0.40	0.15	-0.30	0.18	emotion:task	12.72	5	.026
emotion_amusement-g.m:task_valenceclass-g.m	-0.29	0.16	-1.89	-0.69	0.01	-0.43	-0.19				
emotion_elation-g.m:task_valenceclass-g.m	0.21	0.12	1.73	-0.02	0.46	-0.03	0.35				

Notes: beta = model estimate, SE = standard error of the estimate, CI = lower and upper 95 Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test

Table D10

Statistical results for mean response times by emotion and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	1041.36	43.43	23.98	958.91	1121.84	1023.94	1053.68	(Intercept)	-	-	-	-
emotion_happy-g.m	-39.22	5.25	-7.47	-49.76	-28.80	-41.26	-35.89	emotion	85.57	2	<.001	0.31
emotion_angry-g.m	48.30	5.25	9.20	38.11	59.20	44.82	49.86					
stimsize_large-g.m	13.31	5.25	2.54	3.52	23.72	11.05	15.01	stimsize	7.37	2	.025	0.02
stimsize_medium-g.m	-10.83	5.25	-2.06	-20.97	-1.10	-13.21	-8.41					
emotion_happy-g.m:stimsize_large-g.m	-0.92	7.42	-0.12	-15.58	12.97	-3.45	2.49					
emotion_angry-g.m:stimsize_large-g.m	-5.87	7.42	-0.79	-20.48	9.30	-8.04	-2.89	emotion:stimsize	5.25	4	.263	0.02
emotion_happy-g.m:stimsize_medium-g.m	11.36	7.42	1.53	-3.15	25.56	9.31	13.61					
emotion_angry-g.m:stimsize_medium-g.m	-8.86	7.42	-1.19	-23.50	5.07	-10.62	-4.01					

Notes: β = model estimate, SE = standard error of the estimate, CI = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D11

Statistical results for the P1 mean amplitudes by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	3.49	0.34	10.36	2.80	4.22	3.34	3.60	(Intercept)	-	-	-	-
facescr_face-g.m	-0.36	0.08	-4.69	-0.51	-0.21	-0.38	-0.32	facescr	21.42	1	<.001	0.11
stimsize_medium-g.m	0.15	0.11	1.36	-0.06	0.36	0.11	0.19	stimsize	14.03	2	<.001	0.07
stimsize_large-g.m	0.26	0.11	2.36	0.03	0.46	0.22	0.31					
facescr_face-g.m:stimsize_medium-g.m	-0.04	0.11	-0.34	-0.24	0.16	-0.05	-0.02	facescr:stimsize	3.80	2	.149	0.02
facescr_face-g.m:stimsize_large-g.m	-0.16	0.11	-1.48	-0.37	0.06	-0.18	-0.15					

Notes: β = model estimate, SE = standard error of the estimate, CI = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D12

Statistical results for the P1 peak amplitudes by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	5.24	0.38	13.88	4.49	5.97	5.08	5.37	(Intercept)	-	-	-	-
facescr_face-g.m	-0.43	0.08	-5.37	-0.60	-0.27	-0.47	-0.40	facescr	27.55	1	<.001	0.15
stimsize_medium-g.m	0.16	0.11	1.43	-0.06	0.38	0.13	0.19	stimsize	44.34	2	<.001	0.25
stimsize_large-g.m	0.59	0.11	5.18	0.37	0.82	0.54	0.63					
facescr_face-g.m:stimsize_medium-g.m	-0.01	0.11	-0.08	-0.23	0.21	-0.02	0.01	facescr:stimsize	6.84	2	.033	0.03
facescr_face-g.m:stimsize_large-g.m	-0.25	0.11	-2.22	-0.47	-0.02	-0.28	-0.23					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D13

Statistical results for the P1 peak latency by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	99.64	1.48	67.29	96.49	102.57	99.17	100.06	(Intercept)	-	-	-	-
facescr_face-g.m	0.87	0.51	1.69	-0.10	1.87	0.55	1.19	facescr	2.92	1	.087	0.01
stimsize_medium-g.m	-0.83	0.72	-1.15	-2.27	0.64	-1.17	-0.58	stimsize	37.11	2	<.001	0.20
stimsize_large-g.m	-3.46	0.72	-4.79	-4.87	-2.06	-3.80	-3.25					
facescr_face-g.m:stimsize_medium-g.m	0.44	0.72	0.60	-0.99	2.01	0.19	0.62	facescr:stimsize	3.26	2	.196	0.02
facescr_face-g.m:stimsize_large-g.m	0.84	0.72	1.16	-0.56	2.20	0.60	1.04					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect sizeNote that the inspection of residuals indicated a potential misfit of the model, possibly due to the temporal boundary of the ROI timewindow.

Table D14

Statistical results for the N170 mean amplitudes by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-3.30	0.56	-5.92	-4.41	-2.26	-3.48	-3.00	(Intercept)	-	-	-	-
facescr_face-g.m	-3.45	0.13	-26.29	-3.69	-3.19	-3.54	-3.34	facescr	302.79	1	<.001	3.54
stimsize_medium-g.m	-0.12	0.19	-0.65	-0.49	0.24	-0.14	-0.10	stimsize	1.86	2	.395	0.01
stimsize_large-g.m	0.25	0.19	1.35	-0.13	0.61	0.22	0.30					
facescr_face-g.m:stimsize_medium-g.m	0.15	0.19	0.80	-0.22	0.51	0.13	0.17	facescr:stimsize	3.93	2	.140	0.02
facescr_face-g.m:stimsize_large-g.m	0.22	0.19	1.16	-0.15	0.58	0.18	0.23					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D15

Statistical results for the N170 peak amplitudes by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-7.06	0.64	-11.02	-8.34	-5.84	-7.23	-6.73	(Intercept)	-	-	-	-
facescr_face-g.m	-4.07	0.15	-26.97	-4.37	-3.75	-4.17	-3.95	facescr	310.77	1	<.001	3.73
stimsize_medium-g.m	-0.35	0.21	-1.66	-0.76	0.07	-0.38	-0.33	stimsize	8.01	2	.018	0.04
stimsize_large-g.m	-0.24	0.21	-1.15	-0.63	0.17	-0.28	-0.20					
facescr_face-g.m:stimsize_medium-g.m	0.12	0.21	0.55	-0.28	0.49	0.09	0.14	facescr:stimsize	5.38	2	.068	0.03
facescr_face-g.m:stimsize_large-g.m	0.36	0.21	1.67	-0.11	0.76	0.32	0.38					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D16

Statistical results for the N170 peak latency by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	150.96	1.81	83.52	147.30	154.58	150.12	151.34	(Intercept)	-	-	-	-
facescr_face-g.m	0.68	0.36	1.90	0.00	1.41	0.54	0.91	facescr	3.69	1	.055	0.02
stimsize_medium-g.m	-1.76	0.51	-3.47	-2.79	-0.73	-1.89	-1.66	stimsize	104.57	2	<.001	0.71
stimsize_large-g.m	-3.98	0.51	-7.86	-4.95	-2.93	-4.15	-3.85					
facescr_face-g.m:stimsize_medium-g.m	-0.22	0.51	-0.44	-1.21	0.75	-0.30	-0.11	facescr:stimsize	0.53	2	.766	0.00
facescr_face-g.m:stimsize_large-g.m	-0.14	0.51	-0.27	-1.10	0.87	-0.25	0.01					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect sizeNote that one subject was removed from this model due to influential observations.

Table D17

Statistical results for the EPN mean amplitudes by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	0.42	0.48	0.88	-0.47	1.34	0.17	0.59	(Intercept)	-	-	-	-
facescr_face-g.m	-1.70	0.11	-15.23	-1.92	-1.48	-1.75	-1.58	facescr	156.68	1	<.001	1.19
stimsize_medium-g.m	-0.10	0.16	-0.62	-0.41	0.22	-0.12	-0.08	stimsize	7.76	2	.021	0.04
stimsize_large-g.m	0.42	0.16	2.66	0.09	0.72	0.38	0.44					
facescr_face-g.m:stimsize_medium-g.m	0.13	0.16	0.84	-0.21	0.44	0.11	0.15	facescr:stimsize	2.52	2	.283	0.01
facescr_face-g.m:stimsize_large-g.m	0.12	0.16	0.73	-0.18	0.42	0.09	0.15					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D18

Statistical results for the LPC mean amplitudes by facial intactness and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.07	0.33	12.46	3.44	4.68	3.96	4.17	(Intercept)	-	-	-	-
facescr_face-g.m	1.62	0.09	17.20	1.43	1.81	1.54	1.69	facescr	184.64	1	<.001	1.52
stimsize_medium-g.m	0.00	0.13	0.03	-0.26	0.27	-0.02	0.02	stimsize	2.94	2	.230	0.01
stimsize_large-g.m	0.19	0.13	1.46	-0.06	0.45	0.16	0.21					
facescr_face-g.m:stimsize_medium-g.m	0.02	0.13	0.15	-0.22	0.28	0.01	0.04	facescr:stimsize	0.25	2	.881	0.00
facescr_face-g.m:stimsize_large-g.m	0.04	0.13	0.34	-0.21	0.29	0.03	0.06					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D19

Statistical results for the P1 mean amplitudes by emotion (incl. scrambled) and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	3.31	0.32	10.43	2.68	3.91	3.19	3.42	(Intercept)	-	-	-	-
emotion_scrambled-g.m	0.54	0.09	6.19	0.36	0.72	0.47	0.58	emotion	40.07	3	<.001	0.10
emotion_happy-g.m	-0.29	0.09	-3.31	-0.45	-0.13	-0.31	-0.26					
emotion_angry-g.m	-0.06	0.09	-0.70	-0.23	0.11	-0.08	-0.04					
stimsize_medium-g.m	0.13	0.07	1.80	-0.01	0.27	0.09	0.17	stimsize	18.52	2	<.001	0.04
stimsize_large-g.m	0.18	0.07	2.47	0.03	0.32	0.14	0.22					
emotion_scrambled-g.m:stimsize_medium-g.m	0.06	0.12	0.45	-0.19	0.28	0.04	0.08	emotion:stimsize	8.40	6	.210	0.02
emotion_happy-g.m:stimsize_medium-g.m	-0.05	0.12	-0.43	-0.28	0.17	-0.07	-0.04					
emotion_angry-g.m:stimsize_medium-g.m	0.05	0.12	0.42	-0.20	0.28	0.03	0.07					
emotion_scrambled-g.m:stimsize_large-g.m	0.24	0.12	1.95	-0.02	0.48	0.22	0.27					
emotion_happy-g.m:stimsize_large-g.m	0.04	0.12	0.30	-0.20	0.28	0.00	0.05					
emotion_angry-g.m:stimsize_large-g.m	-0.11	0.12	-0.88	-0.35	0.12	-0.13	-0.08					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D20

Statistical results for the P1 peak amplitudes by emotion (incl. scrambled) and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	5.02	0.35	14.18	4.34	5.73	4.88	5.15	(Intercept)	-	-	-	-
emotion_scrambled-g.m	0.65	0.09	6.99	0.48	0.83	0.60	0.70	emotion	50.20	3	<.001	0.12
emotion_happy-g.m	-0.34	0.09	-3.70	-0.52	-0.15	-0.36	-0.32					
emotion_angry-g.m	-0.08	0.09	-0.87	-0.26	0.10	-0.10	-0.06					
stimsize_medium-g.m	0.16	0.08	2.10	0.01	0.30	0.13	0.19	stimsize	69.08	2	<.001	0.17
stimsize_large-g.m	0.46	0.08	6.12	0.31	0.63	0.42	0.50					
emotion_scrambled-g.m:stimsize_medium-g.m	0.01	0.13	0.10	-0.23	0.27	-0.01	0.04	emotion:stimsize	12.25	6	.057	0.03
emotion_happy-g.m:stimsize_medium-g.m	-0.05	0.13	-0.39	-0.30	0.21	-0.08	-0.03					
emotion_angry-g.m:stimsize_medium-g.m	0.03	0.13	0.25	-0.23	0.27	0.01	0.07					
emotion_scrambled-g.m:stimsize_large-g.m	0.38	0.13	2.89	0.11	0.64	0.35	0.42					
emotion_happy-g.m:stimsize_large-g.m	-0.04	0.13	-0.31	-0.30	0.22	-0.07	-0.02					
emotion_angry-g.m:stimsize_large-g.m	-0.15	0.13	-1.12	-0.39	0.11	-0.17	-0.12					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D21

Statistical results for the P1 peak latency by emotion (incl. scrambled) and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	100.08	1.53	65.57	96.86	102.92	99.62	100.50	(Intercept)	-	-	-	-
emotion_happy-g.m	-0.89	0.62	-1.43	-2.12	0.32	-1.05	-0.75					
emotion_angry-g.m	1.19	0.62	1.92	0.00	2.44	0.97	1.54	emotion	9.64	3	.022	0.02
emotion_neutral-g.m	1.00	0.62	1.60	-0.20	2.23	0.79	1.14					
stimsize_medium-g.m	-0.61	0.51	-1.21	-1.60	0.35	-0.88	-0.38	stimsize	57.24	2	<.001	0.14
stimsize_large-g.m	-3.04	0.51	-6.00	-3.99	-2.04	-3.28	-2.80					
emotion_happy-g.m:stimsize_medium-g.m	-0.18	0.88	-0.21	-1.94	1.45	-0.36	0.06					
emotion_angry-g.m:stimsize_medium-g.m	0.57	0.88	0.64	-1.27	2.22	0.31	0.76					
emotion_neutral-g.m:stimsize_medium-g.m	0.27	0.88	0.31	-1.42	2.06	-0.19	0.61	emotion:stimsize	5.35	6	.499	0.01
emotion_happy-g.m:stimsize_large-g.m	0.68	0.88	0.78	-1.10	2.49	0.29	0.93					
emotion_angry-g.m:stimsize_large-g.m	0.02	0.88	0.02	-1.68	1.75	-0.17	0.21					
emotion_neutral-g.m:stimsize_large-g.m	0.55	0.88	0.63	-1.16	2.30	0.38	0.84					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D22

Statistical results for the N170 mean amplitudes by emotion (incl. scrambled) and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-5.03	0.59	-8.52	-6.14	-3.86	-5.19	-4.74	(Intercept)	-	-	-	-
emotion_scrambled-g.m	5.17	0.14	35.84	4.89	5.45	5.01	5.31					
emotion_happy-g.m	-1.78	0.14	-12.35	-2.07	-1.48	-1.84	-1.73	emotion	611.31	3	<.001	3.01
emotion_angry-g.m	-2.02	0.14	-13.99	-2.30	-1.75	-2.07	-1.96					
stimsize_medium-g.m	-0.05	0.12	-0.39	-0.29	0.19	-0.07	-0.03	stimsize	11.06	2	.004	0.03
stimsize_large-g.m	0.36	0.12	3.04	0.14	0.58	0.32	0.40					
emotion_scrambled-g.m:stimsize_medium-g.m	-0.22	0.20	-1.09	-0.63	0.16	-0.26	-0.20					
emotion_happy-g.m:stimsize_medium-g.m	0.14	0.20	0.71	-0.22	0.56	0.12	0.17					
emotion_angry-g.m:stimsize_medium-g.m	0.06	0.20	0.31	-0.35	0.44	0.04	0.10	emotion:stimsize	7.50	6	.277	0.02
emotion_scrambled-g.m:stimsize_large-g.m	-0.32	0.20	-1.58	-0.71	0.06	-0.35	-0.27					
emotion_happy-g.m:stimsize_large-g.m	0.10	0.20	0.47	-0.31	0.47	0.07	0.12					
emotion_angry-g.m:stimsize_large-g.m	0.10	0.20	0.49	-0.29	0.50	0.07	0.13					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Table D23

Statistical results for the N170 peak amplitudes by emotion (incl. scrambled) and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-9.09	0.71	-12.85	-10.44	-7.75	-9.27	-8.78	(Intercept)	-	-	-	-
emotion_scrambled-g.m	6.10	0.16	38.21	5.81	6.42	5.92	6.25	emotion	653.86	3	<.001	3.42
emotion_happy-g.m	-2.11	0.16	-13.25	-2.43	-1.78	-2.18	-2.05					
emotion_angry-g.m	-2.33	0.16	-14.58	-2.65	-2.02	-2.39	-2.27					
stimsize_medium-g.m	-0.30	0.13	-2.27	-0.55	-0.04	-0.31	-0.28	stimsize	8.90	2	.012	0.02
stimsize_large-g.m	-0.07	0.13	-0.51	-0.32	0.18	-0.10	-0.03					
emotion_scrambled-g.m:stimsize_medium-g.m	-0.17	0.23	-0.77	-0.60	0.27	-0.21	-0.14	emotion:stimsize	11.02	6	.088	0.03
emotion_happy-g.m:stimsize_medium-g.m	0.12	0.23	0.54	-0.31	0.55	0.09	0.15					
emotion_angry-g.m:stimsize_medium-g.m	0.04	0.23	0.19	-0.37	0.49	0.01	0.08					
emotion_scrambled-g.m:stimsize_large-g.m	-0.53	0.23	-2.36	-0.95	-0.08	-0.56	-0.48					
emotion_happy-g.m:stimsize_large-g.m	0.16	0.23	0.69	-0.30	0.60	0.12	0.19					
emotion_angry-g.m:stimsize_large-g.m	0.23	0.23	1.03	-0.19	0.65	0.19	0.27					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D24

Statistical results for the N170 peak latency by emotion (incl. scrambled) and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	151.82	1.76	86.35	148.27	155.19	151.11	152.22	(Intercept)	-	-	-	-
emotion_happy-g.m	0.50	0.43	1.18	-0.38	1.34	0.38	0.69	emotion	11.83	3	.008	0.03
emotion_angry-g.m	0.86	0.43	2.01	0.04	1.70	0.77	1.00					
emotion_neutral-g.m	0.00	0.43	0.00	-0.83	0.79	-0.17	0.21					
stimsize_medium-g.m	-1.68	0.35	-4.79	-2.35	-0.98	-1.87	-1.56	stimsize	229.54	2	<.001	0.68
stimsize_large-g.m	-4.15	0.35	-11.86	-4.85	-3.51	-4.26	-4.05					
emotion_happy-g.m:stimsize_medium-g.m	-0.21	0.61	-0.35	-1.42	1.00	-0.33	-0.08	emotion:stimsize	4.71	6	.582	0.01
emotion_angry-g.m:stimsize_medium-g.m	0.07	0.61	0.11	-1.18	1.21	-0.08	0.17					
emotion_neutral-g.m:stimsize_medium-g.m	-0.54	0.61	-0.89	-1.73	0.67	-0.61	-0.38					
emotion_happy-g.m:stimsize_large-g.m	-0.47	0.61	-0.77	-1.68	0.78	-0.56	-0.34					
emotion_angry-g.m:stimsize_large-g.m	-0.34	0.61	-0.56	-1.62	0.84	-0.43	-0.16					
emotion_neutral-g.m:stimsize_large-g.m	0.33	0.61	0.54	-0.88	1.47	0.16	0.40					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect sizeNote that one subject was removed from this model due to influential observations.

doi: 10.53846/goediss-9841

Table D25

Statistical results for the EPN mean amplitudes by emotion (incl. scrambled) and stimulus size

	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	-0.43	0.52	-0.82	-1.40	0.62	-0.69	-0.26	(Intercept)	-	-	-	-
emotion_scrambled-g.m	2.55	0.12	20.46	2.28	2.82	2.37	2.63					
emotion_happy-g.m	-1.13	0.12	-9.06	-1.37	-0.87	-1.16	-1.05	emotion	306.20	3	<.001	1.01
emotion_angry-g.m	-0.98	0.12	-7.89	-1.24	-0.75	-1.03	-0.91					
stimsize_medium-g.m	-0.03	0.10	-0.32	-0.22	0.17	-0.05	-0.01	stimsize	27.33	2	<.001	0.06
stimsize_large-g.m	0.48	0.10	4.69	0.26	0.67	0.44	0.51					
emotion_scrambled-g.m:stimsize_medium-g.m	-0.20	0.18	-1.12	-0.57	0.16	-0.23	-0.16					
emotion_happy-g.m:stimsize_medium-g.m	0.13	0.18	0.73	-0.22	0.45	0.10	0.15					
emotion_angry-g.m:stimsize_medium-g.m	0.16	0.18	0.90	-0.20	0.51	0.13	0.18	emotion:stimsize	6.68	6	.351	0.02
emotion_scrambled-g.m:stimsize_large-g.m	-0.17	0.18	-0.99	-0.50	0.16	-0.22	-0.13					
emotion_happy-g.m:stimsize_large-g.m	0.17	0.18	0.98	-0.15	0.52	0.15	0.20					
emotion_angry-g.m:stimsize_large-g.m	-0.01	0.18	-0.03	-0.35	0.34	-0.04	0.03					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

Table D26

Statistical results for the LPC mean amplitudes by emotion (incl. scrambled) and stimulus size

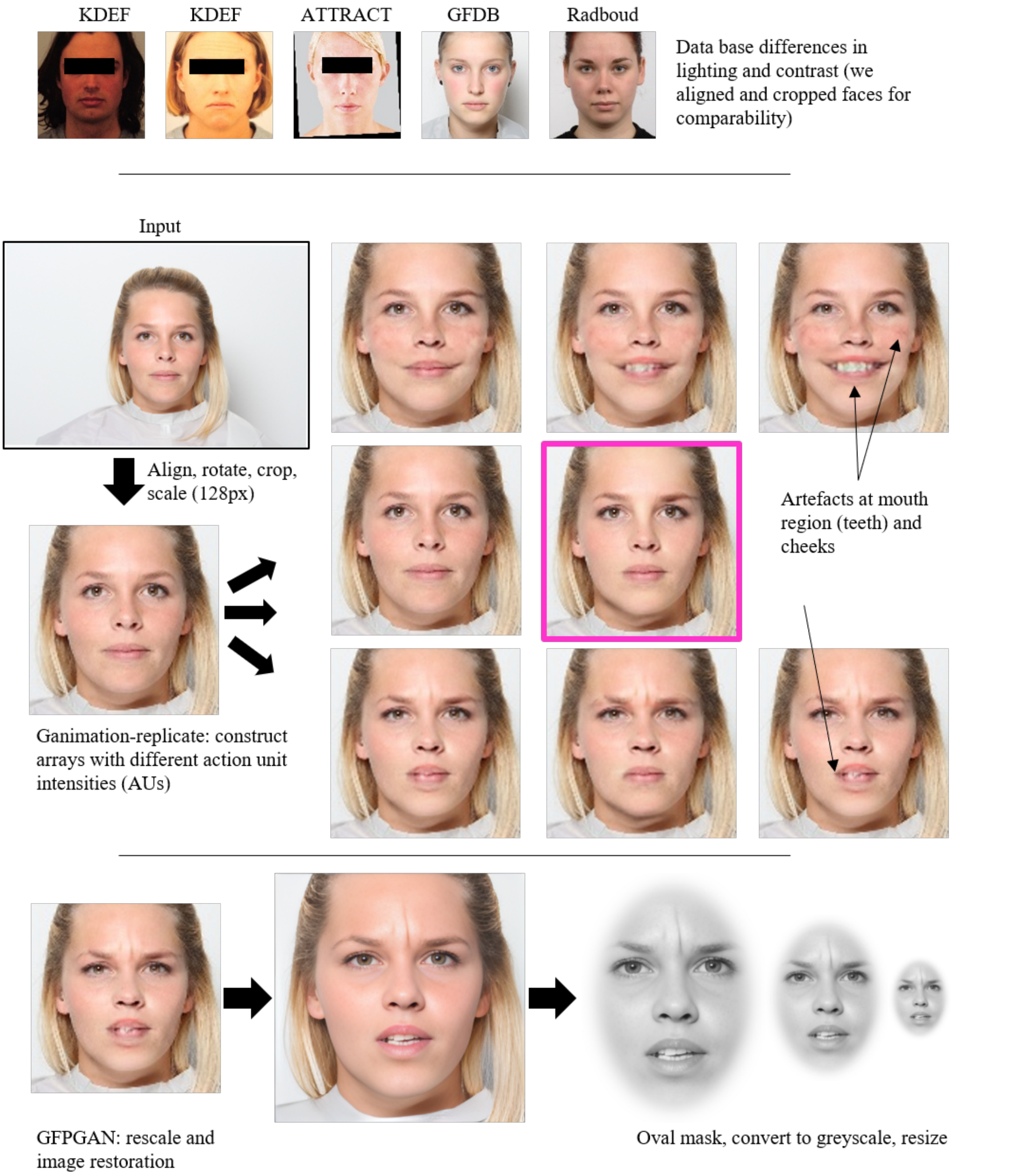
	β	SE	t	CI_l	CI_u	$Stab_{min}$	$Stab_{max}$	LRT:Model	χ^2	df	p	f^2
(Intercept)	4.88	0.37	13.08	4.24	5.59	4.75	4.99	(Intercept)	-	-	-	-
emotion_scrambled-g.m	-2.43	0.11	-22.92	-2.62	-2.22	-2.53	-2.31					
emotion_happy-g.m	1.05	0.11	9.91	0.85	1.25	1.01	1.09	emotion	354.61	3	<.001	1.24
emotion_angry-g.m	0.73	0.11	6.92	0.53	0.95	0.69	0.78					
stimsize_medium-g.m	0.01	0.09	0.17	-0.15	0.19	0.00	0.03	stimsize	9.04	2	.011	0.02
stimsize_large-g.m	0.22	0.09	2.50	0.05	0.38	0.19	0.23					
emotion_scrambled-g.m:stimsize_medium-g.m	-0.03	0.15	-0.20	-0.33	0.26	-0.06	-0.01					
emotion_happy-g.m:stimsize_medium-g.m	0.11	0.15	0.75	-0.17	0.40	0.08	0.13					
emotion_angry-g.m:stimsize_medium-g.m	-0.09	0.15	-0.59	-0.40	0.18	-0.11	-0.05	emotion:stimsize	1.22	6	.976	0.00
emotion_scrambled-g.m:stimsize_large-g.m	-0.07	0.15	-0.45	-0.36	0.21	-0.09	-0.04					
emotion_happy-g.m:stimsize_large-g.m	-0.05	0.15	-0.35	-0.34	0.25	-0.08	-0.01					
emotion_angry-g.m:stimsize_large-g.m	0.10	0.15	0.69	-0.20	0.39	0.07	0.14					

Notes: β = model estimate, *SE* = standard error of the estimate, *CI* = lower and upper 95% bootstrapped confidence intervals, Stab = estimate ranges leaving out one participant at a time, LRT = Likelihood ratio test, f^2 = Cohen's f^2 effect size

doi: 10.53846/goediss-9841

Figure D1

Illustration of the creation of expressive faces for the present study.

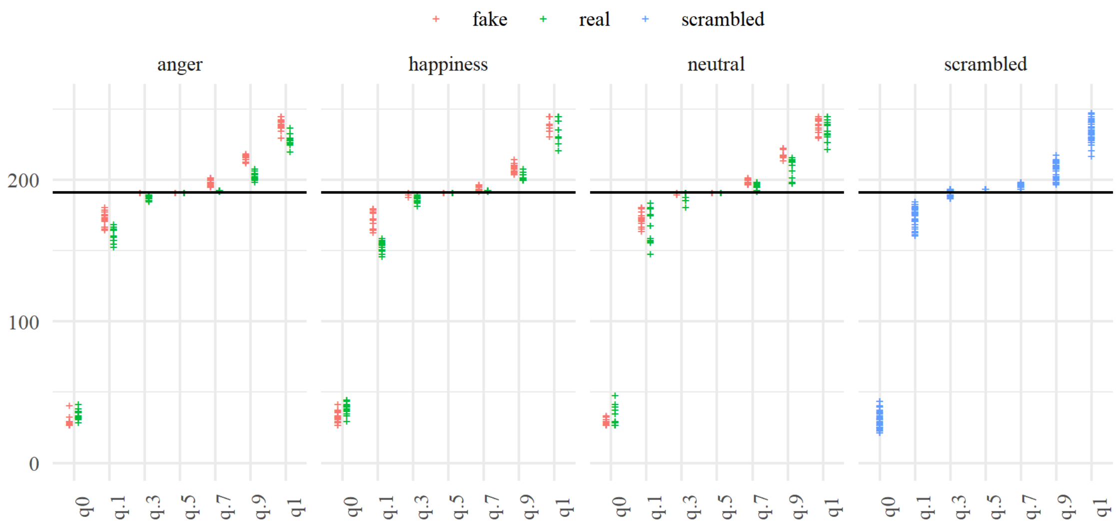


Notes: The top row show example for different image properties between frequently used face databases: KDEF: Karolinska Directed Emotional Faces (Garrido, 2017, *Front Psych, 8*); ATTRACT: face database validated for attractiveness (Schacht, 2008, *CABN, 8*); GFDB: Goettingen Face Database (Kulke, 2017, *osf.io/4knpf*); Radboud: Radboud Faces Database (Langner, 2010, *Cogn Emot, 24*). Stimulus processing starts with alignment, rotation, scaling and cropping of the input image, automatically with *OpenFace*. The 128 $\times$ 128 px image is the input for the first generative adversarial network (GAN), 'Ganimation-replicate'. An array of intensities for 17 available action units has to be specified and passed to the model. Some example outputs are displayed on the right. With increasing intensity, the model produced artefacts, especially at the mouth and cheek region for happy and angry stimuli. Some were diminished by the next step, restoring and upscaling the expressive images using GFPGAN. Finally, an oval mask with gaussian blur was applied to the image, the image was normalized and resized. Note, that we also created instances of model-based neutral expressions with the first GAN (highlighted in pink).

doi: 10.53846/goediss-9841

Figure D2

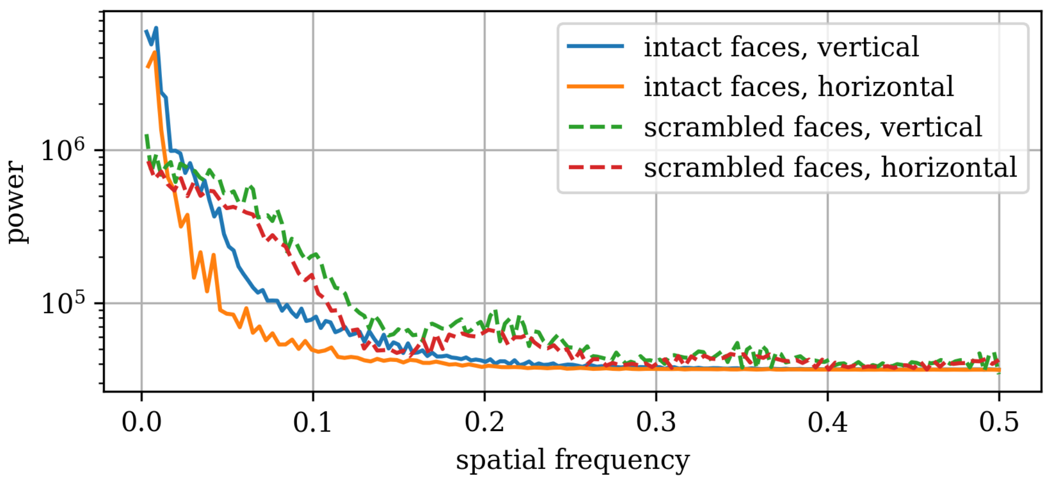
Quantiles of pixel RGB values as a measure of luminance per stimulus category.



Notes: Every cross represents an individual stimulus. The black horizontal line indicates the background luminance which corresponded to the median of stimuli. Medians of scrambled versions had a slight positive offset (+3). Mean luminance and contrast differed between stimuli depicting fake and real expressions despite normalization.

Figure D3

Averaged power spectrum of the spatial frequencies for intact and scrambled faces.



Notes: Scrambling by shuffling chunks of pixels resulted in a different power spectrum and importantly, did not preserve the pronounced low frequencies of the intact facial stimuli.

doi: 10.53846/goediss-9841