Executive Control of Facial Expressions of Emotion

Matthias Beringer, M.Sc.

Dissertation zur Erlangung des Doktorgrades, Dr. rer. nat. an der Universität Hamburg

Fakultät für Psychologie und Bewegungswissenschaft, Institut für Psychologie

Hamburg 2024

Erstgutachter: Prof. Dr. Jan Wacker, Universität Hamburg

Zweitgutachterin: Prof. Dr. Sabina Pauen, Universität Heidelberg

Tag der Disputation: 30.04.2025

Promotionsprüfungsausschuss:

Vorsitz	Prof. Dr. phil. Alexander Redlich
1. Dissertationsgutachten	Prof. Dr. rer. nat. Jan Wacker
2. Dissertationsgutachten	Prof. Dr. phil. nat. Sabina Pauen
1. Disputationsgutachten	Prof. Dr. phil. Juliane Degner-Premraj
2. Disputationsgutachten	Prof. Dr. rer. nat. Tania Lincoln

Table of Contents

Abstract

Automated assessment with computer software offers new, efficient opportunities for measurement of facial expressions of emotion. However, little is known about reliability and validity of these assessment tools.

In **Publication 1**, we investigated the quality of a machine vision algorithm (FACET, iMotions, 2016) using standardized databases of dynamic facial expressions under various conditions (angle, distance, lighting, and resolution). We found high reliability in ratings concordance for facial expressions and went on examining the convergent validity of automated assessment and electromyography (EMG) by measuring reaction times (RTs) during the production of joy and anger expressions in a response priming task. Both EMG and automated assessment data showed similar performance costs in RTs when inhibiting an incorrectly prepared expression and reprogramming the correct one. These results support the use of automated assessment for evaluating experimental effects in facial expressions.

In **Publication 2**, we combined electroencephalography (EEG) and automated facial expression assessment, for the first time. We started examining facial expressions of joy, fear, and disgust in response to different visual stimuli using a go/no-go task. Then we went on focusing on expressions of joy and disgust influenced by gaze direction (with and without eye contact) in a more natural setting with a real person as the stimulus. Analysis of RTs, errors, and an event related potential (ERP) analysis of the no-go P3 suggest that facial expressions modulated by mimicry, emotional reactions, and push factors require greater top-down control, especially for expressions of joy compared to fear and disgust.

In **Publication 3**, we examine the demands of a social situation using a Stroop-like task. Participants took part in a simulated online dating scenario to study possible moderation effects of perceived attractiveness on facial expressions. RTs showed facilitation for deliberate expressions of joy and disgust with congruent pictures, but interference from incongruent expressions only affected joy. Accuracy decreased when participants expressed disgust at a smiling, attractive stimulus, without affecting RTs. ERP data revealed an early posterior negativity (EPN) for better sensory processing of joy over neutral expressions and a more negative N2 indicating conflict detection from mismatched expressions. Our findings demonstrate automatic tendencies to imitate facial expressions of joy and disgust.

Across all experiments, we successfully utilized and demonstrated the benefits of an automated assessment technique for investigating the executive control of facial expressions. However, the lack of process theory of emotion and the experimental study designs, with their numerous trial repetitions, suggest a lack of construct validity, raising doubts about the involvement of emotions in the observed effects.

Keywords: Facial action coding system, Automated facial expression recognition, Reliability, Validity, Executive functions, Top-down control, Go/no-go task, Emotional expression interference, Facial mimicry

Executive Control of Facial Expressions of Emotion

1 Introduction

The ability to monitor and control one's facial expressions of emotion is fundamental for successful social interaction. This includes the recognition of emotion as a fundamental aspect of social intelligence in interactions between people (e.g., Mayer, Roberts, & Barsade, 2008; Scherer, 2009), which has been studied intensively (e.g., Adolphs, 2002; Dimberg, Thunberg, & Elmehed, 2000; Elfenbein & Ambady, 2002; Tottenham et al., 2009). Complex combinations of situational, cultural, and individual factors can be decoded by humans to understand each other or to modulate facial expressions according to actual demands. For example, a smile during a conversation can be a hint for approach and willingness to cooperate (e.g., Frank, 1988), while an anxious face shows others that someone feels uncomfortable and requires help. Facial expressions of emotion reflect an adaptation through evolution (e.g., Darwin, 1998), and complex affect programs of basic emotions help to convey important information between individuals within seconds (e.g., Ekman et al., 1987; Van Kleef, 2009). Much about our own emotional state is conveyed by facial expressions (e.g., Ekman & Friesen, 1982), changes in facial expressions (e.g., Niedenthal, Halberstadt, Margolin, & Innes-Ker, 2000), and influencing factors such as gaze direction (e.g., Bayliss, Frischen, Fenske, & Tipper, 2007).

1.1 Facial Expressions and Executive Functions

Expressions are strategically planned, inhibited, faked, or masked, such as when trying to obtain a favor. Masking rules can be strongly influenced by cultural norms; for example, in some socio-cultural groups, men are encouraged not to cry in public (e.g., Vogel, Heimerdinger-Edwards, Hammer, & Hubbard, 2011). During conversations, facial expressions change rapidly to show approval or disapproval, or intentions of approach or avoidance, driven by our own plans and reactions to the emotional expressions of the interlocutor. Thus, predispositions to show or not to show certain expressions need to be systematically monitored, inhibited, and readjusted. This cognitive management of facial expressions can be seen as a form of executive control and is an emerging field of psychological research.

Hence, executive control over facial expressions of emotion might differ from control over other movements, as there are automatic tendencies to emotionally relevant stimuli, like facial mimicry (e.g., Korb, Grandjean, & Scherer, 2010), social norms (e.g., when to smile; Schmidt, Cohn, & Tian, 2003), and motivational factors like approach (e.g., Frank, 1988). Therefore, one cannot necessarily extrapolate findings from simple hand, finger, or foot movements as responses (e.g., Masaki & Sommer, 2012).

The present studies aim to use new methods of automated assessment of facial expressions to investigate executive functions, focusing on the inhibition of facial movements. In the past, inhibition has been investigated with standard cognitive research paradigms such as the response-priming task, the go/no-go task, and the Stroop task (e.g., Friedman et al., 2008). In the response-priming task, participants prepare a motor response and execute it a few seconds later when a response signal appears (e.g., Rosenbaum, 1980). In some trials, the response signal calls for a different action than the primed one (invalid condition). Previous research has shown costs in reaction times (RTs) for invalidly primed responses, reflecting the demands of inhibiting the planned but inappropriate response and switching to the correct one (e.g., Recio, Shmuilovich, & Sommer, 2014).

The go/no-go task uses the number of commission errors in no-go trials as an indicator of the inhibition of preponderant motor responses (e.g., Korb et al., 2010). The Stroop task requires the inhibition of an automatic perceptual bias in one attribute of the stimulus, classically ignoring the word meaning and naming the color of the ink in which the word is printed (Stroop, 1935). Longer RTs due to greater interference occur when the irrelevant attribute of the stimuli (e.g., word meaning) is processed faster (and hence more automatically) than the relevant attribute (e.g., color), and the unwanted response is therefore available first.

Facial muscle responses have often been studied with electromyography (EMG). EMG studies have identified precise markers for different expressions (e.g., Frank, Ekman, & Friesen, 1993), distinctive features of spontaneous and posed smiles (e.g., Hess, Kappas, McHugo, Kleck, & Lanzetta, 1989), detected automatic mimic reactions to emotional stimuli (e.g., Dimberg, 1982), estimated the impact of social variables on expression (e.g., Fridlund, 2014), and studied the ability to voluntarily suppress facial expressions while viewing emotional stimuli (e.g., Kappas, Bherer, & Thériault, 2000). There is limited EMG literature on the investigation of emotion-related facial expressions with experimental paradigms commonly employed to study motor control (e.g., Korb et al., 2010; Recio et al., 2014). Nevertheless, EMG, as a well-established method for investigating facial movements, is the method of choice to establish the validity of automated assessment tools and ensure these software tools are usable for subsequent research on expressions of emotion.

1.2 Costs and Benefits of Automated Assessment of Facial Expressions

Automated assessment with software offers significant advantages over EMG for experimental research. EMG requires anatomical knowledge to address technical challenges, such as accurately locating specific muscles like the zygomaticus major and corrugator supercilii. Furthermore, it is often unclear whether recorded muscle activities reflect the target muscle or neighboring muscles (Wolf, 2015), and dropout rates can be high (Recio et al., 2014). In contrast, facial movements can be easily recorded with a webcam and analyzed using automated assessment software.

Additionally, participants in EMG studies may recognize the target facial expressions due to electrode placement, potentially leading to increased salience of emotional expressions and affecting the study's aim.

While EMG captures electrical potentials in muscles, including those not visible on the facial surface (e.g., muscle tone), this method may be less sensitive compared to video analysis. For instance, the activation of the zygomaticus major can occur during both chewing and smiling (AU12, lip corner puller, see FACS; Ekman & Friesen, 1978). Video-based analyses, using vector machines trained on diverse facial expressions, may provide a less intrusive and more ecologically valid measure of emotional reactions compared to EMG. In real social interactions, observers likely respond more to visible facial expressions than to subtle changes in muscle tone. However, researchers interested in "micro-expressions" (e.g., Ekman, 2009; Ekman & Friesen, 1969; Matsumoto & Hwang, 2014; Yan et al., 2013) might prefer EMG for its higher temporal resolution and ability to indicate the intensity of facial expressions.

Automated assessment offers greater flexibility than EMG by simultaneously analyzing multiple facial expressions (e.g., different action units and emotions) and enabling double-blind studies. Compared to manual coding, automated software can process large datasets within hours, across many subjects simultaneously. Importantly, software coding is objective, whereas inter-rater reliability can vary, especially with inexperienced or fatigued human coders.

Human categorization of facial expressions typically relies on perceptual and affective processing (e.g., Calvo & Nummenmaa, 2016), while automated assessment depends solely on perceptual matching mechanisms (e.g., Calvo, Avero, Fernández-Martín, & Recio, 2016). Despite this, automated systems show comparable frequencies of prototypical errors (e.g., Cottrell & Hsiao, 2011; Susskind et al., 2007) and can achieve perfect accuracy in categorization under optimal conditions (e.g., 100% for joy; Dailey, Cottrell, Padgett & Adolphs., 2002).

Current literature on the experimental use of automated facial expression assessment is limited, with even fewer studies addressing the reliability and validity of these software tools (e.g., Dente, Küster, Skora & Krumhuber, 2017). Deriso et al. (2012) explored the perception-production link of facial expressions using real-time recognition and feedback with the Computer Expression Recognition Toolbox (CERT; Littlewort et al., 2011). Their findings linked visual-motor associations with perceptual abilities. Susskind et al. (2008) supported the Darwinian hypothesis that facial expressions evolved to alter sensory interactions with the physical world. They utilized active appearance models that match image variations with training set parameters (Cootes, Edwards, & Taylor, 2001).

Other studies have examined affect detection and emotion classification in classroom environments (e.g., Bosch et al., 2015), toddler behavior with social robots (e.g., Malmir et al., 2013), and student engagement through facial expressions of boredom versus engagement (e.g., Whitehill et al., 2014). The field of automated facial expression assessment is expanding, with an anticipated increase in research applications in the near future.

1.3 Neural Underpinnings

Neurophysiological studies have elucidated the foundations of motor planning and control, yet few have specifically investigated facial expressions. Lesion studies suggest distinct motor control systems for voluntary versus spontaneous emotional expressions (Rinn, 1991). Subcortical motor systems are believed to govern automatic, stereotyped facial movements (e.g., sneezing), while the motor cortex is responsible for voluntary facial control. The motor cortex, specialized in motor representation and execution, is hierarchically organized. Frontal and cingulate cortex areas are particularly involved in controlling facial movements and in the interplay between emotional expressions, attention, and cognition (Morecraft, Stilwell–Morecraft, & Rossing, 2004).

The prefrontal cortex (PFC) connects extensively with sensory cortices and motor areas, acting as a bridge between information processing and the selection and execution of appropriate responses. Emotional inputs consistently activate inhibitory control mechanisms in the right PFC and anterior cingulate cortex (ACC; Hooker & Knight, 2006). The PFC also regulates the inhibition of predominant responses. Neuroimaging studies using the go/no-go and Stroop tasks have repeatedly identified involvement of subcortical areas (e.g., basal ganglia) and cortical regions (e.g., right inferior frontal gyrus, ACC, anterior supplementary motor area, preSMA) in motor inhibition (Swick, Ashley, & Turken, 2011). Data suggest that the PFC plays a critical role in the strategic control and effective use of facial expressions in social contexts (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012).

Event-related potentials (ERP) studies using electroencephalography (EEG) have provided valuable insights into the timing of motor response preparation and inhibition. The N200 and P300 ERP components, with frontocentral distribution, are associated with the inhibition of emotion-related facial movements (Recio et al., 2014). Additionally, the early posterior negativity (EPN) component, typically elicited by emotional content 200 to 300 ms post-stimulus at occipital sites, is interpreted as reflecting enhanced sensory processing and reflexive attention to emotional stimuli (Hajcak, Weinberg, MacNamara, & Foti, 2011). Integrating automated assessment of facial expressions, EMG, and EEG allows for a comprehensive investigation of the executive functions underlying emotional facial expressions at various levels of abstraction.

2 The Present Research

Objective 1: Examining the behavioral correlates of executive control over facial expressions. While previous studies have used EMG to assess motor control of facial expressions (e.g., Korb et al., 2010; Recio et al., 2014), in pursuit if this first objective we introduce a novel method for scoring facial expressions. Despite the growing adoption of computer-based facial expression analysis (Cohn & De la Torre, 2015), its recent development raises questions about its efficacy. Specifically, can this method replicate typical experimental effects observed in motor control tasks with other modalities (e.g., finger movements) and methods (e.g., EMG) when applied to facial expressions? Furthermore, do these scores provide reliable and valid measurements of facial expressions? Publication 1 primarily addresses these questions, though our further research will refine and deepen these inquiries.

Objective 2: Investigating the interactions between emotion and executive control over facial expressions. Emotion and cognitive control are interconnected, with different emotional states influencing cognitive control functions and resolving cognitive conflicts by prioritizing specific abilities (Gray, 2004). For example, negative affect might enhance attention and processing speed but impair the suppression of unwanted threat signals (Pessoa et al., 2012). Conversely, positive stimuli might be easier to inhibit but could lead to reduced cognitive control and more commission errors. This objective explores whether executive control varies between facial movements associated with positive, neutral, and negative expressions and how emotional stimuli affect the control of facial expressions. All publications address this question iteratively, with each study offering deeper insights into emotional qualities. We balanced the design to ensure methodological comparability, contrasting positive emotions like joy with various negative emotional qualities.

Objective 3: Examining the automaticity of facial expressions. Basic emotion theories propose that facial expressions arise from evolved, automatic affective programs (Ekman, 1972). In contrast, social theories suggest a greater degree of control over facial expressions. This objective aims to determine whether facial expressions are primarily automatic responses driven by internal factors or if they involve more controlled movements. Publication 3 explores this issue.

Objective 4: Evaluating how social context variables influence the control of facial expressions. Specifically, this objective investigates whether social motives can enhance or hinder facial expression control. Publication 2 explores whether an affiliative motive facilitates the production of smiles and prototypical expressions of disgust or interferes with their control. Publications 2 and 3 address this question.

To address these objectives comprehensively, the following sections present the relevant publications, each contributing to a nuanced understanding of facial expression control and its implications. These publications include multiple studies that examine various facets of the research questions outlined above, providing a detailed view of the methodologies, findings, and their significance within the broader context of this research.

2.1 Publication 1: Beringer, M., Spohn, F., Hildebrandt, A., Wacker, J., & Recio, G. (2019). Reliability and validity of machine vision for the assessment of facial expressions. *Cognitive Systems Research, 56*, 119-132.

I will now present the first publication of our research project. This study serves as an initial evaluation of a new measurement technique, focusing on classical quality criteria in psychology as outlined by classical test theory. Given that these criteria are fundamental for scientific work, it is logical to begin with an examination of this topic.

The publication is divided into two main parts. The first part provides a technical evaluation using simulated data to assess the quality of the automated facial expression assessment tool. The second part connects this evaluation to a commonly used experimental paradigm in emotional facial expression research, with which our research group has prior experience (e.g., response priming task, see Recio et al., 2014), though it is not a direct replication.

The objective of this publication is to evaluate the objectivity, reliability, and validity of the automated assessment tool for facial expressions of emotion and to gain initial insights into its applicability in experimental settings. Ultimately, this assessment aims to determine whether this measurement technique is suitable for use in scientific research.

Reliability and validity of machine vision for the assessment of facial expressions

Matthias Beringer[a], Frank Spohn[a], Andrea Hildebrandt[b], Jan Wacker[a] and Guillermo Recio[a]

[a] Universität Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany

matthias.beringer@uni-hamburg.de, frank.spohn@uni-hamburg.de

jan.wacker@uni-hamburg.de, guillermo.recio@uni-hamburg.de

[b] Carl von Ossietzky Universität Oldenburg, Ammerländer Heerstr. 114-118, D-26129 Oldenburg, Germany, andrea.hildebrandt@uni-oldenburg.de

Corresponding author:
Matthias Beringer
Dillstraße 1
D-20146 Hamburg, Germany
m.q.beringer@gmail.com
Tel: +49 (0)176 30311677

Abstract

Automated assessment of facial expressions with machine vision software opens up new opportunities for the assessment of facial expression in a shrewd and economic way in psychological and applied research. We investigated the assessment quality of one machine vision algorithm (FACET) in a study using standardized databases of dynamic facial expressions in different conditions (angle, distance, lighting and resolution). We found high reliability in terms of ratings concordance across conditions for facial expressions (intraclass correlation, ICC = .96) and action units (ICC = .78). Signal detection analyses showed good classification for both facial expressions (area under the curve, AUC > .99) and action unit scores (AUC = .91). In a second study, we investigated the convergent validity of machine vision assessment and electromyography (EMG) with regard to reaction times measured during the production of smiles (action unit 12) and frowns (action unit 4). To this end, we simultaneously measured EMG and expression classification with machine vision software in a response priming task with validly and invalidly primed responses. Both, EMG and machine vision data revealed similar performance costs in reaction times of inhibiting the falsely prepared expression and reprogramming the correct one. These results support machine vision as a suitable tool for assessing experimental effects in facial reaction times.

Keywords: Reliability, Validity, FACET, Facial expression, Facial action coding system.

1. Introduction

Successful social interactions involve the decoding and integration of complex combinations of situational, cultural and person specific information. Basic emotion theory (e.g., Ekman & Oster, 1979) interprets facial expressions of emotion to reflect an adaptation through evolution (Darwin, 1998), whereas complex affect programs of basic emotions help to convey important information between individuals (Ekman et al., 1987; Van Kleef, 2009), and can serve social predictive functions. For example, a smile during a conversation can be a hint for approach and willingness to cooperate (Frank, 1988), or an anxious face shows others how someone may feel in an uncomfortable situation and requires help. According to this approach, facial expressions convey information about emotional states (Ekman & Friesen, 1982) and subtle changes in them (Niedenthal, Halberstadt, Margolin, & Innes-Ker, 2000). Emotion recognition is a fundamental aspect of socio-emotional intelligence (Blickle, Momm, Liu, Witzki, & Steinmayr, 2011; Schlegel, Fontaine, & Scherer, 2017), and has been studied extensively (Tottenham et al., 2009).

Other theoretical frameworks of emotion describe facial expressions in a dimensional space (e.g., Russell & Mehrabian, 1977) rather than in discrete categories. There is also a debate in the psychological literature regarding the universality of facial expressions across cultures. Some authors argue that prototypical emotional expressions are rarely seen in everyday life, as spontaneous expressions are often more ambiguous than posed expressions (Motley & Camden, 1988). Here, we focus on facial movements that can be related to emotions (e.g., smiling for joy and frowning for anger) in some situations, but we do not investigate the underlying emotional states.

Traditionally, research measuring prototypical facial expressions of up to eight categorical basic emotions (Cohn & Ekman, 2005) used electromyography (EMG; Fridlund, 2014) and human ratings relying on the facial action coding system (FACS; Ekman & Friesen, 1978). FACS is a coding system used to qualify and quantify the activation of a single facial muscle or a combination of several muscles (action units, AUs). The authors define AUs as the smallest visible functional facial movements that humans can observe, such as pulling the corners of the lips to smile (AU12), or frowning with the brows (AU4) to show an angry face. The reliability and validity of manual coding scores with FACS is well documented (Cohn, Ambadar, & Ekman, 2007). It has been argued that configuration of several AUs can then be used to describe specific facial expressions (e.g., joy = activation in [Orbicularis oculi + Pars orbitalis, “cheek raiser”, AU6] + activation in [Zygomaticus major, “lip corner puller”, AU12]) (Friesen & Ekman, 1983). However, the coding procedure is very time-consuming and requires considerable resources in personnel, as usually two coders need an hour for 1 minute of videotaped expressions (Ekman & Oster, 1979).

1.1 Costs and benefits of automated assessment of facial expressions

In recent years a number of machine vision software solutions have appeared as an alternative to manual coding (Baltrušaitis, Robinson, & Morency, 2016; Littlewort et al., 2011; Shafiq, Tauseef, Fahiem, & Farhan, 2017). Although still an active area of research, some studies have found that software coding can be as precise and reliable as assessment done by human raters (Terzis, Moridis, & Economides, 2010).

Automated assessment might also overcome some problems associated with EMG. Quite often, it is unclear whether the recorded EMG activities actually reflect the target muscle or another neighboring muscle (Wolf, 2015), and dropout rates are relatively high in some studies (Recio, Shmuilovich, & Sommer, 2014). Facial movements can be easily recorded using a webcam and analyzed with machine vision. Hence, automated assessment of facial expressions with machine vision software could reduce preparation time relative to EMG. Additionally, electrodes attached to the skin during EMG measurements could direct attention to specific muscles or facial expressions in particular (Cacioppo, Petty, & Marshall-Goodell, 1984; Fridlund & Izard, 1983). On the one hand, EMG captures the measurable electrical potentials of muscle movements not directly observable for the human eye (e.g., muscle tone), presumably making this method more sensitive to weak facial expressions than machine vision. On the other hand, specificity of EMG might be lower than analyses based on video recordings. For example, activation of zygomaticus major can be observed both while chewing and smiling ("lip corner puller", AU12), which might complicate the distinction between these two facial movements from EMG data.

Software analyses are clearly faster than manual coding (e.g., Ekman & Oster, 1979; Littlewort et al., 2011), even though they strongly depend on computer performance capacity. Besides theoretical assumptions made during the training of the software, the automated assessment is more objective, as inter-rater reliability is usually variable for human coders. Humans typically rely on both perceptual and affective processes to categorize facial expressions of emotion in everyday life situations or in laboratory settings (Calvo & Nummenmaa, 2016), whereas automated assessment rests upon perceptual matching mechanisms only (Calvo, Avero, Fernández-Martín, & Recio, 2016).

1.2 The data output of automated assessment

In the present study, we used the software FACET (version 6.1.2667.3, iMotions, 2016) to investigate the automated assessment of facial expressions. FACET is built upon another software called CERT (Littlewort et al., 2011). Probabilistic results are provided as evidence scores for discrete facial expressions of emotion and AUs. An evidence score for a facial expression represents the estimated odds, in decimal logarithmic scale, of that particular expression being present in a given face image. Evidence scores can be transformed to probability (*P*) using the following formula: $P = 1/(1+10^{-\text{evidence score}})$. For example, an evidence score of zero for "joy" (as named in FACET) would indicate, that it is equally likely, that the targeted face either shows or does not show a facial expression of joy. An evidence score of 0.5 represents an estimated probability of 76% for the presence of a facial expression of joy, an evidence score of 1 indicates an estimated probability of 91%, and an evidence score > 2 represents an estimated probability of approximately 100%. Thus, when the face recorded in the input video shows a prototypical facial expression of joy, the evidence scores for "joy" in the output increase to values above zero while it decreases in all other channels (see Figure A in the supplemental material).

Some research questions require a classification indicating the presence or absence of one or several target expressions, however, the quality of the video material may sometimes be compromised. Study 1 investigates recognition performance of machine vision using video stimuli of different quality to assess FACET's performance under controlled conditions, using standardized databases validated with human ratings (e.g., van der Schalk, Hawk, Fischer, & Doosje, 2011). Other research questions commonly need to assess facial reaction times (RTs). Study 2 investigates the use of machine vision software to score the onset of two facial expressions as RTs, to assess whether FACET is suitable to measure experimental effects, i.e., differences in RT between two conditions.

The evidence scores can be dichotomized (i.e., into presence or absence of the facial expression) using a cut-off criterion or threshold, which reflects the onset of the facial expressions (Recio & Sommer, 2018). More generally, from the perspective of signal detection analysis, a threshold provides a value at which it is more appropriate to assume the presence of a signal than of noise. Given an emerging signal of a prototypical expression and little noise, the evidence score in a target expression increases, while evidence score in all counter expressions would decrease. Nevertheless, it would be difficult to define an acceptable value for a facial RT because expressions increase continuously over time until reaching the apex. Moreover, evidence scores of all channels of facial expressions of emotion are not bound to each other, and in empirical data it is possible and plausible that 2 or more expressions emerge at the same time (e.g., for less prototypical, mixed expressions). Thus, a criterion is needed for every channel to define whether evidence score of a target expression is strong enough for unequivocally classifying a hit. This question will be further elaborated in Study 2.

1.3 The present study

Up to now, there are only scarce studies using automated assessment of facial expressions for experimental research (Cohn & De la Torre, 2014) and even less research investigating the reliability and validity of available software tools (Dente, Küster, Skora, & Krumhuber, 2017; Stöckli, Schulte-Mecklenbeck, Borer, & Samson, 2017). Deriso et al. (2012) investigated the perception-production link of facial expressions using real time recognition and automated feedback using CERT (Littlewort et al., 2011) in a dynamic expression recognition task. Susskind et al. (2008) found support for the Darwinian hypothesis that facial expressions are not arbitrary configurations but may have originated in altering the sensory interface with the physical world. For example, the expression of fear enlarges the visual field, accelerates eye movements and increases in nasal volume while disgust expressions follow the opposite pattern. Susskind and colleagues (2008) applied active appearance models, based on an algorithm that matches modes of shape and gray-level variations of the incoming images with parameters of a training set (Cootes, Edwards, & Taylor, 2001). Girard, Cohn, Jeni, Syette, and De la Torre (2015) compared automated and manual coding with FACS and showed high reliability for the proportion of time that each AU occurred, and moderate to high reliability for frame-by-frame analyses. Others investigated affect detection and accuracy of facial expression classification in classroom environments (Bosch et al., 2015), toddler behavior in interaction with social robots (Malmir, Forster, Youngstrom, Morrison, & Movellan, 2013) and student engagement by typical facial expressions for boredom vs. engagement (Whitehill, Serpell, Lin, Foster, & Movellan, 2014).

We observe an expanding field of research questions that can be addressed with automated assessment of facial expressions, however, the reliability and validity of the scores is still unclear. Here we focus on three fundamental issues related to the measurement of variables and constructs in psychological research. In Study 1, we (1) assess the reliability of scores of facial expressions provided by automated assessment under well-controlled conditions, and (2) examine potential influence of light, angle and resolution conditions on measurement accuracy. We used stimuli from standardized databases of dynamic facial expressions of both basic emotions and AUs. We expected high agreements between classifications provided by FACET and the intended classifications of the stimuli obtained from human raters in the samples used for the standardization of the databases. Next, we digitally manipulated the original stimuli from the databases to compromise optimal conditions of lighting, resolution, distance and head position. We expected high correlations of facial expressions of emotion and AU evidence scores for the manipulated stimuli across all conditions.

In Study 2, we (3) address the crucial issue of the validity of the classifications provided by automated assessment. Here, we tested convergence between EMG measurements and FACET in a replication of the experimental effects captured by facial RTs during a response priming task (Recio et al., 2014). We choose EMG to maximize comparison with this study and because it is a reliable method widely used in research on facial expressions (Hess et al., 2017). We expected similar experimental effects in RTs measured with EMG and FACET, and high correlations between both measures.

2. Study 1

2.1 Procedure and Stimuli

We used dynamic facial expressions of seven posed facial expressions of basic emotions (joy, anger, surprise, fear, contempt, disgust, and sadness) and the neutral faces of 21 young adults (9 female, 52% Caucasian, 48% Moroccan) trained by FACS experts and portrayed in frontal view from the standardized Amsterdam Dynamic Facial Expression Set (ADFES; van der Schalk, Hawk, Fischer, & Doosje, 2011). Each dynamic stimulus starts with a neutral face, changes into a prototypical facial expression of emotion, and returns at the end to a neutral state.

We additionally used a database with face videos showing posed activation in 19 AUs provided by the Max Planck Institute for Biological Cybernetics in Tuebingen, Germany (MPI, Troje & Bulthoff, 1996) to estimate FACET classification performance for AUs. The MPI database captures one professional AU performer, repeating each AU three times, starting with a neutral face evolving to maximum activation. We excluded AU1 because it has been only captured twice, instead of three repetitions. The Videolab technology used to capture this database provided synchronized recordings of facial movements from six different viewpoints at the same time (Kleiner, Wallraven, & Bülthoff, 2004). From these viewing angles we used two angles (+9°, -9°) within the range provided by FACET's developers for an optimal functioning (i.e., +15°,-15°), and two angles (+27°, -27°) outside this range, to challenge FACET performance and test possible impact on the measurement of the remaining 18 AUs. However, the outer angles provided by the MPI database (+45°, -45°) are overly outside the suggested range, so that the face detector does not identify the face and therefore FACET does not provide any analyses.

We then manipulated the original video material of both databases using FFmpeg (version 3.1.3, FFmpeg Developers, 2016), and created four additional variations of the facial stimuli according to the factors: *horizontal angle*, *distance*, *lighting* and *resolution*. Figure 1 shows all factors and their levels, including *vertical angle*, which was already provided in the MPI database, but not in the ADFES. The final set consisted of 64 stimuli for each of the 19 different AUs from the MPI database, each performed 3 times, resulting in 3648 trials in total. In case of ADFES, there were 17 trials for each of 21 subjects showing seven different facial expressions of emotion (2499 trials in total). All videos were analyzed with FACET (version 6.1.2667.3, iMotions, 2016) using post hoc processing after applying a manual baseline correction, defined as the neutral face in case of the ADFES stimuli. No such baseline correction procedure is available for AUs.

Figure 1: Design Study 1

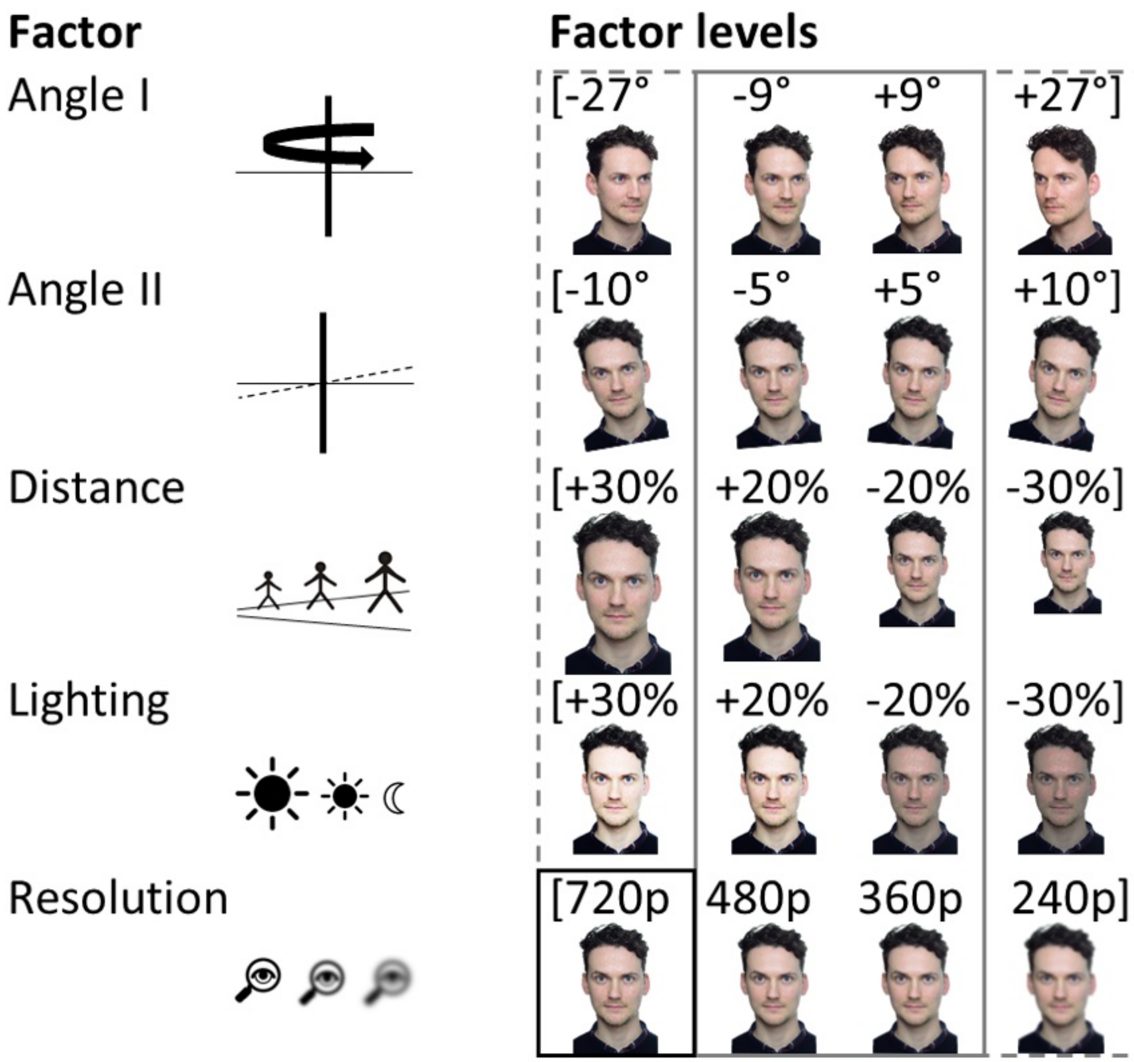


Figure 1. Example of the different levels for the factors manipulated in Study 1. Dark solid line (**–**) indicates the original stimuli. Grey solid line (**–**) indicates factor levels that are within recommended range provided in product information of FACET. Dashed line (- -) indicates factor levels that are outside acceptable deviations considering product information of FACET. The factor Angle I shows 4 different camera viewpoints of MPI and was not available for the ADFES. All other factor levels were prepared for ADFES and MPI in exactly the same way. Original posers of the ADFES and MPI databases were replaced with an example in this figure to respect copyrights.

2.2 Data Analyses

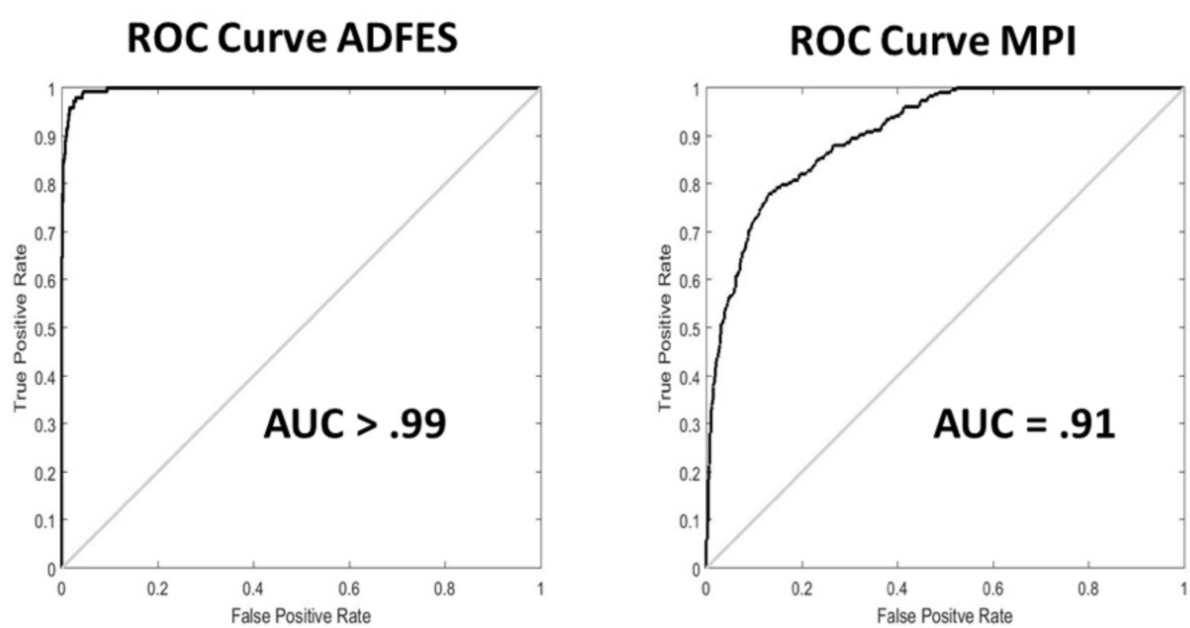
The FACET output data were preprocessed and analyzed using MATLAB (R2016a, The MathWorks, 2016). To compare each original expression with its digitally manipulated duplicates following the factor levels mentioned above, we computed the maxima for every expression and AU channel of the original trials. Since all stimuli were exactly of the same length, timestamps of those maxima were used to identify corresponding evidence scores of all channels in the manipulated conditions. We excluded the trials of one condition provided by the MPI database (vertical angle -27° $\wedge$ horizontal angle -10°) because the face detector of FACET would not work for this particular condition. It is important to point out, that both factor levels were outside the recommendations provided in the user manual. Analyses were calculated for the remaining 3591 trials of the MPI and the 2499 trials of the ADFES stimuli.

As a starting point, we compared calibrated databases of facial expressions (ADFES and MPI) with FACET and first conducted a signal detection analysis with the calculated maxima of the original stimuli to evaluate the classification performance of FACET (e.g., coding joy in a joy trial, and AU2 in an AU2 trial). To evaluate classification accuracy, we plotted receiver operating characteristic curves (ROC) and computed the area under the curve (AUC) to provide a single value indicating the classification performance of FACET. To compute the ROC curves, we varied thresholds of facial expressions of emotion channels between -18 and +18 with an iteration of 0.001 and a smaller interval for thresholds of AUs between -8 and +8 with an iteration of 0.001, because evidence scores of ADFES maxima had a wider range than MPI maxima. AUC were calculated using the trapezoidal rule between two data points. Reliability of evidence scores was estimated as intraclass correlation (ICC) of type C,1, indicating consistency agreement between factors and between factor levels across trials (McGraw & Wong, 1996). As sampling errors and repeated measurement effects within participants should be reduced to a minimum by the use of an automated assessment tool, we used the manipulated factors to simulate expressions in different conditions, as they could occur in some laboratory settings. We sum up results using a mean ICC (see table 1, first row), because we expected reliability to be high, being easily expressed with one index, especially after excluding the resolution factor (see table 1, last row). We are happy to share our data and code of Study 1 for data analyses upon request[1].

2.3 Results

Signal detection analyses. Overall, ROC curves for the original trials of ADFES and MPI stimuli revealed very high classification accuracy for both scores for facial expressions of emotion (AUC > .99, see Figure 2) and AUs (AUC = .91).

Figure 2: Receiver operating characteristic curves of Study 1



[1] For requests please contact m.q.beringer@gmail.com.

Figure 2. Left side: Area under the curve (AUC) over all facial expressions of emotion channels and all original trials of ADFES database. Right side: AUC over all AUs of the MPI database. ROC curves for each facial expression and each AU can be found in Figure B in the supplemental material.

Reliability. Overall, ICCs (type C, 1) showed high agreement for facial expressions of emotion: ICC = .85, and a moderate agreement for AUs: ICC = .71. Of the manipulated factors, resolution impacted reliability the most. Excluding the resolution factor, ICCs increased for both, facial expressions ICC = .96 and AUs ICC = .98. Excluding other manipulation factors did not affect reliability (see Table 1).

Table 1

ICCs Study 1

	FACET	
	ADFES**	MPI**
Over all conditions*	.85	.71
4 conditions - following condition excluded:		
.Angle horizontal	.82	.65
.Distance	.82	.65
.Brightness	.82	.65
.Resolution	.96	.98

Note: * To exploit variance given by the maximum of available data points ICCs were computed comparing trials over all conditions (3591 face stimuli from the MPI database and 2499 of the ADFES; see *Over all conditions*). To explore the influence of the individual factors, each factor was excluded before computing ICCs (denoted by "." before the respective condition name, e.g. *.Resolution* means that ICC was computed over all conditions without the resolution factor).
** ICC (C,1), Single score intraclass correlation coefficient for two way random effects models giving the consistency among measurements (McGraw & Wong, 1996).
*** The factor angle vertical could not be excluded for the MPI database, because there was no frontal camera. ICCs represent mean of both middle cameras (-9° and +9). There was no manipu-

lation of vertical angle for the ADFES database, as it is not possible to manipulate this factor digitally because information from the sides of the face of the models is missing in the original stimuli of the MPI database, as they were recorded with a single frontal camera.

2.4 Discussion

Signal detection analyses of the original stimuli from the ADFES and MPI face databases show high classification accuracy of the analyses with FACET for facial expressions of emotion and AUs. We calculated ICCs (C,1) to complement empirical evidence on reliability based on the data from Study 1, and results indicate satisfactory performance of FACET taking all different factors into account (angle vertical, angle horizontal, distance, lighting and resolution). Classification agreement across conditions was high for all facial expressions and AUs, after excluding the factor resolution, which is in line with findings observed for spontaneous expressions (e.g., see Girard, Cohn, Jeni, Sayette, and De la Torre, 2015).

Both angle and distance comprise naturally occurring variations (e.g., head movements), which are difficult to entirely control in experimental settings, however, the classification accuracy of FACET shows resilience to these factors when they vary within the range studied here. Possible limitations of this approach will the discussed in the general discussion.

3. Study 2

3.1 Participants

An a priori G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) analysis revealed a total sample size of seven participants for a repeated measurement, within factors design, $\eta_p^2 = 0.5$ (Recio et al., 2014), with an alpha error probability of .05, desired power of 0.99, one group, three measurements, a correlation of .70 among repeated measurements and nonsphericity correction of 0.5. With respect to anticipated dropout rates when testing a new measurement technique and with respect to effect size overestimation due to publication bias (Szucs & Ioannidis, 2017), we decided a priori for an analysis sample of 30 participants. All participants gave informed consent and received course credits (41%) or 14 Euro (59%) for their contribution. The study was approved by the Ethics Committee of the Deutsche Gesellschaft für Psychologie.

In order to make sure that participants' facial expressions would be correctly measured without any visual barriers for FACET and to obtain a good signal from the EMG electrodes, we recruited only participants without full beard or glasses, who had less than ±1.00 diopter or were wearing contact lenses. Additionally, participants were all right handed to control for functional asymmetry effects (KelesL, Díyarbakirli, Tan, & Tan, 1997). To retain good data quality we excluded three participants, who had more than 2 *SD* above the mean of false positives over all trials (Leonhart & Lichtenberg, 2009) and four additional participants, because they performed the target expressions with activation of both muscles at the same time and hits for EMG data could not be classified. One further participant was excluded, because he did not understand the task, as he reacted to the prime stimuli instead of the response signal. Eight participants could not be analyzed at all due to a technical issue (i.e., missing stimuli triggers). After an additional recruitment (total sample: 46 healthy adults, 36 women, M_{age} = 24.4 years, SD_{age} = 5.2) the final sample consisted of 30 participants (23 women, M_{age} = 25.2 years, SD_{age} = 5.8).

3.2 Procedure and Stimuli

In Study 2 we used a response priming task similar to Recio et al. (2014). Participants were primed with a "W" or an "M" to mentally prepare to either smile or frown, and hold the facial expression, which could either be valid or invalid with the response signal, i.e., valid "W/M =", or invalid "W/M ≠". For example, if "W" indicates to mentally prepare to smile, the valid response signal asks participants to produce a smile as fast as possible. The invalid response signals ask participants not to show the mentally prepared smile, but the other expression, thus, frowning. Figure 3 displays the temporal sequence of events in a trial. “W” and “M” were randomly assigned between subjects to prepare to smile or to frown. There were four different task conditions: (1) in *joy valid*, participants prepared joy and produced a smile; (2) in *joy invalid*, participants prepared joy but produced a frown; (3) in *anger valid*, participants prepared anger and produced a frown; and (4) in *anger invalid*, participants prepared anger but produced a smile. All stimuli were presented in color black (Arial, front size 80) on a gray background (RGB = 230/230/230).

Figure 3: Trial scheme Study 2

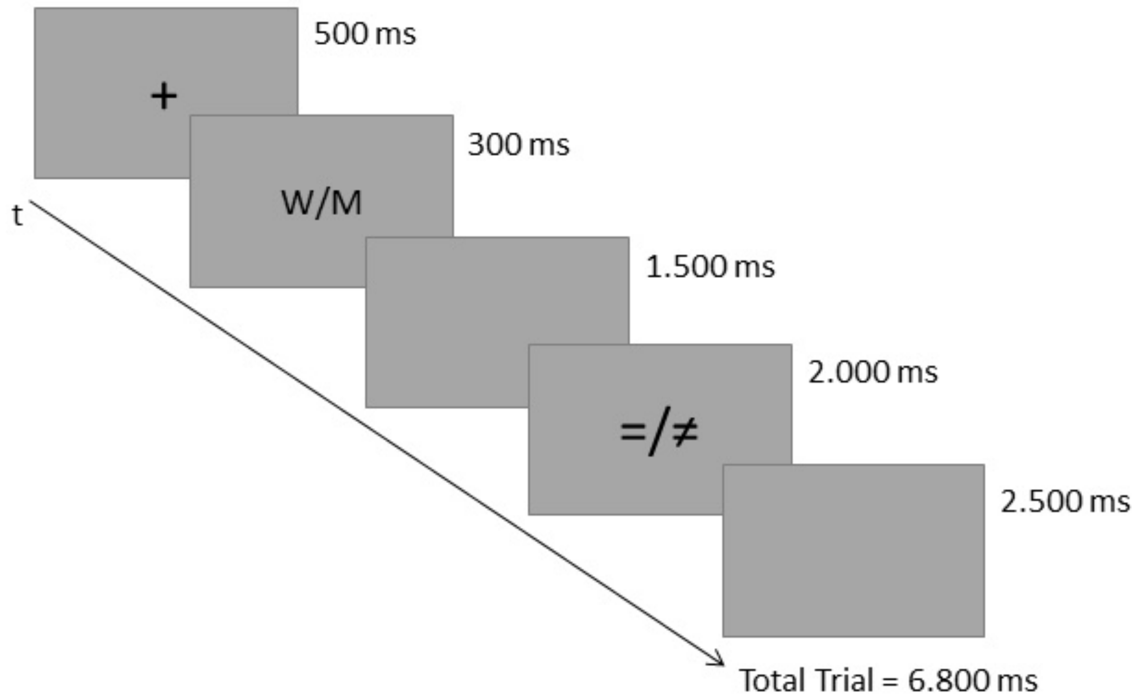


Figure 3. (+) Fixation cross; (W, M) Prime in random order; (=) Response signal with valid prime; (≠) Response signal with invalid prime. Response signals were assigned in random order, with 80% validly primed. Preceding instructions informed participants whether "W" or "M" served as prime for joy or anger (randomly assigned). Instruction during training trials asked participants to show all expressions as intensely as possible and quickly return to a neutral face.

The experiment consisted of 200 trials, half of them instructing frowns, the other half smiles (80% validly primed for each facial expression), with a 5-minutes break in between. There were voluntary breaks (maximally 3 minutes duration) after completing 25% and 75% of the trials. Including preparation for EMG, calibration for FACET and practice trials, the experiment lasted around 40 minutes. Participants performed the task in an electrically shielded room, sitting approximately 80 cm away from a 21-inch LCD display (75 Hz refresh rate). Experimenters (n = 3) were all male and tried to minimize interaction with the participants. All instructions were standardized and presented in written form. Experimenters systematically monitored the correct recording of EMG electrodes and the video for FACET (e.g., attachment of electrodes, impedances, face detection, etc.).

The response priming and response switching tasks were randomly assigned across participants as the first or the second task. Participants also completed the German version of the NEO-PI-R questionnaire (Ostendorf & Angleitner, 2004) for measuring extraversion and neuroticism the day before completing the experiment. Data from the personality questionnaire and the switching task are out of the scope of the current study and will not be discussed.

3.3 Electrophysiological and video recordings of the face

The EMG was recorded from the muscles Zygomaticus major, typically involved in raising mouth corners during smiling (corresponding to AU12), and Corrugator supercilii, typically involved in pressing eyebrows together for frowning (corresponding to AU4). Using 2 Ag/AgCl electrodes for each muscle attached to the skin with adhesive pads on the left side of the face, we followed the guidelines by Fridlund and Cacioppo (1986). The ground electrode was placed on the upper half of the right forehead and impedances were kept below 10 kΩ. Data was sampled at 2048 Hz, DC-1.6 kHz bandwidth using a BioSemi Active Two amplifier. The raw signal was digitized, full-wave rectified, segmented and baseline-corrected using the pre-response signal period with a sliding average window of 3 time frames with the MATLAB Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and sampled at a frequency of 1 kHz.

For video recording we used a Logitech HD Pro Webcam C920 with a sampling rate of 30 fps, and digital zoom function to standardize face size, to optimally fit participants' faces to the face detector area of FACET. Individual baselines were also applied to control for individual differences in emotionality during neutral state due to individual differences in facial morphology, using a 6 seconds interval without any head and face movements before the training phase

(Olderbak, Hildebrandt, Pinkpank, Sommer, & Wilhelm, 2014). Exported data with UTC timestamp of stimuli events was preprocessed with MATLAB.

3.4 Preprocessing of FACET and EMG data

For each subject, the preprocessing of the FACET baseline-corrected and EMG data included four steps. (1) For every trial we segmented the continuous data according to a segment of 2,500 ms beginning at stimulus onset (response signal). Within this step, we rectified evidence scores for sequence effects by computing an additional baseline for target channels, joy and anger, within every segment, using the sliding average of -230 ms to 0 ms right before the fixation cross appeared. Those averages – individual for every segment – were substracted from the following evidence scores during the segment to correct for mood-related variability (e.g., participants might be more excited at the beginning and more annoyed toward the end of the session).

We used an offline filtering for EMG data (19 Hz lowpass butterfly filter of second order) and subtracted values of medial electrodes from values of lateral electrodes. Segmentation was analogue to FACET data. After a full wave rectification EMG data were smoothed by means of a moving average with a window length of 23 ms. Subsequently, each data point in each segment was z-standardized relative to all trials regardless of condition. For every segment an additional baseline was established at 230 ms before stimulus onset.

(2) We calculated Youden indices for FACET data by running signal detection analyses to identify optimal thresholds to define RTs (Youden, 1950). The idea behind this procedure was to use data-driven ROC curves to determine which threshold gives the best proportion between sensitivity and specificity. Optimal cut-off points for the underlying data are given by the AUC subtended by a single operating point or graphically as the height of ROC curve above the chance line. In this way, we obtained additional indicators to those recommended by FACET,

namely a threshold of 1.0 evidence score for intense expressions and a threshold of 0.5 for weak expression across channels of different facial expressions of emotion and AUs. For joy we established the threshold at 2.5 evidence scores, which corresponds to approximately 100% probability that joy rather than another expression is displayed. For anger we established the threshold at 0.5 evidence scores, which indicates a 75% certainty that the categorized expression belongs to anger and not to any other facial expression.

We followed the EMG data processing described in Recio et al. (2014). The threshold of Corrugator supercilii was defined as 25% of the maximum. Because mean activity in Zygomaticus major was overall lower, the 25% threshold produced an unrealistic number of errors in this channel (in some participants even the slightest changes in amplitude would reach the threshold). Therefore, we defined a threshold of 50% of the maximum as a more liberal threshold for the Zygomaticus major.

(3) For both FACET and EMG data, we then defined as *hits* those trials with activation (i.e., evidence scores and amplitude in microvolts, respectively) above the threshold in the target channel (90.85% of the FACET data, 82.37% of the EMG data), as *false positives* those trials with activation above the threshold in the counter channel (1.77% of the FACET data, 1.43% of the EMG data), as *omissions* those trials with activation below the threshold in the target and the counter channel (0.47% of the FACET data, 7.48% of the EMG data), and as *inhibition errors* those trials in which the threshold was beaten in the target channel within 260 ms before the response signal (2.58% of the FACET data, 4.9% of the EMG data). RTs were calculated as the temporal point after stimulus onset when threshold was first exceeded. The analyses of RTs included hits only.

(4) Subsequently, outlier analyses detected and excluded trials with standard deviations of RTs twice above or below mean RT on participant level within condition. Outlier analyses were repeated until no further outliers were detected. No data was corrected more than twice (4.33% of the FACET data, 3.82% of the EMG data).

3.5 Experimental effects and association between FACET and EMG

Preprocessed data were imported to SPSS (Version 23, IBM Corp., 2012) to run repeated measurement analyses of variance (rmANOVA) with factors facial expression (joy, anger) and validity (validly primed, invalidly primed). Post hoc *t*-tests were used to estimate the difference in RTs between valid and invalid conditions for expression joy and anger. All post hoc tests were Bonferroni-corrected for multiple testing.

Finally, RTs of hits calculated from FACET data were related with RTs extracted from EMG data on trial by trial level using Linear Mixed Effects Modeling with the package lme4 (Bates, Mächler, Bolker, & Walker, 2014) in the R environment (R Core Development Team, 2017). In a model series RTs measured with FACET were predicted by RTs measured with EMG. The RTs were z-transformed within participants (standardized) in order to facilitate parameter interpretation. We estimated a random intercept and random slope model, meaning that the relationship between FACET and EMG identified RTs was assumed to vary across persons, using the maximal random effects structure. The average relationship and the variation of this relationship indicate how well the two RT estimations concur.

4. Results

Figure 4 depicts mean FACET evidence scores and EMG amplitudes (in microvolts) for hit responses only. As expected, RTs of both measurements show shorter RTs for validly than for invalidly primed trials for both smiles in joy evidence scores and EMG of the Zygomaticus major, and frowns in the anger evidence scores and the activation of Corrugator supercilii. Accordingly, rmANOVA on RTs revealed a significant main effect of validity for both FACET data, $F(1,29) = 100.79$, $p < .001$, $\eta_p^2 = .78$; and EMG data, $F(1,29) = 112.67$, $p < .001$, $\eta_p^2 = .80$; and also a significant main effect of facial expression for FACET data, $F(1,29) = 29.46$, $p < .001$, $\eta_p^2 = .50$, and for EMG data, $F(1,29) = 51.67$, $p < .001$, $\eta_p^2 = .64$. The validity by facial expression interaction was significant for FACET data, $F(1,29) = 33.73$, $p < .001$, $\eta_p^2 = .54$, but not for EMG data $F(1,29) = .80$, $p = .38$, $\eta_p^2 = .03$. Post hoc *t*-tests for FACET data presented in Table 2 show an overview of the significant effects for all pair-wise comparisons (e.g. *joy valid* vs. *joy invalid*, *joy valid* vs. *anger valid*). All comparisons for the FACET data except from joy valid vs. anger valid clearly reached the level of significance with medium to large effect sizes ranging between $d = .63$ and $d = 1.79$. For EMG only joy valid vs. joy invalid reached the level of significance with an effect size of $d = .30$.

Figure 4: Mean FACET and EMG amplitudes for hits of Study 2

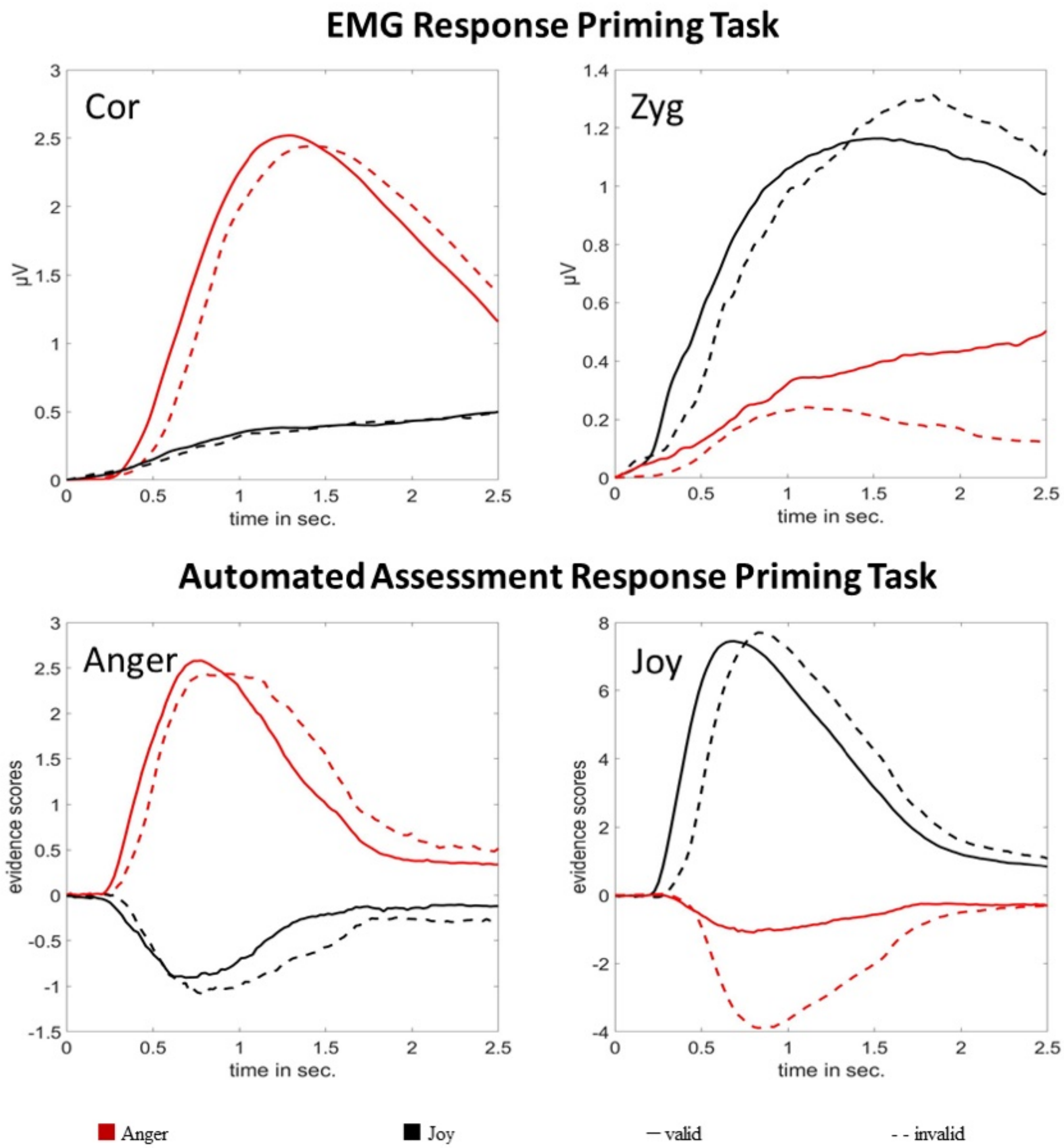


Figure 4. Top – mean EMG amplitudes (microvolts) of the Corrugator supercilii (Cor) and Zygomaticus major (Zyg). Bottom – mean FACET evidence scores for anger and joy channels. Time zero refers to the onset of response signal.

Table 2

Post-hoc *t*-tests of Study 2

FACET

Condition	*M**	*SD**	Comparison	*t*(29)	*p***	Cohen's *d*
(1) Joy valid	577.11	112.34	1 vs. 2	-16.02	< .001	1.79
(2) Joy invalid	805.52	141.54	1 vs. 3	-1.46	.08	0.22
(3) Anger valid	601.87	111.10	3 vs. 4	-6.69	< .001	0.89
(4) Anger invalid	716.11	143.73	2 vs. 4	-4.08	< .001	0.63

EMG

Condition	*M**	*SD**	Comparison	*t*(29)	*p***	Cohen's *d*
(1) Joy valid	683.17	381.55	1 vs. 2	-3.32	.002	0.30
(2) Joy invalid	811.36	488.71	1 vs. 3	0.46	.65	-0.01
(3) Anger valid	657.17	183.80	3 vs. 4	-0.56	.58	0.08
(4) Anger invalid	678.82	332.81	2 vs. 4	2.18	.04	-0.32

Note: *Unit in milliseconds. ** Uncorrected p values. All comparisons are one-tailed. After correction for multiple testing all p-values smaller than .05/4 = .0125 are significant (Bonferroni correction). *Cohen's d* = .20 corresponds to a *small effect*, d = .50 corresponds to a *medium effect* and d = .80 corresponds to a large effect (Cohen, 1988).

Figure 5 shows the relationship between the FACET and EMG estimates of RTs averaged across subjects and depicts the variation of this relationship between persons. As statistical modeling of these relations, we estimated a random intercept and random slope model predicting RTs obtained from FACET by RTs obtained from EMG. The model assumed that the expected value in FACET scores and the relationship between the two scores varies across persons. This model revealed an average standardized regression weight of .49 ($p < .01$). The variation of this relationship across persons was not statistically significant. The ICC estimated in a null model additionally indicated that only 20% of variance of RTs obtained from FACET were due to between-person differences.

Figure 5: Linear Mixed Model of Study 2

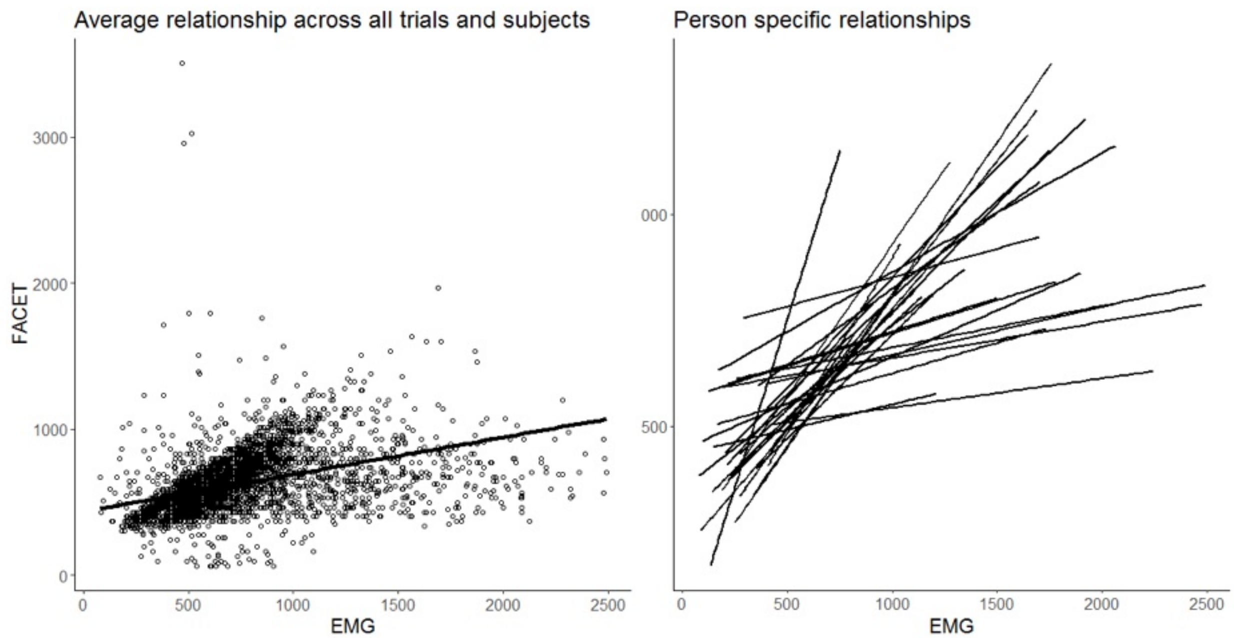


Figure 5. RTs estimated with FACET and their relationship with RTs estimated with EMG. Left side: Average descriptive relationship between FACET estimated RTs and EMG estimated RTs across all trials and persons. Right side: Person specific relationships in the observed sample.

5. Discussion

Study 2 shows that automated assessment of facial expression (here FACET) can replicate experimental effects observed with EMG, like longer RTs for invalidly primed facial motor responses (i.e., smiles and frowns). As observed by Recio et al. (2014), for FACET data the main effects of validity were larger for smiles than frowns. The estimated effect sizes show similar validity effects for FACET and EMG data. Within participants regressions of RTs obtained from FACET and EMG show a moderate correlation, indicating that the two measurements converge whereby the standardized regression weight indicates a medium to big effect (Cohen, 1988).

Nevertheless, measurement with automated assessment of videotaped muscle movements seems in some way different from EMG of the electrical potentials of moving facial muscles.

As mentioned in the introduction, automated assessment is arguably more specific than EMG (e.g., discrimination between chewing and smiling). Moreover, video-based analyses of facial expressions could be a less intrusive and more ecologically valid method to measure facial expressions than EMG. For example, in real social interactions an interlocutor probably reacts to visually salient expressions and less to the linearly scaled muscle tonus of moving muscles. The unique variance of each measurement can be due to different technical error sources of video recording vs. EMG, differences in data structure and parameterization (e.g., logarithmic probability scores vs. linear scale) and specificities of data analyses (e.g. adaptive thresholding vs. 25% of maximum). Furthermore, the fact that effect sizes were larger from FACET than from EMG (see table 2), suggests inherent differences between machine vision, relying on the face as a whole, and EMG analysis, relying on single muscle (e.g., AU4 for pressing eyebrows together). Whether this explains the methodical differences between FACET and EMG remains unclear and deserves further investigation.

6. General Discussion

We conducted two studies to estimate the reliability and validity of automated assessment of facial expressions. As shown in Study 1, classification performance of FACET is excellent for facial expressions of emotion and for AUs. FACET shows a robust functioning across different manipulations of facial stimuli from standardized databases, when they are varied within a range slightly exceeding the software recommendations. Factors influencing the quality of the video recordings (e.g., head movements) do not compromise a reliable measurement, at least within the values investigated in our study. However, it is crucial to control for technical factors like resolution. Such factors are easier to control for, as compared to participant's movements. In addition, the results from Study 2 demonstrate that FACET captures experimental effects in RTs of facial movements. We can conclude that quantitative research with facial expressions acquired in experimental settings is possible in a reliable and valid way using automated assessments (e.g., FACET).

6.1 Reliability

Study 1 demonstrated the reliable detection of facial expressions of emotion and AUs across different conditions applied to facial recordings. Several sources of noise-related variance can be eliminated using automated assessment as compared with human raters. For example, for machine vision there is no unwanted variance due to fatigue, mood, previous experience or sequence effects, as might be the case for human raters. Taking one step further, most of the measured noise-related variance originates from stimuli and the way FACET deals with them. Because the algorithm does not change over time, re-test reliability and objectivity should also be larger (implementation reliability of machine vision is approximately 1) compared to human raters classifying facial expressions using FACS, or for EMG-based classifications of non-emotional movements as emotional expressions (e.g., chewing would be not distinguishable from smile). The high reliability holds up even when FACET analyses data of non-optimal quality. Our first study showed a robust processing when the same stimuli were presented with different head positions (vertical and horizontal angles) and at different sizes. Nonetheless, the 2D transformations to simulate head positions used here is just an approximation of actual out-of-plane rotation and does not include complex interactions of different angles in 3D.

A benefit of the automated assessment in comparison to EMG, is that there is no need to instruct examiners with anatomical information about facial muscles. In addition, Study 1 suggests the necessity to control for technical factors (i.e., resolution), which is easy to achieve with a standard camera, at least in case of lab experiments, where recording conditions are easier to control than in field studies. Further research on this topic should include weak and less prototypical expressions in a more naturalistic environment (e.g., during a walk through a park). Facial expression classification algorithms may struggle in less optimal recording circumstances (e.g., flickering light). Here we found first evidence for proper classification quality under very controlled conditions slightly exceeding the optimal range, indicating that FACET represents a suitable alternative for experimental psychological investigations in well-controlled lab recording conditions.

Whereas we observed satisfactory reliability for FACET, caution is warranted concerning generalization to other software solutions. Standards for evaluating psychometric properties of automated assessment within the framework of modern test theory would be helpful to support psychological researchers, as already developed for technical details of AUs recognition (Valstar et al., 2015) and head pose recognition (Valstar et al., 2017). Other studies showed satisfying functioning of FACET, too. Dente et al. (2017) showed varying classification performance for different standardized databases and Stöckli et al. (2017) showed acceptable performance of FACET for spontaneous (non-posed) expressions in a group of participants, who watched emotion-inducing pictures. The contribution of our study is to provide an estimate of reliability and influencing factors (angle, brightness, resolution) of emotional expressions, and for the first time, of AU classification using FACET.

6.2 Validity

Our second study showed that it is possible to replicate experimental effects like reprogramming-costs in facial RTs during the production of facial expressions (Recio et al., 2014) with automated assessment coding. RTs were larger for invalidly primed than for validly primed smiles and frowns. Besides correct classification of expressions of emotion, it is important to point out that both expressions were prepared between prime and response signal and differences between RTs indicate inhibition of activated motor plans as well as reprogramming of alternative expressions. Altogether, the results of Study 2 suggest that FACET is a viable alternative to EMG for measuring experimental effects using facial expressions as responses – at least for those with large effect sizes. We saw convergence between results from EMG and automated assessment with effect sizes comparable with a former study using a very similar task (Recio et al., 2014).

Nevertheless, the medium to high association between RTs measured from automated assessment of videotaped facial movements and from EMG of electrical impulses of moving facial muscles also suggests differences between the two methods. While EMG rests upon linear increases of voltage in a muscle, intensity of facial expressions increases depending on the combined changes in different AUs. In other words, while muscle tension of Zygomaticus major is still low and increases over time during a smile, other parts like AU6, the cheek raiser (Orbicularis oculi and Pars orbitalis), might be near to maximal activation. Furthermore, the interplay between muscles might vary between different repetitions of smiling and frowning within every person. As a consequence, automated assessment in comparison with EMG may detect the onset of the movement a little earlier or later. This might explain the moderate to high association between RTs measured with EMG and FACET observed here. However, this difference does not seem to disturb the measurement of experimental effects, i.e., difference in RT between two experimental conditions. Therefore, we conclude that results of automated assessment converge with findings using EMG.

6.3 Limitations and future perspectives

There are some limitations in Study 1. First, we used simple manipulations of illumination, which might not reflect the complexity and contrast shifts between light and shadow in natural conditions. Also, we used portraits of posed expressions from standardized databases validated with an external criterion for expression classification. Although the MPI database provides different angles with additional information of the side of the face, not all possible rotations were considered. This limits the ecological validity, because it does not reflect spontaneous expressions during life interactions (e.g. ambiguous expressions in flickering sun light). Thus, additional work is needed to determine whether our results on reliability will also generalize to spontaneous expressions.

Although we estimated reliability for seven facial expressions and 19 AUs, we focused our investigation of validity on just two expressions, which are easy to identify for humans and probably for algorithms, too. We chose smiles and frowns as examples, because they are easily investigated with EMG (Wolf, 2015). We are not aware of any empirical study addressing the question of whether other expressions of basic emotions measured with automated assessment show similar or different performance as compared with EMG research results. Automated assessment gives us plenty of opportunities and overcomes problems associated with EMG (e.g., vague conductance of small muscles; Wolf, 2015).

Automated assessment has some limitations per se. For example, it cannot be more valid than trained human raters, as long as algorithms learn on datasets using human ratings as external criterion. That way, we tested for prototypical apex expressions only, possibly increasing the recognition performance of the machine vision software. Classification of expressions at the apex with maximal intensity is a fundamental demand that software solutions should generally satisfy. Results of this study refer to controlled manipulations of stimuli showing intense expressions. We aimed at identifying the factors that could influence precise facial expression measurement, but these factors varied in a relatively small range compared with all possible variations given outside laboratory settings. Although FACET comes with a detailed manual concerning the interpretation of evidence scores, a complex preprocessing of the data is necessary to analyze the data on a trial basis.

A next step would thus be to investigate performance of machine vision for facial expression recognition in more naturalistic situations, with non-trained posers, displaying less prototypical expressions of lower intensity. Beyond these limitations, automated assessment might be helpful to validate standardized databases of facial expressions or to provide additional controls in a set of face stimuli (e.g., Calvo, Fernández-Martín, Recio, & Lundqvist, 2018), as variance of inter- and intra-rater agreement is eliminated (Calvo et al., 2016). It might also be possible to program online feedback on performance posing facial expressions and integrate this feedback in standardized trainings for experimental instructions, or therapy programs.

Exclusion of participants in this study was rather conservative (e.g., no glasses, no beards etc.). Even though Study 1 includes Caucasians and Moroccans with no differences detectable between groups, there is empirical evidence for cultural differences of expression and decoding of emotion (Elfenbein & Ambady, 2002; Jack, 2013), for morphological differences between women and men (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007), and clearly, effects of age on emotion differentiation in both expression and decoding for children (Ganchrow, Steiner,

& Daher, 1983; Nelson, 1987). Finally, the influence of speech movement on classification is important in terms of external validity of automated assessment, like a smile during a flirty conversation, or differences in expression of emotion for diverse patient groups (Bantum & Owen, 2009). These aspects are still unknown and could be a topic of future research.

Several questions still remain open, regarding for instance optimal thresholds in other samples (thresholds were optimized for our particular sample); the smoothing of the signal (Olderbak et al., 2014); or whether FACET evidence scores can be understood as a linear function of expression intensity. Girard and colleagues (2015) investigated how to estimate expression intensity with computer analyses and found high reliability for intensity-trained multiclass classification. Further, for all research questions dealing with differences between emotions (e.g., joy can be inhibited faster than fear), it will be essential to determine whether evidence scores for all facial expressions are measured on the same scale. All these issues could be topics of future research.

Besides, in Study 2 we distinguished between two responses (i.e., smile vs. frown) for both FACET and EMG. Nonetheless, classification of facial expressions is a multiclass problem, which has to deal with multiple emotion-related expressions. When using automated assessment without any external criteria or second measurement (e.g., EMG), the information of all counter channels should be considered (e.g., checking thresholds for all possible counter channels of FACETS data output).

Finally, there is a need for applications to measure facial expressions, for research and for practical solutions like training programs for people with autism (Cockburn et al., 2008), automated assessment of facial expressions in patients with psychiatric disorders (Wang et al., 2008), automated detection of driver fatigue (Gu & Ji, 2004), or automated feedback for intelligent tutoring systems (D'Mello et al., 2008; Kapoor, Burleson, & Picard, 2007). The presented study contributes to previous efforts to find solutions for investigating humans' facial expression, by assessing the reliability and validity of automated assessment of facial expression with software. Our findings argue in favor of using resource-saving machine vision tools for assessment of facial expressions in psychological research. To conclude, we have shown that expression and AU assessment provided by automated assessment, here FACET, provides reliable and valid measurements of prototypical facial expressions of emotion and AU activation in standard experimental conditions.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft [grant number RE 3721/2-1 to GR]. The authors thank Christopher Dannheim, Daron Rother and Espen Henken for their help with data collection and analyses. We verified license agreements of iMotions before publishing the present study.

Competing interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplemental Material

Figure A: Example of FACET's evidence scores

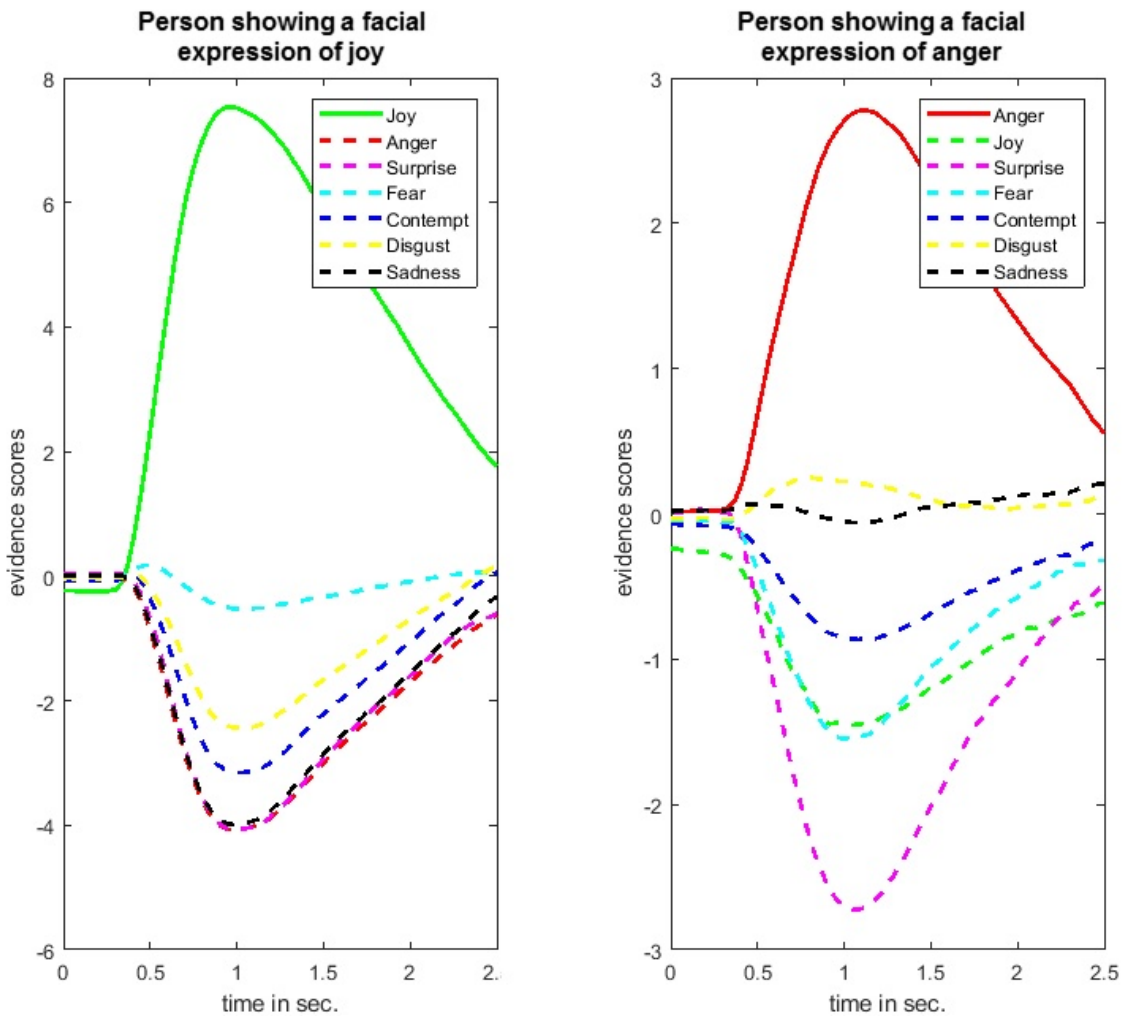
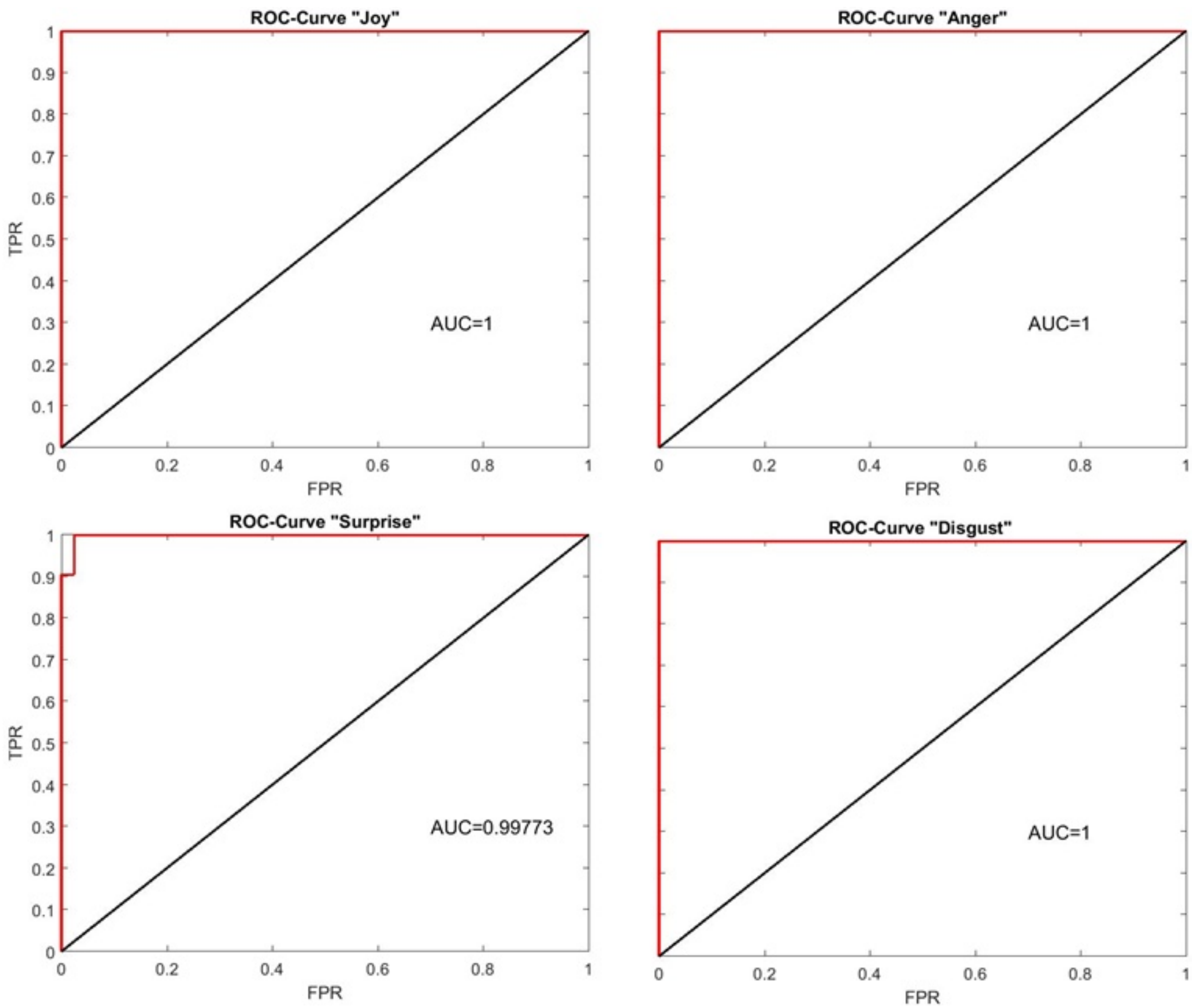
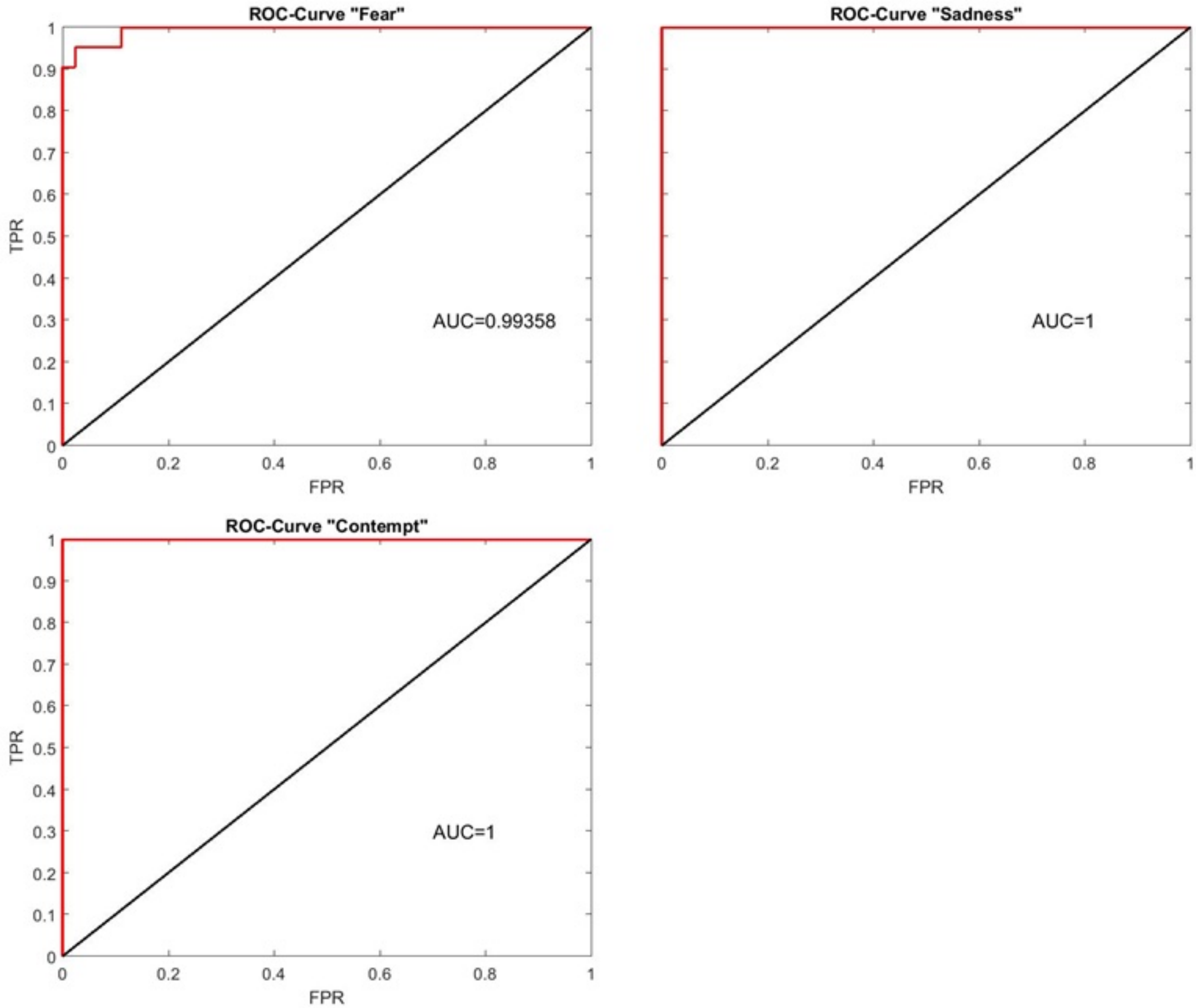


Figure A. Two examples showing increases in evidence scores as a function of time for fully intense expressions of joy (left) and anger (right). The evidence scores for the expression shown in that frame increase above zero, whereas the scores for other expressions decrease below zero.

Figure B: Receiver operating characteristic curves of Study 1




ROC-Curve "Fear"
AUC=0.99358
ROC-Curve "Sadness"
AUC=1
ROC-Curve "Contempt"
AUC=1
TPR
FPR

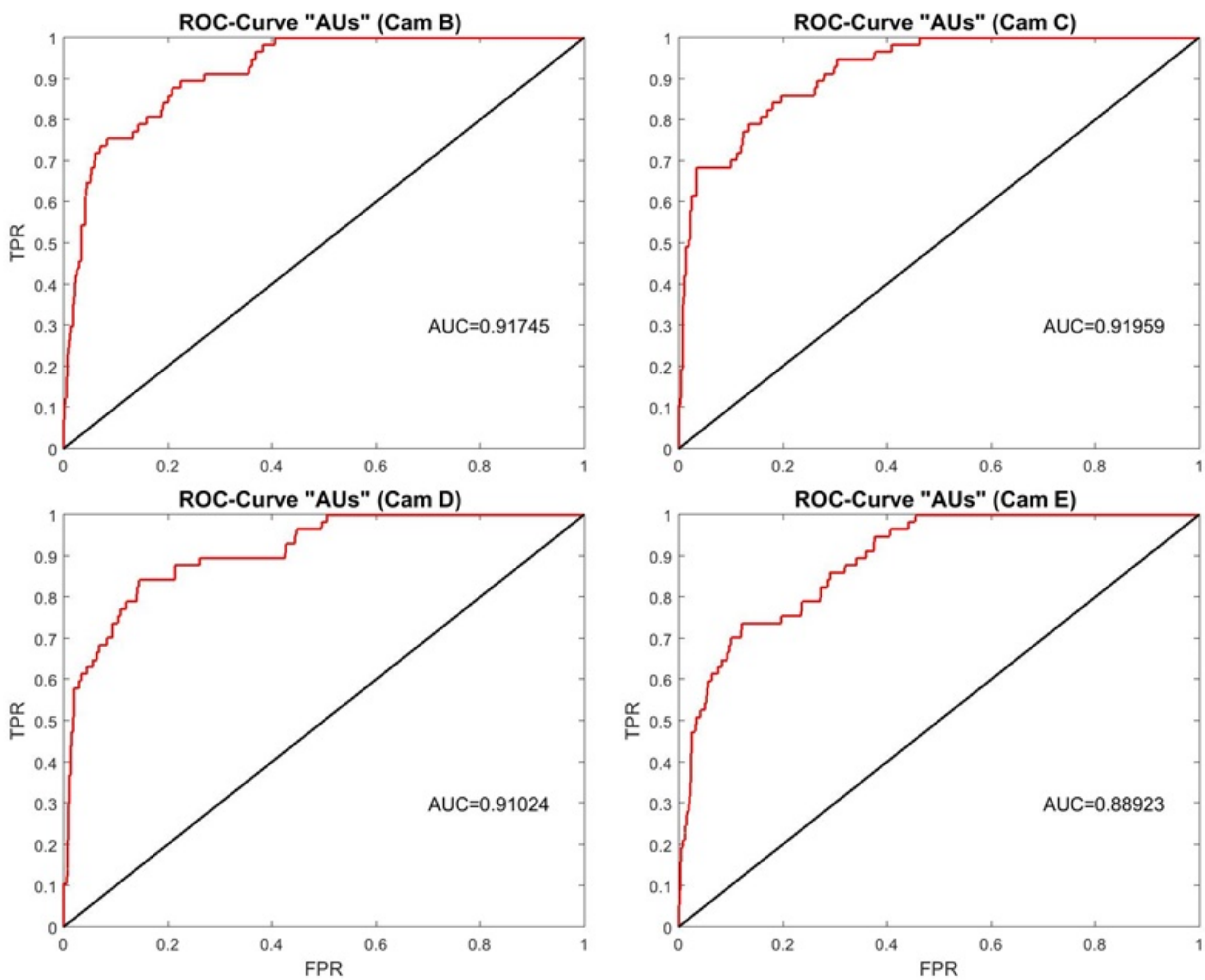


Figure B. Area under the curve (AUC) for each emotion of all original trials of ADFES database and for each camera viewpoint of the MPI database. Cam B = + 9°, Cam C = -9°, Cam D = +27°, Cam E = -27°, TPR = True Positive Rate, FPR = False Positive Rate.

References

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). *Openface: an open source facial behavior analysis toolkit.* Paper presented at the Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on.

Bantum, E. O. C., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological assessment, 21*(1), 79.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of personality and social psychology, 92*(2), 179.

Blickle, G., Momm, T., Liu, Y., Witzki, A., & Steinmayr, R. (2011). Construct validation of the Test of Emotional Intelligence (TEMINT). *European Journal of Psychological Assessment*.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., . . . Zhao, W. (2015). *Automatic detection of learning-centered affective states in the wild.* Paper presented at the Proceedings of the 20th international conference on intelligent user interfaces.

Cacioppo, J. T., Petty, R. E., & Marshall-Goodell, B. (1984). Electromyographic specificity during simple physical and attitudinal tasks: Location and topographical features of integrated EMG responses. *Biological psychology, 18*(2), 85-121.

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion, 16*(8), 1186.

Calvo, M. G., Fernández-Martín, A., Recio, G., & Lundqvist, D. (2018). Human observers and automated assessment of dynamic emotional facial expressions: KDEF-dyn database validation. *Frontiers in psychology, 9*, 2052.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion, 30*(6), 1081-1106.

Cockburn, J., Bartlett, M., Tanaka, J., Movellan, J., Pierce, M., & Schultz, R. (2008). *Smilemaze: A tutoring system in real-time facial expression perception and production in children with autism spectrum disorder.* Paper presented at the ECAG 2008 Workshop Facial and Bodily Expressions for Control and Adaptation of Games.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. *Hillsdale, NJ*, 20-26.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The handbook of emotion elicitation and assessment*, 203-221.

Cohn, J. F., & De la Torre, F. (2014). Automated face analysis for affective. *The Oxford handbook of affective computing*, 131.

Cohn, J. F., & Ekman, P. (2005). Measuring facial action. *The new handbook of methods in nonverbal behavior research*, 9-64.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE transactions on pattern analysis and machine intelligence, 23*(6), 681-685.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., . . . Picard, R. (2008). *AutoTutor detects and responds to learners affective and cognitive states.* Paper presented at the Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems.

Darwin, C. (1998). *The expression of the emotions in man and animals*: Oxford University Press, USA.

Dente, P., Küster, D., Skora, L., & Krumhuber, E. (2017). *Measures and metrics for automatic emotion classification via FACET.* Paper presented at the Proceedings of the Conference on the Study of Artificial Intelligence and Simulation of Behaviour (AISB).

Deriso, D., Susskind, J., Tanaka, J., Winkielman, P., Herrington, J., Schultz, R., & Bartlett, M. (2012). *Exploring the facial expression perception-production link using real-time automated facial expression recognition.* Paper presented at the Computer Vision–ECCV 2012. Workshops and demonstrations.

Ekman, P., & Friesen, W. (1978). *Manual for the facial action coding system*: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of nonverbal behavior, 6*(4), 238-252.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *J Pers Soc Psychol, 53*(4), 712-717.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology, 30*(1), 527-554.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin, 128*(2), 203.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.

FFmpeg-Developers. (2016). ffmpeg tool (Version beld324) [Software].

Frank, R. H. (1988). *Passions within reason: the strategic role of the emotions*: WW Norton & Co.

Fridlund, A., & Izard, C. E. (1983). Electromyographic studies of facial expressions of emotions and patterns of emotions. *Social psychophysiology: A sourcebook*, 243-286.

Fridlund, A. J. (2014). *Human facial expression: An evolutionary view*: Academic Press.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23*(5), 567-589.

Friesen, W. V., & Ekman, P. (1983). EMFACS-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco, 2*(36), 1.

Ganchrow, J. R., Steiner, J. E., & Daher, M. (1983). Neonatal facial expressions in response to different qualities and intensities of gustatory stimuli. *Infant Behavior and Development, 6*(2-3), 189-200.

Girard, J. M., Cohn, J. F., & De la Torre, F. (2015). Estimating smile intensity: A better way. *Pattern recognition letters, 66*, 13-21.

Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behavior research methods, 47*(4), 1136-1147.

Gu, H., & Ji, Q. (2004). *An automated face reader for fatigue detection.* Paper presented at the Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on.

Hess, U., Arslan, R., Mauersberger, H., Blaison, C., Dufner, M., Denissen, J. J., & Ziegler, M. (2017). Reliability of surface facial electromyography. *Psychophysiology, 54*(1), 12-23.

Jack, R. E. (2013). Culture and facial expressions of emotion. *Visual Cognition, 21*(9-10), 1248-1286.

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International journal of human-computer studies, 65*(8), 724-736.

KelesL, P., Díyarbakirli, S., Tan, M., & Tan, Ü. (1997). Facial asymmetry in right-and left-handed men and women. *International Journal of Neuroscience, 91*(3-4), 147-159.

Kleiner, M., Wallraven, C., & Bülthoff, H. H. (2004). The MPI VideoLab-A system for high quality synchronous recording of video and audio from multiple viewpoints. *MPI-Tech. Reports*(123).

Leonhart, R., & Lichtenberg, S. (2009). *Lehrbuch Statistik*: H. Huber.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). *The computer expression recognition toolbox (CERT).* Paper presented at the Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.

Malmir, M., Forster, D., Youngstrom, K., Morrison, L., & Movellan, J. (2013). *Home alone: Social robots for digital ethnography of toddler behavior.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision Workshops.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*(1), 30.

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Communication (includes Communication Reports), 52*(1), 1-22.

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: Mechanisms of development. *Child development*, 889-909.

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes-Ker, A. H. (2000). Emotional state and the detection of change in facial expression of emotion. *European Journal of Social Psychology, 30*(2), 211-222.

Olderbak, S., Hildebrandt, A., Pinkpank, T., Sommer, W., & Wilhelm, O. (2014). Psychometric challenges and proposed solutions when scoring facial emotion expression codes. *Behavior research methods, 46*(4), 992-1006.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*, 1.

Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R)*: Hogrefe Göttingen.

Recio, G., & Sommer, W. (2018). Copycat of dynamic facial expressions: Superior volitional motor control for expressions of disgust. *Neuropsychologia*, *119*, 512-523.

Recio, G., Shmuilovich, O., & Sommer, W. (2014). Should I smile or should I frown? An ERP study on the voluntary control of emotion - related facial expressions. *Psychophysiology, 51*(8), 789-799.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality, 11*(3), 273-294.

Schlegel, K., Fontaine, J. R., & Scherer, K. R. (2017). The Nomological Network of Emotion Recognition Ability. *European Journal of Psychological Assessment*.

Shafiq, S., Tauseef, H., Fahiem, M., & Farhan, S. (2017). AN ALGORITHM FOR FACIAL EXPRESSION BASED AUTOMATIC DECEPTIVE PAIN DETECTION. *Pakistan Journal of Science, 69*(1).

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2017). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior research methods*, 1-15.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature neuroscience, 11*(7), 843-850.

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS biology, 15*(3), e2000797.

Terzis, V., Moridis, C. N., & Economides, A. A. (2010). *Measuring instant emotions during a self-assessment test: the use of FaceReader.* Paper presented at the Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research, 168*(3), 242-249.

Troje, N. F., & Bulthoff, H. H. (1996). Face recognition under varying poses: the role of texture and shape. *Vision Res, 36*(12), 1761-1771.

Valstar, M. F., Almaev, T., Girard, J. M., McKeown, G., Mehu, M., Yin, L., . . . Cohn, J. F. (2015). *Fera 2015-second facial expression recognition and analysis challenge.* Paper presented at the Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on.

Valstar, M. F., Sánchez-Lozano, E., Cohn, J. F., Jeni, L. A., Girard, J. M., Zhang, Z., . . . Pantic, M. (2017). FERA 2017-Addressing Head Pose in the Third Facial Expression Recognition and Analysis Challenge. *arXiv preprint arXiv:1702.04174*.

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion, 11*(4), 907-920. doi:10.1037/a0023853

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current directions in psychological science, 18*(3), 184-188.

Wang, P., Barrett, F., Martin, E., Milonova, M., Gur, R. E., Gur, R. C., . . . Verma, R. (2008). Automated video-based facial expression analysis of neuropsychiatric disorders. *Journal of neuroscience methods, 168*(1), 224-238.

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Transactions on Affective Computing, 5*(1), 86-98.

Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in clinical neuroscience, 17*(4), 457.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35.

2.2 Publication 2: Beringer, M., Wacker, J., & Recio, G. (2022). Deliberate control of facial expressions in a go/no-go task: An ERP study. *Acta Psychologica*, 230, 103773.

I will now present the second publication of our research project. This publication addresses Objective 2: Investigating the interactions between emotion and executive control over facial expressions. Additionally, it tackles Objective 4: Evaluating how social context variables influence the control of facial expressions. Specifically, this objective examines whether social motives can enhance or hinder facial expression control.

Deliberate control of facial expressions in a go/no-go task: An ERP study

M. Beringer [a], J. Wacker [a], G. Recio [b]

[a] Universität Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany

[b] Universitat de Barcelona, Insitute of Neuroscience, Passeig de la Vall d'Hebron 171, ES-08035 Barcelona, España

m.q.beringer@gmail.com, jan.wacker@uni-hamburg.de, guillermo.recio@ub.edu

Funding: This work was supported by the German Research Foundation [grant number RE 3721/2-1 to GR].

Corresponding author:
Matthias Beringer
Dillstraße 1,
D-20146 Hamburg, Germany
m.q.beringer@gmail.com
Tel: +49 (0)176 303 116 77

Abstract

In two studies we investigate the role of affective factors and top-down processes underlying production and deliberate control of emotional facial expressions and its neural underpinnings. In Study 1 we examine facial expressions of joy, fear and disgust depending on the emotional content of the visual stimuli (upright faces, inverted faces, emotion inducing pictures without faces). In Study 2 we focus on expressions of joy and disgust depending on gaze direction (with and without eye contact) in a more natural setting with a real person as stimulus. We hypothesized that the more automatic processes are induced by stimuli (e.g., arousal, mimicry or social cues like eye contact) the harder it is to control facial expressions; particularly expressions of joy compared to fear and disgust. In both studies we used go/no-go tasks and showed shorter RTs for conditions with upright faces or eye contact, respectively. We also found shorter RTs for expressions of joy than of fear and disgust. In Study 1 participants showed more errors in no-go trials for expressions of joy than for expressions of fear and disgust, indicating worse top-down control for expressions of joy than of fear or disgust. An ERP analysis of the no-go P3 in Study 1 revealed larger amplitudes for upright faces compared with both inverted faces and emotion inducing pictures and larger amplitudes for expressions of joy than for disgust. This indicates greater demand of top-down control when automatic mimicry processes are activated and some degree of specificity for certain facial expressions. In Study 2 more errors in no-go trials in conditions with eye contact only for expressions of joy indicate mimicry could be larger for expressions with high affiliative intent like expressions of joy, and reduced mimicry for negative expressions. All results indicate that facial expressions buffered by automatic processes (e.g., mimicry) have a greater need for top-down control, especially expressions of joy compared to expressions of fear and disgust.

Keywords: Facial expression, top-down control, go/no-go task, Gaze, Mimicry

Highlights

- We investigated deliberate control of emotional facial expressions in a go/no-go task.
- We found differences for expressions of joy compared to those of fear and disgust.
- ERPs indicate greater demand of control when automatic processes are activated.
- Eye contact modulates automatic processes for expressions of joy.

Deliberate control of facial expressions in a go/no-go task: An ERP study

Control over facial expressions in social interactions is essential in our daily context. Examples of deliberate control of nonverbal communication include keeping an expression of joy in public moments (e.g., as a politician despite feeling nervous), as well as restraining expressions of fear (e.g., when a policeman breaks up a fight and needs to respond to the facial expression of his opponents with determination) and disgust (e.g., when your date invited you for a self-made dinner but accidentally burned the meal).

Realizing what kind of expression of emotion is adaptive to what extent in a given social situation is fundamental for successful interaction and needs to be planned, monitored and readjusted. This strategic use of voluntary control point to the fact that facial expressions are not only spontaneous epiphenomena of emotional states (Krumhuber & Kappas, 2005), but can be used consciously to communicate complex information (Hayes & Metts, 2008), and to push the interlocutor to perform a specific behavioral or emotional response (Knapp & Daly, 2002; Manstead & Fischer, 2001). Production and perception of emotion are fundamental aspects of social intelligence (Mayer, Roberts, & Barsade, 2008; Scherer, 2009) and have been studied intensively (Adolphs, 2002; Dimberg, Thunberg, & Elmehed, 2000; Elfenbein & Ambady, 2002; Tottenham et al., 2009).

Complex combinations of situational, cultural, and individual factors can be decoded by humans to understand each other, or to modulate facial expressions according to current demands. Much about our own emotional state is conveyed by facial expressions, e.g., spontaneous expression of positive emotion versus deliberate attempts to appear as if positive emotion is felt versus acknowledgements of feeling miserable but not intending to do much about it (Ekman & Friesen, 1982). Moreover, emotion specific changes in facial expressions enhance perceptual processing of information congruent with our own emotional state, e.g., such that happy people perceive happy expressions rather than neutral or sad expressions (Niedenthal, Halberstadt, Margolin, & Innes-Ker, 2000). The complex interaction of facial expressions with other cognitive functions is even true for factors like gaze direction, which can be used as a cue for information of interest or aversion (Bayliss, Frischen, Fenske, & Tipper, 2007). These examples show the importance of the interplay between facial expressions of emotion and cognitive processing. Whereas facial expressions as tools of nonverbal behavior are quite interesting per se (e.g. for communication research, see Mehrabian, 2017, for an overview), they are not to be confused with emotions, e.g., emotions are no necessary or sufficient preconditions of certain spontaneous expressions (Fernández-Dols & Ruiz-Belda, 1997), even though they are strongly related to the emotional state (Frijda, Scherer, & Sander, 2009).

In his early research on facial expressions of emotion Charles Darwin (1872; 1998), observed typical mimic movements for discrete emotional states, often termed “basic emotions”. A formalized atlas of functional face movements of individual muscles or groups of muscles (so called actions units, AUs) related to basic emotions, the *emotional facial action coding system*, emFACS-7 (Friesen & Ekman, 1983), has been shown to be valid for assessing facial expressions with considerable commonalities across cultures (Ekman, 1971; Ekman et al., 1987). Ekman (1999) describes basic emotions as distinctive universal signals that all humans across cultures have in common, with an emotion-specific physiology and automatic appraisal mechanisms activated through universal antecedent events.

In sharp contrast to the idea of basic emotions, other frameworks describe emotions in a dimensional way rather than as discrete categories (Russell, 1980; Russell & Mehrabian, 1977), considering emotions as neither necessary nor sufficient for facial expressions (Russell &

Fernández-Dols, 1997), but rather as a vehicle in social interaction (Schmidt & Cohn, 2001). There is an ongoing debate whether emotions differ on dimensions of valence, dominance, and appraisal rather than being distinctive in their experience, physiology, and expressions per se. It is also clear that emotion related expressions are just one of several determinants of facial movements besides speech, postural changes, and inner experience. For instance, facial mimicry has been shown to be dependent on emotional and social context (Hess & Bourgeois, 2010). In sum, these perspectives and phenomena argue not only for different functions of facial expressions beside the relation to emotion (e.g., Scherer, 1992) but also for different perspectives on cognitive control mechanisms of facial expressions of emotion. We understand the pushing factor (see Scherer, 1992) as a driving impulse to express the underlying emotion promptly, which should be largest for disgust. From an evolutionary perspective, this makes sense, as the disgust expression is associated with behavior to reject something potentially infectious by e.g., decreasing the surface of mucosa.

Mimicry, executive functions, and facial expressions

Accurate perception and adequate production of facial expressions of emotion are two important aspects of social interaction, commonly inextricably linked to each other. Therefore, deliberate control of inappropriate expressions occupies a key role, too, and is conceptualized as one aspect of the broad framework of executive functions. These encompass different cognitive top-down processes to adapt and control behavior, often divided into updating, switching, and inhibition (Diamond, 2013), which describe the voluntary control of a prepotent or automatic motor response (Miyake et al., 2000). The two studies presented here focus on the top-down control over facial expressions as a form of motor control (e.g., Morecraft, Stilwell–Morecraft, & Rossing, 2004).

Prepotent responses are created, for example, by the tendency to imitate perceived actions, e.g., finger movements (Brass, Zysset, & von Cramon, 2001) and facial expressions (Achaibou, Pourtois, Schwartz, & Vuilleumier, 2008; Postma & Postma-Nilsenová, 2016). They are involuntary activations in the same muscle groups of subjects corresponding to the facial expressions of a perceived model, even when subjects are told not to react (Dimberg, Thunberg, & Grunedal, 2002). Facial mimicry reflects automatic processes, as it has been observed within the first 500 ms after stimulus onset, robustly over several successive trials (Harrison, Morgan, & Critchley, 2010), and even when expressions of emotion were task irrelevant (Cannon, Hayes, & Tipper, 2009).

Furthermore, event-related potential (ERP) studies provide insights of cognitive processes involved in perception, production and deliberate control of facial expressions (e.g., Korb, Grandjean, & Scherer, 2008; Recio & Sommer, 2018). The go/no-go task is a common paradigm to investigate behavioral inhibition (e.g., Falkenstein, Hoormann, & Hohnsbein, 1999; Schulz et al., 2007), used by Korb, Grandjean and Scherer (2010), to investigate mimicry and inhibition of facial expressions.

The general idea of the go/no-go task in the present study is to show participants different types of visual stimuli (e.g., pictures of smiling faces, inverted faces, emotional images) and a cue (e.g., a colored dot in the center of the stimuli) which indicates whether they need to react (e.g., smile = go trial) or not react (e.g., hold a neutral facial expression). A ratio of more go trials (e.g., 70%) to no-go trials (e.g., 30%) generates a preponderant preparedness to respond in go trials, which needs to be inhibited in no-go trials. Automatic processes are driven by the induced emotional arousal of the stimuli. The more emotional arousal is induced (e.g., faces vs. inverted faces), the larger the facilitation effect will be in go trials. With Study 1 we investigate whether mimicry can be a component that automatically enhances the ability to react promptly in go trials (e.g., when smiling back) that needs to be inhibited in a top-down process in no-go trials. As we use stimuli of faces, inverted faces and emotion inducing pictures we can separate effects of emotional arousal and mimicry.

The no-go P3, defined as a larger P3 in no-go trials than in go trials, has been found as a robust ERP, associated with response inhibition in general (e.g., Fallgatter, Brandreis, & Strik, 1997; Luck & Kappenmann, 2011) and more specifically with motor inhibition of hand movements (e.g., Falkenstein, Hoormann, & Hohnsbein, 1999). However, later components like P3, peaking around 350 ms after stimulus onset (Spencer, Dien, & Donchin, 2001), come with larger amplitudes for conditions which need more cognitive resources. The benefit of the no-go P3 as a difference wave (P3 in no-go trials – P3 in go trials) is that it allows to compare the P3, associated with response inhibition, between four conditions in one step (e.g., response inhibition of expressions of joy, P3 in no-go trials with upright smiling faces – P3 in go trials with upright smiling faces vs. response inhibition of expressions of joy, P3 in no-go trials with inverted smiling faces – P3 in go trials with inverted smiling faces). To the present, the use of facial responses in ERP studies is rare, with little studies on the neuronal basis of the control of facial expressions (Recio, Shmuilovich, & Sommer, 2014; Recio & Sommer, 2018). One lingering question is whether the motor programs of facial expressions are governed by automatic processes or can be modulated by executive control (e.g., reflected by P3) and whether these effects impact all types of facial expressions of emotion in the same way.

There is little literature about emotion specificity of executive control and automatic mimicry processes (e.g., Hess, & Fischer, 2013; Hess, & Fischer, 2014). For example, while expressions of joy are mimicked regardless to the group of the expresser, mimicry of expressions of sadness can only be observed between ingroup members (Burgeois & Hess, 2008). Further literature suggested distinct neural systems that motivate appetitive behavior (e.g., expressions of joy) and avoidance behavior (e.g., expression of disgust; Cacioppo, & Gardner, 1999; Davidson, 1995) supporting hypotheses of emotion specific effects for the control over facial expressions of emotion. In other words, our control over facial expressions of emotion could build on selective effects over automatic processes such as mimicry and can therefore help to resolve cognitive conflicts by prioritizing executive control (Gray, 2004). On the one hand, expressions of negative affects might be associated with an enhancement of attention and a faster processing but might impair the control of unwanted signals of threat from awareness (Pessoa, Padmala, Kenzer, & Bauer, 2012). On the other hand, expressions of joy might be faster to produce over all conditions and harder to inhibit, decrease cognitive control and increase commission errors. The open question remains whether those expression specificity effects can be observed on a positive-negative valence scale (as Russel's theory implies; Russel, 1980) or must be drafted for distinct basic emotions. Scherer (1992) suggests using the term "push effect" for the internal mechanism of an emotion pushing its expression to the surface by the operation of physiological changes in the service of adaptation. Scherer's framework would predict faster reactions for expressions of disgust than for other negative facial expressions as of disgust comes with the clearest internal push factor (Scherer, 1992). With data presented here, we bring behavioral effects of the control over facial expressions (e.g., reaction times and errors) together with a neural underpinning of top-down control and automatic processes of the production of facial expressions.

Most previous studies employed controlled stimuli (e.g., pictures of standardized databases, e.g., Beringer et al., 2019) in order to reduce noise and eliminate impact of various third variables. Nonetheless, it remains unclear whether findings of this kind of studies will generalize to more natural settings (e.g. encounters between subjects including social cues like gaze direction), too (e.g., Risko, et al, 2012). Fischer and van Kleef (2010) argue that emotional interactions and social cues are rarely captured in emotion research and theoretical frameworks. Pönkänen et al. have shown that being in a face-to-face situation can enhance the processing of facial information, when another persons' gaze addresses the interlocutor (Pönkänen, Alhoniemi, Leppänen, & Hietanen, 2010). Intuitively that makes sense, as for social cues it is quite important to distinguish between a situation in which an interlocutor gazes at you while showing anger or gazes at something or someone else (i.e. may be preparing to attack you or someone else). In other words, both gaze direction and expression of emotion can indicate whether someone is liked or disliked. It seems plausible that different types of expressions are related to different tendencies of gaze direction (e.g., showing joy to someone with eye contact versus turning away when feeling disgust). Bayliss, et al. (2007) have shown that expressions of joy and disgust modulate the use of gaze cues relating to affective evaluations of objects. To this point, the question remains how top down control over facial expressions of emotion interacts with gaze direction.

It is important to investigate the above-mentioned research questions, because the link between cognitive theories of deliberate control and emotion theories is still unclear. In sum, in Study 1 we investigated behavioral and ERP correlates of deliberate control of three different facial expressions of emotion (joy, fear, disgust), using a go/no-go task in a commonly controlled laboratory setting. Participants posed facial expressions while observing three stimulus types, namely, upright faces, inverted faces or emotion inducing pictures without faces, presuming we induced mimicry and emotional states. We were also interested in differences in control between expressions. Study 2 investigated the impact of gaze direction on two facial expressions of emotion (joy, disgust), using a similar go/no-go task but including social cues in a face-to-face situation.

Study 1

Does stimulus type modulate production and top-down control of specific expressions of emotion? Generally, we expected shorter RTs for conditions inducing more emotional arousal, because automatic processes facilitate expressions in go trials (e.g., RTs faces < RTs inverted faces, RTs emotion inducing pictures). However, for disgust inducing pictures an internal push factor might accelerate RTs and facilitate the production of face expressions relative to upright faces, based on automatic reactions to modify the organism in response to emotional events as described by Scherer (1988) (RTs of facial expressions of disgust in response to emotion inducing pictures < RTs of facial expressions of disgust in response to faces, inverted faces). To separate this hypothesis from valence effects (Russell, 1980), we also included a condition with expressions of fear, which are also negative in valence, but the internal push factor should be diminished (RTs of facial expressions of fear in response to faces < RTs of facial expressions of fear in response to inverted faces, emotion inducing pictures).

Even though emotional arousal of emotion inducing pictures can be a source of errors we expect that mimicry has the bigger impact on error rates. Regarding errors in no-go trials we expected more errors in conditions inducing mimicry than in conditions without mimicry (error rates faces > error rates inverted faces, emotion inducing pictures). There is growing evidence for effects of positive affects decreasing cognitive control (e.g., Hefer, & Dreisbach, 2020). Here we hypothesized more errors for expressions of joy than for expressions of fear or disgust. For inhibition errors we did not expect interaction effects between the factors stimulus type (faces, inverted faces, emotion inducing pictures) and type of expression of emotion (expressions of joy, disgust and fear) as display rules allow expressions of joy in many situations, but strongly oppose showing disgust or fear, especially to people of an outgroup (Matsumoto, 1990).

In Study 1 we compared stimuli of upright faces with inverted faces to isolate the effect of mimicry (e.g., Knight & Johnston, 1997). But as faces with expressions of emotion not only launch mimicry processes but induce emotion themselves (e.g., Moody, McIntosh, Mann, & Weisser, 2007), we also used emotional pictures without faces, to induce emotions but no mimicry. That way, we wanted to separate the impact of mimicry from pure emotion induction. Based on former studies, we expected larger no-go P3 amplitudes for conditions inducing mimicry than for those without mimicry (no-go P3 upright faces > no-go P3 inverted faces, emotion inducing pictures) and larger no-go P3 for expressions of joy than for expressions of fear or disgust (e.g., Schulz et al., 2007), reflecting more cognitive resources for conditions which demand more facial control (e.g., Recio & Sommer, 2018; no-go P3 expressions of joy > no-go P3 expressions of fear, disgust).

Methods Study 1

Participants

An a priori power analysis conducted with G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) suggested a total sample size of 26 participants for a repeated measurement design, estimated $\eta_p^2 = 0.1$ (Korb et al., 2010), an alpha error probability of 0.05, desired power of 0.80, one group, three measurements, a correlation among repeated measurements of .50, and nonsphericity correction of 0.5. Anticipating some dropouts, we recruited a sample of 40 participants (total sample: 40 healthy adults, 50% women, $M_{age} = 25.2$ years, $SD_{age} = 4.0$). In order to make sure that participants' facial expressions could be correctly measured without any visual barriers for automated assessment tools of facial expressions, we followed the same inclusion and exclusion criteria as in our previous publication (Beringer et al., 2019; exclusion of participants with full beard or glasses, with more than ±1.00 diopter unless they wear contact lenses). All participants provided informed consent. Psychology students (20% of the sample) received course credits and all others 22 Euro for their contribution. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Deutsche Gesellschaft für Psychologie (protocol number GR 112016_amd_082014).

To retain good data quality we excluded six participants who had a number of incorrect expressions in go trials over all facial expressions (Leonhart & Lichtenberg, 2009) (more than 2 *SDs* incorrect expressions higher than the mean) and another participant who had participated in a previous experiment with a very similar task. The final test sample consisted of 33 participants (52% women, M_{age} = 25.0 years, SD_{age} = 4.1).

Procedure and Apparatus

After signing consent forms, participants sat on a fixed chair in a quiet and electromagnetically shielded chamber. We presented all instructions including stimuli on a 21-inch LCD display (75 Hz refresh rate), approximately 80 cm from participants' eyes and used indirect light of three LED stripes (40-60cm length) to illuminate participants' faces homogeneously. For video recording we used a Logitech HD Pro Webcam C920 on a sample rate of 30 frames per second, fixed at the bottom of the monitor. We used the software FACET (version 6.1.2667.3, iMotions, 2016) to analyze videos of the participants' facial expressions. The software provides results of the video analyses frame by frame as evidence scores for each expression (joy, surprise, anger, disgust, sadness, contempt, fear) in decimal logarithmic scale. For example, an evidence score of zero in joy indicates that in this frame it is equally likely, that the targeted face shows joy, as that it does not. Evidence scores can be transformed to probability (*P*) using the following formula: $P = 1/1+10^{-\text{evidence score}}$. Research on reliability and validity of FACET demonstrated satisfying psychometric quality (Beringer et al., 2019).

Participants were instructed to either show expressions of joy, disgust or fear as fast as possible after stimulus presentation in go trials and to return to a neutral face right after apex. In no-go trials participants were instructed to maintain a neutral expression. Go trials had a higher occurrence than no go trials, which evokes a preponderant response to show an appropriate expression.

Participants began producing facial expressions in a short training phase with up to five calibration trials per expression, being asked to start with an expression of joy, followed by expressions of disgust and fear. During each trial participants received visual feedback from FACET evidence scores in form of percentages increasing with intensity of the specific facial expression – getting more difficult (+25%) in the next calibration trial when participants reached the top and easier (-25%) when participants failed to reach the intended intensity, similar to the training procedure described in Recio and Sommer (2018). That way, we wanted to make sure that participants were able to show prototypical expressions detectable by FACET. Indeed, all participants passed the training phase, and the software was able to score their facial expressions.

Then, three experimenters (two women), randomly assigned to participant numbers, prepared EEG recordings. They left the participants alone in the shielded chamber to systematically monitor the correct recordings of EEG electrodes and the video for FACET (e.g., impedances, face detection, etc.). Instructions on the monitor emphasized not to move the head and to show all expressions as intensely as possible and to return to a neutral face after apex immediately.

In the following part of the experiment, we implemented a go/no-go task, organized in three blocks. In go trials of block 1 participants expressed joy to pictures of models with expressions of joy, displayed fearful expressions to pictures of fear-showing models, and disgust expressions to pictures of disgust-showing models – as fast as possible after stimulus onset. Meanwhile in no-go trials, participants inhibited their expression. The temporal sequence of the events in a trial is presented in Figure 1. In block 2 we presented pictures of models' faces upside down to reduce automatic response tendencies like mimicry (Bruce, 2017). In block 3 we used pictures without faces from the OASIS picture set (Kurdi, Lozano, & Banaji, 2017) to induce emotional states of joy, disgust and fear aiming for minimal mimicry. We instructed participants in go trials to show a corresponding facial expression to the presented picture, and in no-go trials to keep a neutral face. We used "W" and "M" (randomly assigned between subjects) as small symbols on the noses of the models' pictures and in center of the emotion inducing pictures respectively to inform participants whether a trial is a go or no-go trial. To sum up, there were eighteen different task conditions: (1-3) in *go joy upright faces, go joy inverted faces, go joy emotion inducing pictures* participants saw a joy-associated picture and produced an expression of joy; (4-6) in *go fear upright faces, go fear inverted faces, go fear emotion inducing pictures*, participants saw a fear associated picture and produced a fear expression; (7-9) in *go disgust upright faces, go disgust inverted faces, go disgust emotion inducing pictures*, participants saw a disgust associated picture and produced a disgust expression; and (10-18) in no-g*o joy/fear/disgust upright faces, no-go joy/fear/disgust inverted faces, no-go joy/fear/disgust emotion inducing pictures*, participants saw a joy/fear/disgust associated picture and maintained a neutral expression. In every block every expression was requested 120 times, go trials had an occurrence of 75% (no-go trials 25%). Block's sequences were assigned randomly.

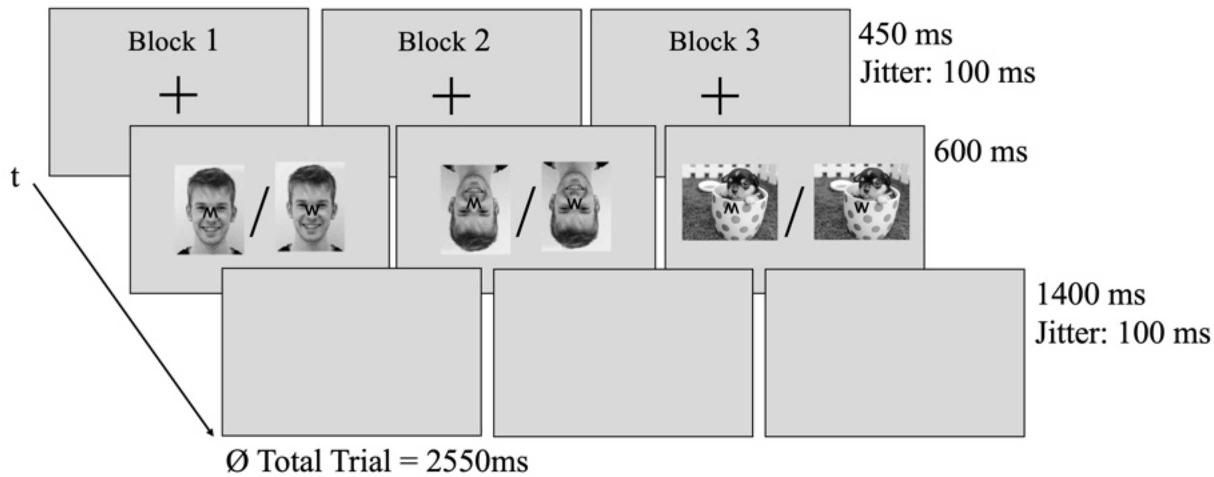


Figure 1. Trial scheme of the go/no-go task. Before the task participants were instructed whether "W" or "M" is a signal for go or no-go (randomly assigned, 75% go trials). Additional information during training trials asked participants to show all expressions as intensely as possible and quickly return to a neutral face. (+) Fixation cross. During the experiment participants saw the pictures in color.

Data Analyses

Video-data processing.

New technology allows to investigate facial expression without hours of manual coding using emFACS-7 (Friesen & Ekman, 1983) or interfering electrodes using electromyography (EMG; Beringer, et al., 2019). With machine vision for the assessment of facial expression video material can be coded efficiently, which is why common studies started to use it more and more often (e.g., Küster, et al., 2020). We analyzed behavioral effects with a machine vison software based on video recordings with FACET (FACET, version 6.1.2667.3, iMotions, 2016).

As there are no standards of data preprocessing for automated software analyses, we pre-registered ours on Open Science Framework (Beringer, 2018) before analyzing data, to maintain transparency and an a priori perspective on the data (Beringer, 2018). In the online supplementary material we give a MATLAB (R2016a, The MathWorks, 2016) example of how the code deals with an experimental condition in more detail[2].

(1) Baseline: we computed individual baselines to control for individual differences in emotionality during neutral state due to individual differences in facial morphology, using a 6 sec interval before the training phase (Olderbak, Hildebrandt, Pinkpank, Sommer, & Wilhelm, 2014). Additionally, we used the mean of the 7 frames (210 ms) before target stimulus was presented and subtracted it from every following evidence score of a given trial controlling for mood effects (e.g. Schmidt-Atzert, Stemmler, & Peper, 2014).

(2) Classification: (a) we defined the *expression onset* using a threshold of evidence scores greater or equal 1, as recommended by software developers to measure clear expressions, for at least 7 frames (210 ms, lower limit for brief expression; Yan, Wu, Liang, Chen, & Fu, 2013) in a trial for all expressions in the same way. Although expressions of fear and disgust come with a generally lower absolute value in evidence scores than expressions of joy, all expressions clearly exceed evidence scores of 1. So we fixed the threshold at 1, as a fixed threshold at 2 or 3 would not affect the classification of hits and errors.

---

[2] https://osf.io/wh6rx/wiki/home/

(b) We classified all trials without any data as *NaNs* (< 0.001% of data); all trials with an expression onset within the first 7 frames after stimulus onset as *too early reactions* (1% of data), because participants reacted before they had seen the stimulus. We classified all trials without an onset in any expression as *omissions* (4% of go trials) and *correct inhibitions* (80% of no-go trials); all trials with an onset in target expression but not in non-target expressions as *hits* (52% of go trials) and *inhibition errors* (10% of no-go trials); all trials with an onset in any non-target expression but not in target expression as *incorrect expressions* (9% of go trials and 6% of no-go trials). (c) For trials with less prototypical expressions, with an onset in more than one expression, we calculated the median of the frames between onset, at which the target expression reached threshold for the first time in a trial and the *offset*, at which evidence scores fell below an evidence score of 1 the first time after onset. If the median of evidence scores in the target expression was greater than the median of evidence scores in non-target expressions, we classified a trial as a *blended hit* (34% of go trials) – a trial showing a blended expression – and as a *blended inhibition error* (4% of no-go trials) when the median of evidence scores in the target expression was equal to or lower than the median of non-target expressions.

(3) RTs and error rates: Visual inspection revealed the presence of blended expressions in some trials in some participants. We therefore considered both as correct answers but differentiated between pure and blended expressions for the hits and reported both as it could be interesting for other researchers. Therefore, we calculated RTs for all prototypical and less prototypical *correct expressions* (hits + blended hits; 86% of go trials) as the difference in time between stimulus onset and expression onset. Subsequently, outlier analyses detected and excluded trials with two or more standard deviations above or below the mean RTs on participant level within condi-

tion (4% of correct expressions). Thus, we calculated error rates of prototypical and less prototypical *errors* (inhibition errors + blended inhibition errors; 14% of no-go trials) as the quotient of errors to all no-go trials for each condition separately.

Electrophysiological recordings and signal processing

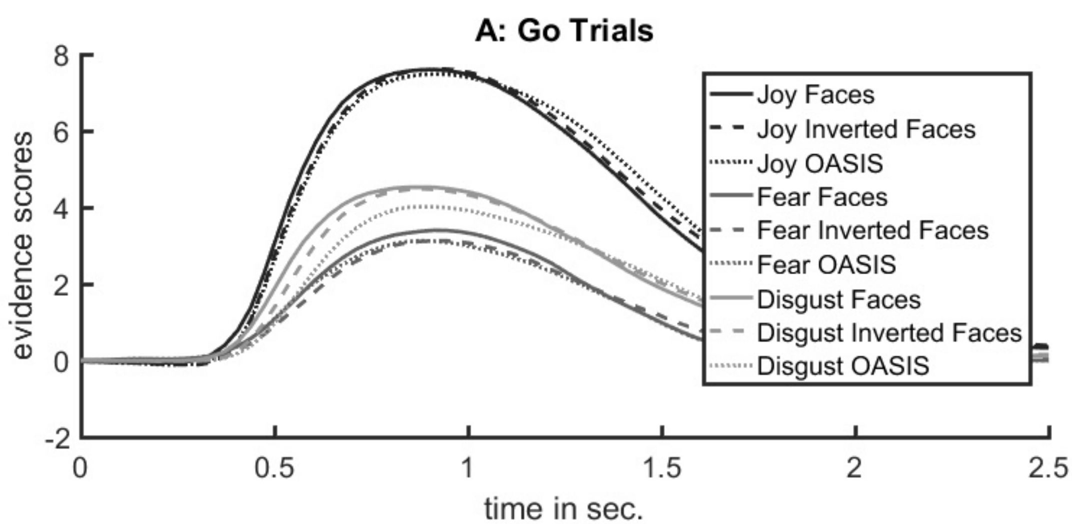
For EEG recordings we used 64 two-wire active pin electrodes (10-20 system) with a BioSemi ActiveTwo Mk2 amplifier, sampled at 2048 Hz with DG-2kHz bandwidth. Raw EEG data was down-sampled to 512 Hz with PolyRex (Kayser, 2003) and processed using Brain Vision Analyzer (version 2.1, Brain Products GmbH, Munich, Germany). Offline, continuous signals were high pass filtered (0.03 Hz) and segmented into 1 sec epochs starting 200 ms before and ending 800 ms after stimulus onset. We considered only correct expressions in go trials and correct inhibitions in no-go trials by exporting event markers of stimulus onsets to MATLAB to add information about results of classification of video data. After reimporting adjusted event markers to Brain Vision Analyzer we removed blinks and muscle artifacts from segmented data by means of independent component analyses. Then, we identified channels with artifacts and excluded all channels that showed an amount of artifacts higher than 2 SDs above the mean of trials marked as having an artifact across participants from further analyses (namely Fp1, Fp2, AF7, T7, T8, TP7, TP8). We interpolated the excluded channels individually for each participant using topographic interpolation. After re-calculation to average reference, we applied a baseline correction for the average activity in the 200ms before stimulus onset. We marked segments with voltage steps larger than 100 µv/ms, amplitude shifts of 200 µv within a period of 200 ms, or amplitudes exceeding ± 200 µv as artifacts (0% of the data) and applied a 30 Hz low-pass filter. Finally, we averaged segments for different conditions separately.

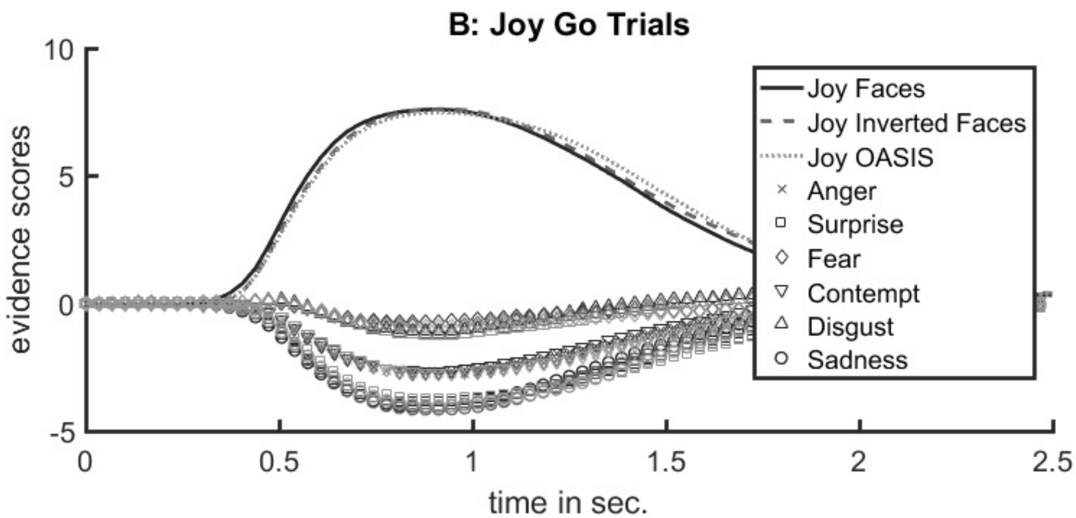
For statistical data analyses we performed analyses of variance for repeated measures (rmANOVA) for all three central measures (RTs, errors, P3 amplitudes) separately. If the first step revealed significant main effects or interactions, we performed post-hoc *t*-tests with Bonferroni corrected *p* values to ascertain which conditions drive the effect.

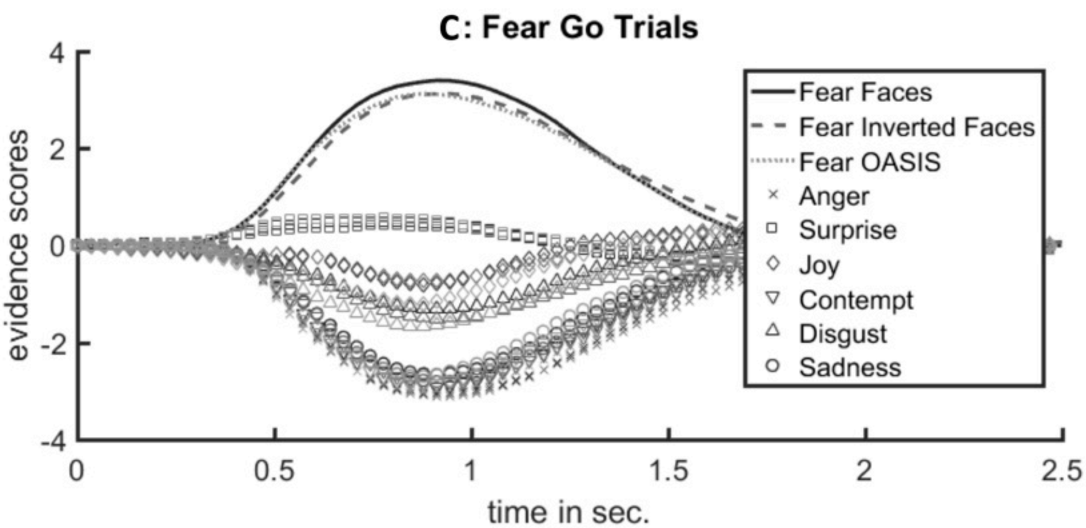
Results Study 1

Behavioral performance

RTs in go trials. Figure 2 shows time course of correct expressions in go trials for different facial expressions in each experimental block. We report *p* values corrected for multiple testing, as well as additional information about number of tests at the end of every section. The rmANOVA over RTs with the factors stimulus type (upright face, inverted face, emotion inducing pictures) and facial expression of stimuli (joy, disgust, fear), revealed a significant main effect for both stimulus type, $F(1,32) = 5.80$, $p = .007$, $\eta^2_p = .27$; and facial expression, $F(1,32) = 17.96$, $p < .001$, $\eta^2_p = .54$; and a significant interaction of both factors, $F(1,32) = 6.40$, $p = .001$, $\eta^2_p = .47$.




B: Joy Go Trials
evidence scores
time in sec.
Joy Faces
Joy Inverted Faces
Joy OASIS
Anger
Surprise
Fear
Contempt
Disgust
Sadness


C: Fear Go Trials
evidence scores
time in sec.
Fear Faces
Fear Inverted Faces
Fear OASIS
Anger
Surprise
Joy
Contempt
Disgust
Sadness

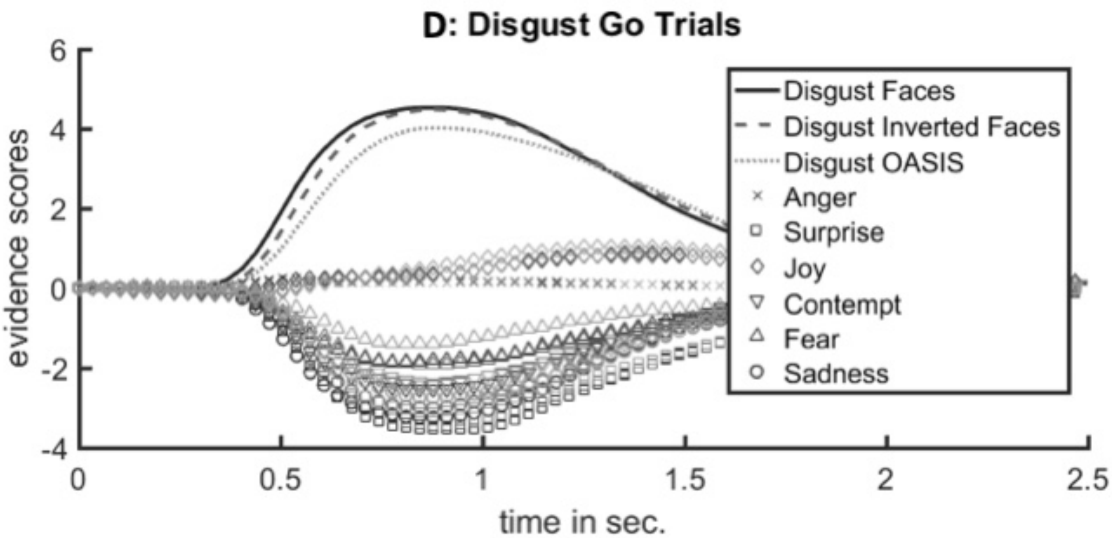


Figure 2. Evidence scores for target expressions (-, --, :) and other expressions (×, □, ◊, ∨, ∧, ○). Time zero refers to the onset of stimulus (upright face, inverted face, emotion inducing pictures without faces from the Open Affective Standardized Image Set (OASIS, Kurdi et al., 2017)).

Post-hoc *t*-tests (*p* value multiplied by number of tests to correct for multiple testing, 3 in total) showed shorter RTs for upright faces than for inverted faces or emotion inducing pictures, $t(32) = -2.83$, $p = .02$, $d = 0.4$; $t(32) = -3.17$, $p < .01$, $d = 0.57$; and no difference in RTs between inverted faces and emotion inducing pictures, $t(32) = -0.05$, $p = 2.88$, $d < 0.01$.

Three additional post-hoc *t*-tests (*p* value multiplied by number of tests to correct for multiple testing, 3 in total) following up on the main effect of expression revealed shorter RTs for showing joy relative to fear or disgust expressions, $t(32) = -5.03$, $p < .01$, $d = 0.94$, and $t(32) = -5.98$, $p < .01$, $d = 1.14$, respectively. The difference in RTs between fear and disgust was not significant $t(32) = -0.83$, $p = 1.23$, $d = 0.15$.

Difference in RTs between facial expressions reflect the activation of different muscles. Therefore, our main interest regarding the significant two-way interaction was the comparison between stimulus types within each expression. For joy expressions the pairwise comparisons revealed shorter RTs for upright faces than for inverted faces. For fear or disgust expressions the same comparison was not significant (see Table 1, *p* value multiplied by number of tests to correct for multiple testing, 18 in total).

RTs for expressions of disgust were significantly longer for emotion inducing pictures relative to upright and inverted faces, $t(32) = -4.08$, $p < .01$, $d = 0.76$ and $t(32) = -3.30$, $p = .04$, $d = 0.64$. The comparison of emotion inducing pictures vs. faces (upright or inverted) was not significant for joy and fear (see Figure 3 and Table 1 for details).

The difference in RT between joy and fear expressions was significant for inverted faces, but not for emotion inducing pictures (see Table 1). Furthermore, we observed shorter RTs for expressions of joy than for showing disgust to inverted faces. This difference was also significant for emotion inducing pictures, but not for upright faces. Finally, we found shorter RTs for showing fear than for showing disgust when viewing emotion inducing pictures, but not when viewing either upright or inverted faces.

Table 1

Post-hoc *t*-tests of interaction effects in RTs

Comparison	*t*(32)	*p*	*d*
Between Stimulus Type Effects within Expression			
joy upright faces – joy inverted faces	-3.19	.05*	0.55
joy upright faces – joy inducing pictures	-1.73	1.62	0.29
joy inverted faces – joy inducing pictures	1.11	4.86	-0.19
fear upright faces – fear inverted faces	-1.75	1.62	0.33
fear upright faces – fear inducing pictures	0.19	15.30	-0.03
fear inverted faces – fear inducing pictures	0.85	1.62	-0.30
disgust upright faces – disgust inverted faces	-2.21	.54	0.38
disgust upright faces – disgust inducing pictures	-4.08	< .01*	0.76
disgust inverted faces – disgust inducing pictures	-3.30	.04*	0.64
Between Expression Effects within Stimulus Type			
joy upright faces – fear upright faces	-5.48	< .01*	1.00
joy upright faces – disgust upright faces	-2.56	.36	0.48
fear upright faces – disgust upright faces	1.50	2.54	-0.27
joy inverted faces – fear inverted faces	-3.92	< .01*	0.77
joy inverted faces – disgust inverted faces	-3.75	.01*	0.69

Comparison	*t*(32)	*p*	*d*
fear inverted faces – disgust inverted faces	1.42	3.06	-0.24
joy inducing pictures – fear inducing pictures	-2.73	.18	0.53
joy inducing pictures – disgust inducing pictures	-7.70	< .01*	1.78
fear inducing pictures – disgust inducing pictures	3.89	< .01*	-0.64

Note: We conducted nine tests to clarify the effect of facial expression given the same stimulus type and nine tests for the effect of stimulus type given the same expression. * Indicates significant tests after correction (*p* values reported here are corrected for multiple testing by multiplying *p* values by number of tests, 18 in total).

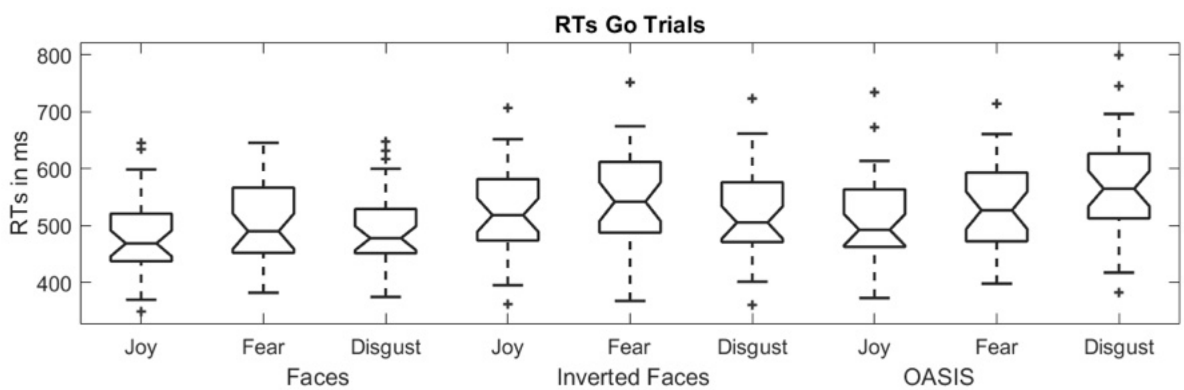


Figure 3. RTs in ms in go trials. OASIS = emotions inducing pictures without faces from the Open Affective Standardized Image Set (OASIS, Kurdi et al., 2017).

Error rates in no-go trials. Figure 4 shows the *error rates* (number of errors divided by number of no-go trials within condition) for different facial expressions. We conducted a rmANOVA with the factors, stimulus type (upright face, inverted face, emotion inducing pictures) and facial expression (joy, disgust, fear). The main effect for the factor stimulus type was not significant, $F(1,32) = 2.70$, $p = .08$, $\eta_p^2 = .15$; the main effect for the factor expression was significant, $F(1,32) = 19.11$, $p < .001$, $\eta_p^2 = .55$; the interaction effect was not significant, *F*(1,32) = 0.38, *p* = .82, η_p^2 = .05. Post-hoc *t*-tests (*p* value multiplied by number of tests to correct for multiple testing, 3 in total) confirmed our prediction of higher error rates for joy than showing either fear or disgust expressions, *t*(32) = 6.27, *p* < .01, *d* = -0.93, and *t*(32) = 3.63, *p* < .01, *d* = -0.61, respectively; and no difference between error rates when showing fear compared with showing disgust, *t*(32) = -1.73, *p* = .27, *d* = 0.38.

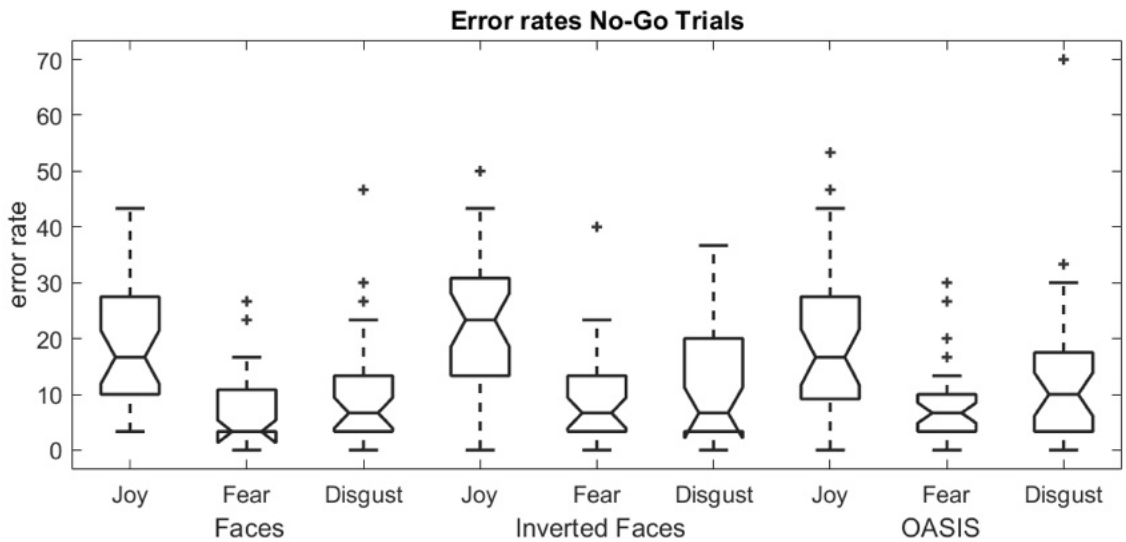


Figure 4. Error rates as percentages of errors in no-go trials. OASIS = emotions inducing pictures without faces from the Open Affective Standardized Image Set (OASIS, Kurdi et al., 2017).

Event related potentials

We computed mean ERP amplitudes of electrode clusters (regions of interest, ROI), following the guidelines published by Keil et al. (2014) and a previous study (Recio & Sommer, 2018). We measured average P3 amplitudes at Fz, Fcz, Fc1, Fc2 and Cz between 430 and 540 ms (e.g., Luck & Kappenmann, 2011). In line with our hypothesis P3 amplitudes were overall larger in no-go trials than in go-trials, *F*(1,32) = 50.15, *p* < .001, η_p^2 = .62. For all further analyses, we computed difference waves (no-go trials – go trials = “no-go P3”, e.g., Falkenstein, Hoormann, and Hohnsbein, 1999) for each expression within stimulus type before averaging amplitudes across all electrodes at the ROI. Figure 5 shows positive components, in the time window of P3 after stimulus onset.

The rmANOVAs over mean ERP amplitudes with the factors stimulus type (upright face, inverted face, emotion inducing pictures) and expression (joy, disgust, fear), revealed a significant main effect for both the factor stimulus type, $F(1,32) = 4.54$, $p = .02$, $\eta_p^2 = .23$, and the factor expression, $F(1,32) = 4.99$, $p = .01$, $\eta_p^2 = .25$, but no significant interaction, $F(1,32) = 1.12$, $p = .37$, $\eta_p^2 = .14$.

Follow up *t*-tests (*p* values reported here are corrected for multiple testing by multiplying *p* values by number of tests, 6 in total), showed larger no-go P3 amplitudes for upright faces than for inverted faces, $t(32) = 2.92$, $p = .04$, $d = -0.49$, and no difference in no-go P3 amplitudes between both faces and inverted faces than for emotion inducing pictures, $t(32) = 2.43$ $p = .12$, $d = -0.40$; $t(32) = -.35$, $p = 4.38$, $d < 0.06$, respectively.

Following up on the effect of expression we performed three post-hoc *t*-tests which showed no difference between no-go P3 amplitudes for expressions of joy and fear, $t(32) = 0.16$, $p = 5.40$, $d = -0.04$; larger no-go P3 amplitudes for expressions of joy than for showing disgust, $t(32) = 2.96$, $p = .04$, $d = -0.57$; and no difference between no-go P3 amplitudes when showing fear compared with showing disgust, $t(32) = 1.64$, $p = .66$, $d = .0.26$.

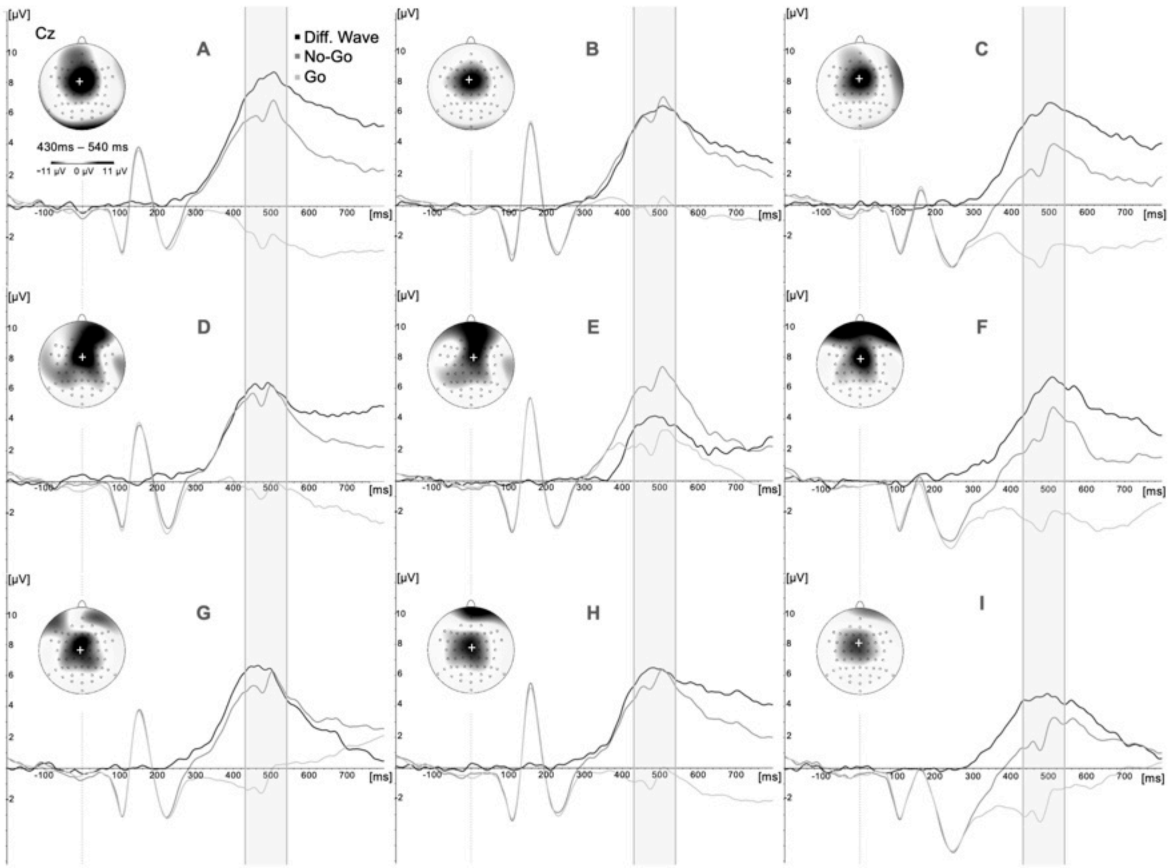


Figure 5. ERPs at a central electrode (Cz) and scalp distribution in the P3 (430 ms – 540 ms, gray marked in time course) after onset of stimulus. Top panels represent the blocks with joy expressions as responses (A joy upright face, B joy inverted face, C joy inducing pictures), middle panels represent the blocks with fear expression as responses (D fear upright face, E fear inverted face, F fear inducing pictures) and bottom panels the blocks with disgust expressions as responses (G disgust upright face, H disgust inverted face, I disgust inducing pictures). Positive voltage is plotted up. In the scalp topographies + marks positive polarity.

Discussion

In Study 1 RTs were in line with our hypotheses (shortest RTs for expressions of joy and upright faces), while error rates differed significantly only for expressions of joy compared to those of fear and disgust in accordance with valence theories of emotion. While error rates probably were not sensitive enough to detect effects of stimulus type, ERP amplitudes over all showed significant differences for the no-go P3 as predicted. These effects were driven by the differences between faces vs. inverted faces and expressions of joy vs. disgust. These findings support the idea of emotion specific effects rather than valence specific effects of facial expressions of emotion, and the specific role of faces as stimuli (e.g., inducing mimicry) compared to inverted faces. For us, the most plausible explanation is the influence of mimicry on inhibition processes when „smiling back". Nonetheless, regarding the no-go P3 we cannot distinguish the inhibition of arousal induced by emotion inducing pictures from mimicry and arousal effects induced by upright faces. Lastly, the special character of Study 1 is the multi-measurement design with machine vision and EEG simultaneously in a laboratory setting. The next step is to investigate these effects during a face-to-face interaction in Study 2.

Study 2

The main research questions of this study were (1) whether the previously observed differences between facial expressions of joy compared to negative expressions of disgust can also be found in a face-to-face interaction, and (2) whether gaze direction impacts the control of facial expressions in a communicative context. We expected shorter RTs for trials with eye contact, as this additional social cue could increase automatic processes to respond promptly, compared to trials without eye contact. We expected more inhibition errors for conditions with a greater need for control, specifically, more errors for joy than disgust expressions, indicating mimicry could be larger for expressions with high affiliative intent. Moreover, we expected more errors when making eye contact compared with averted gaze as social cues must be inhibited in addition.

Methods Study 2

Participants

An a priori G*Power analysis (Faul et al., 2009) suggested a total sample size of 11 participants for a repeated measurement design, estimated $\eta_p^2 = 0.24$ (Adams, & Kleck, 2005), an alpha error probability of 0.05, desired power of 0.80, one group, three measurements, a correlation among repeated measurements of .50, and nonsphericity correction of 0.5. Anticipating some dropouts, we recruited a sample of 19 participants (total sample: 19 healthy women, $M_{age} = 23.4$ years, $SD_{age} = 3.9$), not overlapping with the study 1 sample. In order to make sure that participants' facial expressions could be correctly measured and to maximize comparability between Study 1 and Study 2, we followed the same inclusion and exclusion criteria as in Study 1 and in previous work (Beringer et al., 2019). All participants provided informed consent prior to the study. Psychology students received course credits (79%) and all others 10 Euro (21%) for their contribution. The study was conducted in accordance with the Declaration of Helsinki and was approved by Faculty Ethics Committee.

We excluded one participant who had a number of incorrect expressions in go trials more than 2 *SDs* above the mean across all facial expressions (Leonhart & Lichtenberg, 2009) and one participant, because she reported not to have understood the task. The final test sample consisted of 17 participants ($M_{age} = 23.7$ years, $SD_{age} = 3.9$).

Procedure and Apparatus

We used the same apparatus and similar procedure as in Study 1. After a training phase, (passed by all participants), a female confederate of the experimenter entered the chamber and was introduced to each participant as another participant. The confederate sat opposite to the participant, separated from each other by an electrically controlled glass (Privacy SmartGlass, 27 inches), which can be switched from clear to opaque or vice versa in approximately 0.01 sec (at room temperature; 20°Celsius controlled by air conditioner). The glass was approximately 80 cm from participants' and confederates' eyes. At the beginning, the confederate used the display to read out loud the standardized instructions. Before the experiment the confederate was trained to display the target expressions and reach the apex, using visual feedback from FACET. This training helped the confederate to pose prototypical expressions. Trial sequence was randomized in three possible gaze directions, (a) eye contact, (b) without eye contact looking to the right, (c) without eye contact looking to the left, and two expressions (a) joy, (b) disgust. The monitor was placed on the table facing the confederate, so the participant could not see the instructions for the confederate, while the latter could easily read instructions.

Similar to Study 1 the main task was a go/no-go task, organized in two blocks (expressions of joy and disgust), with a real person instead of standardized pictures on a screen as in Study 1. The confederate prepared her facial expression before each trial while the glass was opaque with help of a display reminding her which condition had to be prepared. She showed the apex of expression when the glass turned transparent (see Figure 6). She made eye contact with the participant in 50% of trials, or she looked to the left (25% of trials) or to the right (25% of trials).

A red LED light placed on the forehead of the confederate indicated the participant go and no-go trials (e.g., on = go and off = no-go; randomly assigned between subjects). Hence, in go trials of block 1 participants expressed joy to a confederate showing an expression of joy and in block 2 they displayed expressions of disgust to the confederate showing an expression of disgust as fast as possible after stimulus onset (i.e., after the glass turned transparent). In no-go trials, participants were instructed not to show any facial expression. The temporal sequence of the events in a trial is presented in Figure 6. To sum up, there were eight different task conditions: (1) in *go joy with eye contact* and (2) *go joy without eye contact* participants saw a confederate with an expression of joy and produced an expression of joy; (3) in *go disgust with eye contact* and (4) *go disgust without eye contact*, participants saw a confederate showing an expression of disgust and produced a disgust expression; (5-8) in no-g*o joy/disgust with/without eye contact*, participants saw a confederate showing an expression of joy or disgust and kept a neutral expression. In every block every expression was requested 180 times, whereby go trials had an occurrence of 66% (no-go trials of 33%) and block's sequences were randomly assigned.

Figure 6: Trial scheme

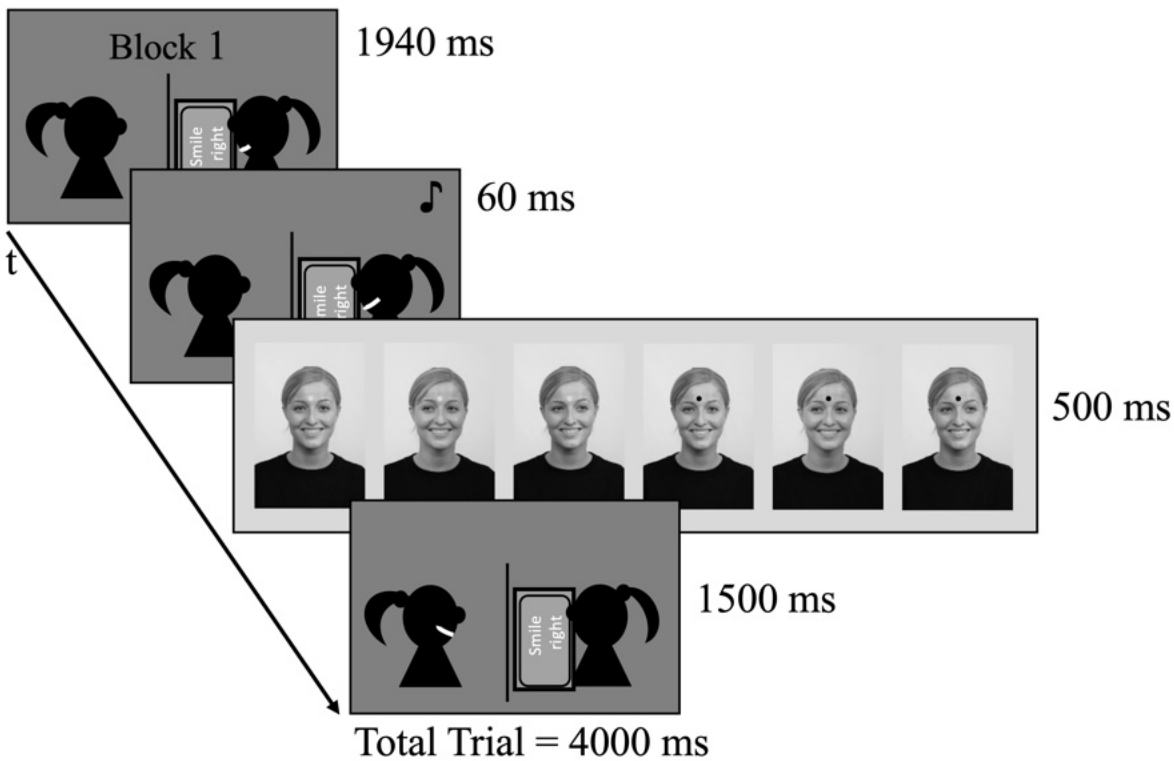


Figure 6. Trial scheme for go/no-go task in the face-to-face situation. Dark grey background indicates the glass between participant and confederate is opaque; light grey indicates that the glass is transparent; (♪) instead of a fixation cross we used a short beep to catch participants' attention; (•, •) go/no-go LED light signal in random order, whereby 66% were go trials. Prior instructions informed participants whether "•" or "•" is a signal for go or no-go (randomly assigned). The light gray box shows all six possible stimuli within the block of joy expressions. Additional information during training trials asks participants to show all expressions as intensely as possible and quickly return to a neutral face. The figure shows block 1 with joy expressions as an example.

Manipulation check

Catch trials were presented to increase participant's attention during the task. In catch trials, the confederate was instructed to show an expression of surprise (18 times within all 360 trials of both blocks, varying randomly within a sequence of 20 trials). All catch trials were go-trials, in which participants should respond with an expression of surprise. That way we examined whether participants continuously encoded the emotion related expression of the confederate or habitually responded to the stimulus with an expression of joy or disgust. All participants included in the analyses solved more than 71% of the catch trails. In a manipulation check after the experimental task 61% of the participants stated to have believed that the confederate was a real participant (M = 2.78, SD = 1.8 on a Likert scale ranging from 0 = “didn’t believe at all” to 5 = “totally believed”).

Data Analyses

Video-data processing.

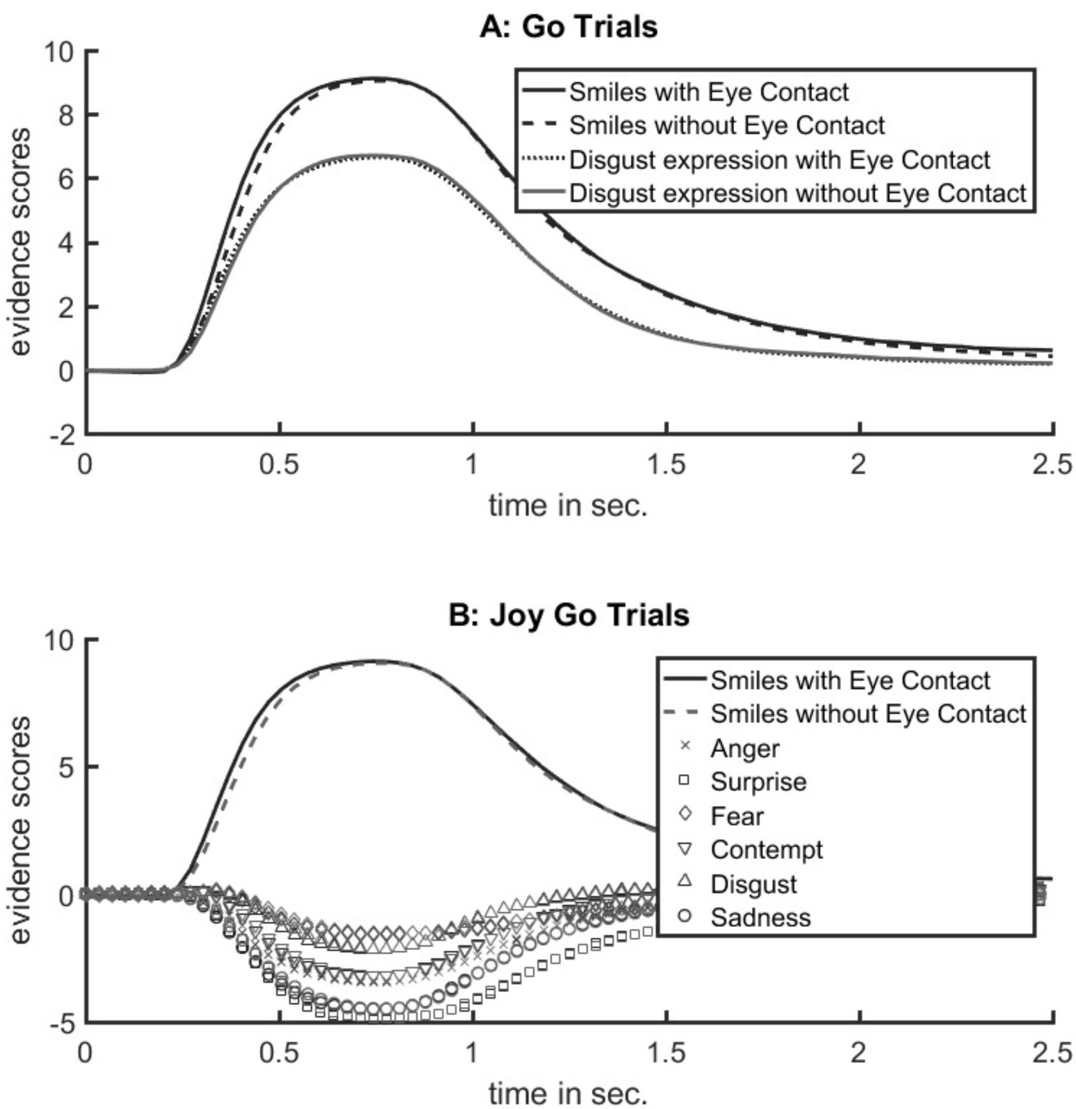
In Study 2 we used the same code to analyze data from facial expressions and video data processing as preregistered for Study 1.[3] The code classified data as *NaNs* (1% of data); *too early reactions* (4% of data), *omissions* (5% of go trials), *correct inhibitions* (55% of no-go trials); *hits* (60% of go trials), *blended hits* (25% of go trials), *inhibition errors* (9% of no-go trials), *blended inhibition errors* (3% of no-go trials); and *incorrect expressions* (6% of go trials and 5% of no-go trials). We calculated RTs and error rates as *correct expressions* (hits + blended hits; 84% of go trials), outlier analyses excluded 4% of correct expressions, *errors rates* (inhibition errors + blended inhibition errors; 12% of no-go trials). rmANOVA were computed to analyze the data statistically (followed by post hoc contrasts with Bonferroni correction, when necessary).

Results Study 2

[3] https://osf.io/wh6rx/wiki/home/

Behavioral performance

RTs in go trials. Figure 7 shows the RTs of correct expressions in go trials for different facial expressions in each experimental block. The rmANOVA over RTs with the factors gaze direction (with eye contact, without eye contact) and expressions (of joy, disgust), revealed significant main effects of gaze direction, $F(1,16) = 10.05$, $p = .006$, $\eta_p^2 = .37$; and expressions, $F(1,16) = 5.32$, $p = .03$, $\eta_p^2 = .24$; but no significant interaction effect, $F(1,16) = 3.87$, $p = .07$, $\eta_p^2 = .19$: RTs for trials with eye contact were shorter than for trials without eye contact and RTs were shorter for joy than for disgust expressions.



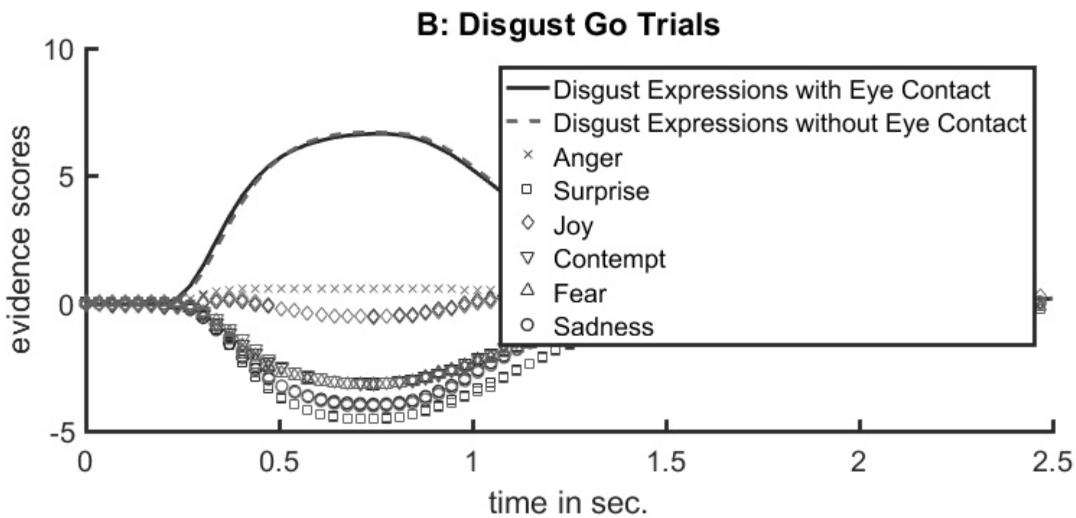


Figure 7. Mean FACET evidence scores for target expressions (-, --, :) and non-target expressions (×, □, ◊, ∨, ∧, ○). Time zero refers to the onset of stimulus.

Error rates in no-go trials. Figure 8 shows the *error rates* (number of errors divided by number of no-go trials within condition) for different facial expressions in each experimental block. The rmANOVA revealed that the main effect for both gaze direction and expression was not significant, $F(1,16) = 4.30$, $p = .06$, $\eta_p^2 = .21$; $F(1,16) = 1.42$, $p = .25$, $\eta_p^2 = .08$, respectively; and a significant interaction effect, $F(1,16) = 10.83$, $p = .005$, $\eta_p^2 = .40$. For a pairwise comparison we corrected *p* values for multiple testing using Bonferroni correction (*p* value multiplied by number of tests, 4 in total). A pairwise comparison revealed a significant effect of gaze direction only for expressions of joy such a way that expressions of joy come with more errors in conditions with eye contact, $t(16) = 4.34$, $p = .004$, $d = 0.08$.

Figure 8. Error rates as percentages of errors in no-go trials.

Discussion

In Study 2 RTs were in line with our hypotheses (shortest RTs for expressions of joy and eye contact), while error rates showed a significant interaction between expression type and eye contact. More errors in no-go trials in conditions with eye contact only for expressions of joy indicate mimicry could be larger for expressions with high affiliative intent like expressions of joy, and reduced mimicry for negative expressions. These findings support the idea of emotion specific effects of facial expressions of emotion, and the specific role of eye contact in face-to-face interactions. The special character of Study 2 is the intersubjective design involving a face-to-face interaction, aiming to approximate a situation of social interaction.

General Discussion

Study 1 investigated differences in deliberate control between facial expressions of three different emotions (joy, disgust and fear), and the impact in facial control of pictures of standardized databases as stimuli (upright faces, inverted faces and emotion inducing pictures without faces), including EEG measurements to probe the neural underpinnings of cognitive processes during the control over facial expressions in a go/no-go task. In Study 2 a person performed facial expressions that served as distractor stimuli in a go/no-go task, aiming to approximate a situation of social interaction. Results of both studies showed worse top-down control for expressions of joy than for disgust. Also, Study 2 revealed that the deliberate control over expressions of joy was affected by gaze direction in the face-to-face setting.

Study 1 is in accordance with our hypothesis and shows shorter RTs in conditions with upright faces as stimuli, compared to inverted faces and emotion inducing pictures. On one hand, perception of faces is a holistic process (Behrmann, Richler, Avidan, & Kimchi, 2015) and induces mimicry (Hatfield, Cacioppo, & Rapson, 1994) thereby speeding encoding and preparation of the production of a compatible expression. On the other hand, inverted faces and emotion inducing pictures launch a detail-oriented visual encoding (Valentine, 1988), which needs more time and does not induce any mimicry, which would automatically prepare an expression (Dimberg et al., 2002). As predicted by our hypothesis on emotion specificity, expressions of joy showed more inhibition errors than expressions of fear and disgust and larger no-go P3 amplitudes for expressions of joy compared to expressions of disgust.

In accordance with our hypothesis in Study 2, RTs were shorter in conditions with eye contact than without eye contact, and shorter for expressions of joy than for expressions of disgust. In addition, a significant interaction effect indicates more errors for expressions with eye contact for expressions of joy. This emphasizes that is harder control an expression of joy when one is smiled at. Again, this points out the role of automatic processes (e.g., mimicry) in the production and control of expressions of joy in a face-to-face situation.

Both studies address our first question whether top-down control of facial expressions interfere with automatic processes. Study 1 showed shorter RTs despite comparable rates of inhibition errors when stimuli are pictures of upright faces, associated with processes like mimicry, compared to inverted faces (inducing less mimicry and less emotional arousal) and pictures without faces (inducing less mimicry but emotional arousal). Notably, we replicated this result in Study 2 with a physically present person, showing the role of automatic face perception processes in conditions with eye contact and without eye contact. These findings can be interpreted as facilitation effects, indicating automaticity, e.g., facilitation of mimicry as shorter RTs in conditions with upright faces compared with inverted faces. On the contrary, it would have been conceivable to find facilitation effects driven by emotional arousal or push factors (especially for disgust expressions) shown as shorter RTs in conditions with emotion inducing pictures compared to upright faces and facilitation for emotional arousal in general in conditions with shorter RTs for emotion inducing pictures than for inverted faces. Study 1 supports our hypothesis of a facilitation effect of mimicry as upright faces came with shorter RTs than inverted faces and emotion inducing pictures. These effects show the relevance of face perception for ones' own facial expressions and ability to react promptly when automatic processes (e.g., facial mimicry) help to produce a response (e.g., Harrison, Morgan, & Critchley, 2010).

Furthermore, larger no-go P3 amplitudes in conditions with upright faces compared to inverted faces indicate higher cognitive cost when top-down control is needed to inhibit induced mimicry which underpins behavioral effects from a neurophysiological perspective and is in line with earlier findings for control of facial expressions (e.g., Recio & Sommer 2018) and hand movements (e.g., Smith, Johnstone & Barry, 2008). However, we expected higher error rates in conditions with upright faces compared to inverted faces, but these effects were not significant.

The absence of an effect for stimulus type in error rates hinders the interpretation of automatic response as the task may have been too easy to produce a sufficient number of inhibition errors for meaningful differences to emerge.

In contrast, inhibition errors are valid for interpretations of expression specificity and in Study 1 expressions of joy came with more errors than expressions of fear and disgust with medium to big effect sizes. Taken together, expressions of joy might be associated with a more liberal response criterion, in terms of Russell's theory and due to the positive arousal (e.g., Russell, 1980). As Hess and Fischer propose, mimicry could be larger for expressions with high affiliative intent like expressions of joy, as opposed to negative expressions. The latter explains why expressions of joy come along with more errors as they tend to induce more mimicry (e.g., Hess, & Fischer, 2013; Hess, & Fischer, 2014). Alternatively, Ekman's' theory of display rules could explain this pattern as negative expressions might be more important to be controlled and therefore come with less inhibition errors (e.g., Ekman, & Friesen, 1975).

Regarding the non-significant interaction of expression and stimulus type, on one hand, the current error rates do not provide empirical support for a larger internal push factor of disgust compared to other emotional expressions, as suggested by appraisal theories (e.g., Scherer, 1992). On the other hand, interpretability of this finding might be limited due to the possibility that the facilitating effects of mimicry (with face stimuli) and a putative pushing factor of disgust (with picture stimuli) may have been of similar magnitude.

Additionally, RTs in Study 2 showed faster reactions for expressions of joy than for expressions of disgust and faster reactions in conditions with eye contact than without eye contact. This addresses our third question and points out the impact of social variables again with an effect of facilitation for positive but less for negative emotional expressions. As both studies reveal similar effect sizes for RTs, these automatic processes might also play a big role in face-to-face situations, in which eye contact occurs, as much as in controlled tasks with standardized stimuli inducing automatic processes like mimicry with pictures of upright faces versus inverted faces.

According to the significant interaction effect in Study 2 there were more errors for expressions of joy with eye contact than without eye contact. This is a very interesting finding as it indicates that social variables like gaze direction affect the performance of expressions of joy but probably not of disgust expressions during social interaction. This fits in with the model by Hess and Fischer (e.g., Hess, & Fischer, 2013; Hess, & Fischer, 2014), which postulates that mimicry is larger for positive affect because in social contexts it indicates affiliative intention. The main finding here is that a social variable (e.g., eye contact) more strongly affects the top-down control of facial expressions of joy than those of disgust in a face-to-face situation. This could be the focus in the further-reaching question of how individuals are influenced differently by different emotions in social contexts and vice versa (e.g., Van Kleef, 2009).

Limitations and future perspectives

Several limitations of the present studies deserve to be mentioned. First, because of the relatively small number of participants and a fixed identity of the stimulus person in Study 2, this might be regarded as somewhat preliminary, illustrating how to investigate deliberate control of facial expressions in a face-to-face situation. Second, the comparability between Study 1 and Study 2 is limited by different no-go rates, a smaller set of investigated facial expressions of emotion, and the presence/absence of EEG recordings. Altogether, this is why we did not perform any statistical comparisons between the two studies. Future research should increase the number of stimulus performers in face-to-face situations, of facial expressions of emotion, and include EEG measurements (possibly even outside the laboratory; see e.g., Aspinall, Mavros, Coyne, & Roe, 2015).

Since our Study 2 used a relatively simple form of social cues, prospectively, it could enhance the external validity of laboratory investigations to transfer paradigms such as the go/no-go task to more complex social situations. Also, we used portraits of posed expressions from standardized databases validated with an external criterion for expression classification and posed expressions from a trained performer. This limits the ecological validity, because it does not reflect the aspect of spontaneous expressions (e.g., ambiguous and less prototypical expressions in different levels of intensity). Thus, additional work is needed to determine whether our results on deliberate control of facial expressions will also generalize to spontaneous expressions of emotion.

In Study 2 we aimed to approximate a situation of social interaction. Having a person physically present while keeping high experimental control is an advantage of the present study. Even though we limited the investigated exchange of expressions of emotion to the go/no-go task it is a crucial benefit of Study 2 to show analogous effects as Study 1. Surely the confederate and the participants had an exchange of expressions of emotion while talking before and after the task, but we did not investigate those parts. Nonetheless, studies have shown that having a person physically present might change the effects observed with portraits of people (e.g., Pönkänen et al., 2010; Risko et al., 2012).

One can argue whether showing visual stimuli and posing facial expressions really induces emotions (e.g., smiling vs. joy, see e.g., Barrett, et al., 2019). Obviously, there must be a lack of experiencing emotions during the hundreds of trials quantitative research needs e.g., for ERP analyses, but we did not measure emotional states of participants and therefore cannot answer this question empirically. Nonetheless, our research shows big differences between expressions of joy and e.g., disgust. Following the idea of Ekman's distinct basic emotion theory these differences can be carefully attributed to differences in the quality of expressions e.g., of joy vs. disgust. Other authors point out, that emotions are expressed in a more complex way in everyday life than in a linear correlation of feeling e.g., joy and smiling (e.g., Barrett, et al., 2019). Barrett and colleagues (2019) provide a detailed list of recommendations how to consume scientific literature about facial expressions of emotion and give a current overview of scientific knowledge being much more cautious about premature conclusions about facial expressions and underlying emotional states.

Furthermore, we used upright, inverted faces and emotion inducing pictures as stimuli that differ in their perceptual processing in a classical go/no-go paradigm in Study 1. The effect of shorter RTs in go trials could be partly driven by differences in perception processes between types of stimulus, as faces are processed more holistically and faster compared to the other visual stimuli we used in Study 1. Nevertheless, inhibition errors in incongruent conditions should be widely uninfluenced by this effect as they refer to the inhibition of prepared reactions considered after perceptual processes. As we computed the no-go P3 (no-go trials – go trials) as it is usually done over all trials independent from the behavioral reaction (e.g., Falkenstein, Hoormann, & Hohnsbein, 1999), effects of no-go P3 might be affected by differences in perceptual processing when participants failed to inhibit their reaction in no-go trials and when reacting correctly to upright faces in go trials. As these questions lead to an open and more fundamental question in the go/no-go paradigm, future research needs to address this methodological question specifically.

In Study 1, we did not investigate a go/no-go N2, which one probably would expect. Typically, stimuli that elicit a prepotent response that needs to be inhibited are associated with enhanced N2. However, recent studies report the absence of go/no-go N2 especially when stimuli base on different perception modalities (e.g., visual and auditory, e.g., Nieuwenhuis, Yeung, & Cohen, 2004). Other researchers argue whether the go/no-go N2 rather reflects conflict monitoring than response inhibition (e.g., Donkers, & Van Boxtel, 2004). Our participants rather had to inhibit a prepotent response than monitor a conflict, as there is no visual conflict in the perception e.g., of an M/W in the face of a smiling person. The visual stimulus rather induced a prepotent motor response associated to a specific emotion. We found very inconsistent N2 data in our former studies, maybe because our stimuli base on two modalities of emotional and visual processing, which is why we decided not to focus on the go/no-go N2 in the present study.

Moreover, future research should investigate the specific role of social cues in the control of facial expressions. Our findings show that paradigms as the go/no-go task, which are commonly used in laboratory settings, can be used to show effects of deliberate control of facial expressions in face-to-face situations. From an information processing perspective, we expect even bigger effects of inhibition costs in form of bigger RT effects and more inhibition errors when a more complex social situation requires more cognitive capacities. From an emotion specificity perspective, we expect expressions of joy to be harder to inhibit (e.g., shorter RTs because of more mimicry and more errors) in complex social situations than expressions of negative emotions. However, our design cannot address the comparison of motor control with vs. without social context. It only can extend our findings of Study 1 to a situation with a real person physically present, which we believe is an important contribution of Study 2. Besides, both the performer and participants in Study 2 were women and therefore the results cannot be generalized to men.

Furthermore, from a technical point of view there are some limitations, as measurement of facial expressions of emotion with machine vision software has not come to a point with published gold standards yet. For example, we never analyzed amplitudes of FACET scores (e.g., peak information), even though visual inspection suggests big effects between types of expression (see figure 2 and figure 7). To the authors' knowledge, no other team of researchers uses this kind of information and published any data about intensity effects measured with machine vision software. Unfortunately, all publications which measure facial expressions with machine vision software, including ours, use part of this information implicitly for thresholding, multiclass classification and calculating RTs and errors. This is a very critical point; RTs could be influenced by different gradients of expression types as technical artifacts. Future research needs a wider discussion about technical understanding of machine vision software and peer reviewed gold standards for analyses with an ongoing discussion as they are formulated e.g., for EEG and EMG.

Finally, participants might engage in different strategies to accomplish the task. For example, some participants could prioritize the perception of the superimposed letters we used to differentiate go from no-go trials and ignore the emotion related content in the background (faces, inverted faces, emotion inducing pictures). Also, individual differences in emotion regulation strategies might affect performance in this task, but this was not the scope of the present study. Since we did not control these variables, we cannot rule out a potential confound. However, the fact that we found behavioral and ERP effects makes it unlikely that a big number of participants used strategies to bypass the task relevant information.

Summing up, two studies suggest that observing different facial expressions and emotional pictures may be associated with differences in constructing and buffering the respective expressions. The current research supports this later intuition by demonstrating that it is rapidly executed when challenged and comes with a high need of cognitive resources, an invest of time to be inhibited and an increasing number of inhibition errors. Using machine vision for the assessment of facial expressions, we found the difference between joy and disgust to be also observable in a laboratory situation involving an actual person displaying emotional expressions as distractors to the task. Besides, automatic mimicry and cognitive processes induced by gaze direction strengthen our disposition to quickly respond, i.e., accelerate reactions but need to be inhibited with a cost of time and cognitive resources, not only in highly controlled laboratory studies but in face-to-face situations, too. Future research is needed to probe whether expressions of joy more than other facial expressions of emotion trigger an automatic process (maybe compared with a gaze and other social variables) that one can hardly control voluntary.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft [grant number RE 3721/2-1 to GR]. The authors thank Kristina Bien, Espen Henken, Jeanne Desbuleux and My Nguyen for their help with data collection and analyses. We verified license agreements of iMotions before publishing the present study.

Competing interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplemental Material

Stimuli and manipulation check in Study 1

In order to select stimuli, we conducted a manipulation check with an independent sample, before conducting the experiment mentioned above (N = 86, 80% women, M_{age} = 24.0 years, SD_{age} = 7.0). We presented 230 pictures of OASIS (Kurdi, Lozano, & Banaji, 2017) in an online study using UniPark (Questback, 2016). All participants provided informed consent and received course credits (100% psychology students) for their contribution. Similar to the images from the international affective picture system, IAPS (Lang & Bradley, 2007), pictures of OASIS are rated on scales based on the circumplex model of affect (Russell, 1980). We selected pictures without faces to induce distinct emotional states (joy = high arousal + high valence; fear = high arousal + low valence; disgust = low arousal + low valence). Participants watched pictures for 4 s (alone in a quiet room) and rated all pictures instantly on pseudo-continuous scales ranging from 0-100, one scale for each basic emotion (joy, surprise, anger, disgust, sadness, contempt, fear) and a neutral scale, whereby participants divided 100 points on the eight scales (controlled by backend) – all pictures were randomly assigned. Hence, a single picture rating of a single participant resulted in a vector with eight dimensions (e.g., (100|0|0|0|0|0|0|0) for an unambiguous joyful picture (e.g., little penguin), or (70|17|0|0|0|0|0|13) for an ambiguous joyful, surprising and fearful picture (e.g., skydiver). To check participants' attention, we randomly assigned 24 catch trials, in which they watched a picture as usual but instead of rating it afterwards on emotion scales, they had to answer open field questions about the content (e.g., "What was the color of the left bird in the picture before?").

To retain good data quality, we excluded three participants who did not pass more than 80% of the catch trials. Also, we excluded twelve participants with a working time more than 3 SDs above the mean (test sample: N = 71, 80% women, M_{age} = 23.7 years, SD_{age} = 5.9).

The matrix with data from ratings consisted of 130640 data points (71 participants x 230 pictures x 8 scales), 98.7% of them remained after an outlier detection (i.e., ratings of a picture more than 2 SDs under or above the mean ratings of a given emotion scale, for example, one participant reported, she rated a picture of a bike highly on sadness, because her bike was stolen the week before). We used the mean vector information across participants of every picture in an Euclidean vector space (orthogonal axes: joy, surprise, anger, disgust, sadness, contempt, fear, neutral), where similar pictures are close, and different pictures are far from each other (see clustering methods, Kaufmann & Rousseeuw, 1990). Visually you can see clouds of data points (mean vector information per picture) close to the joy axis, disgust axis, fear axis and neutral axis (see 3D plots online[4]). The distance from such a mean vector information to an axis, namely the Euclidean distance, can be interpreted as the extent to which a picture is inducing distinct emotion. We plotted this distance in ascending order in a scree plot and chose all pictures on the left of the first sharp bend, for distinct emotion induction without faces (29 different pictures for joy, 28 for disgust and 30 for fear).

For emotion induction with faces, we used the pictures from the Amsterdam Dynamic Facial Expression Set, ADFES (van der Schalk, Hawk, Fischer, & Doosje, 2011), showing expressions of joy, disgust and fear of 20 different models, trained to show prototypical expressions.

[4] https://emotion.app.baqend.com/v1/file/www/index.html, https://emotion.app.baqend.com/v1/file/www/index2.html?BCB

RTs and errors (Study 1)

Table 2

Condition	RT in ms, M (SD)	error rate, M (SD)
upright faces - joy	473.15 (71.25)	17.30 (11.83)
upright faces - fear	522.40 (78.34)	7.12 (8.61)
upright faces - disgust	507.81 (85.81)	9.19 (10.52)
inverted faces - joy	499.92 (71.53)	21.08 (11.86)
inverted faces - fear	548.08 (84.46)	8.02 (8.22)
inverted faces - disgust	534.74 (73.85)	9.91 (9.57)
inducing pictures - joy	488.11 (69.00)	17.75 (12.89)
inducing pictures - fear	519.39 (82.62)	7.66 (7.49)
inducing pictures - disgust	561.57 (89.98)	10.90 (10.47)

A graphic presentation of RTs and errors can be seen in figure 3 and figure 4.

References

Achaibou, A., Pourtois, G., Schwartz, S., & Vuilleumier, P. (2008). Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia, 46*(4), 1104-1113.

Adams Jr, R. B., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, *5*(1), 3.

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1*(1), 21-62.

Aspinall, P., Mavros, P., Coyne, R., & Roe, J. (2015). The urban brain: analysing outdoor physical activity with mobile EEG. *Br J Sports Med*, *49*(4), 272-276.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest, 20*(1), 1-68.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition, 104*(3), 644-653.

Behrmann, M., Richler, J. J., Avidan, G., & Kimchi, R. (2015). Holistic face perception. *Oxford handbook of perceptual organization,* 758-774.

Beringer, M., Spohn, F., Hildebrandt, A., Wacker, J., & Recio, G. (2019). Reliability and validity of machine vision for the assessment of facial expressions. *Cognitive Systems Research*, *56*, 119-132.

Beringer, M. (2018, March 14). Go-NoGo Facial Expression. Retrieved from osf.io/wh6rx.

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological Psychology, 77*(3), 343-352.

Brass, M., Zysset, S., & von Cramon, D. Y. (2001). The inhibition of imitative response tendencies. *Neuroimage, 14*(6), 1416-1423.

Bruce, V. (2017). *Recognising faces*: Routledge.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology, 50*(1), 191-214.

Cannon, P. R., Hayes, A. E., & Tipper, S. P. (2009). An electromyographic investigation of the impact of task relevance on facial mimicry. *Cognition and Emotion, 23*(5), 918-929.

Davidson, R. J. (1995). Cerebral asymmetry, emotion, and affective style In Davidson RJ, editor; & Hugdahl K., editor.(Eds.), *Brain asymmetry,* 361-387.

Darwin, C. (1998). *The expression of the emotions in man and animals*: Oxford University Press, USA.

Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*, 135-168.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*(1), 86-89.

Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition & Emotion, 16*(4), 449-471.

Donkers, F. C., & Van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and cognition, 56*(2), 165-176.

Ekman, P. (1971). *Universals and cultural differences in facial expressions of emotion.* Paper presented at the Nebraska symposium on motivation.

Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion,* 98(45-60), 16.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology, 17*(2), 124.

Ekman, P., & Friesen, W. V. (1975). Unmasking the face Englewood Cliffs. *Spectrum-Prentice Hall*, New Jersey.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of nonverbal behavior, 6*(4), 238-252.

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*(3875), 86-88.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin, 128*(2), 203.

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta psychologica, 101*(2-3), 267-291.

Fallgatter, A., Brandeis, D., & Strik, W. (1997). A robust assessment of the NoGo-anteriorisation of P300 microstates in a cued Continuous Performance Test. *Brain Topography, 9*(4), 295-302.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160.

Fernández-Dols, J. M., & Ruiz-Belda, M.-A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. *The Psychology of Facial Expression, 2*, 255-274.

Fischer, A. H., & Van Kleef, G. A. (2010). Where have all the people gone? A plea for including social interaction in emotion research. *Emotion Review*, *2*(3), 208-211.

Friesen, W. V., & Ekman, P. (1983). EMFACS-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco, 2*(36), 1.

Frijda, N. H., Scherer, K. R., & Sander, D. (2009). Emotion definitions (psychological perspectives). *The Oxford companion to emotion and the affective sciences*, 142-144.

Gray, J. R. (2004). Integration of emotion and cognitive control. *Current Directions in Psychological Science*, *13*, 46–48.

Harrison, N. A., Morgan, R., & Critchley, H. D. (2010). From facial mimicry to emotional empathy: A role for norepinephrine? *Social Neuroscience, 5*(4), 393-400.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). Emotional contagion: Cambridge studies in emotion and social interaction. *Cambridge, UK: Cambridge University Press. errors-in-variables regression model when the variances of the measurement errors vary between the observations. Statistics in Medicine, 21*, 1089-1101.

Hayes, J. G., & Metts, S. (2008). Managing the expression of emotion. *Western Journal of Communication, 72*(4), 374-396.

Hefer, C., & Dreisbach, G. (2020). The volatile nature of positive affect effects: opposite effects of positive affect and time on task on proactive control. *Psychological research, 84*(3), 774-783.

Hess, U., & Bourgeois, P. (2010). You smile–I smile: Emotion expression in social interaction. *Biological Psychology, 84*(3), 514-520.

Hess, U., & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and social psychology review, 17*(2), 142-157.

Hess, U., & Fischer, A. (2014). Emotional mimicry: Why and when we mimic emotions. *Social and personality psychology compass, 8(2)*, 45-57.

Kaufman, L., & Rousseeuw, P. (1990). *Finding groups in data*. Hoboken: NJ: John Wiley & Sons, Inc.

Küster, D., Krumhuber, E. G., Steinert, L., Ahuja, A., Baker, M., & Schultz, T. (2020). Opportunities and challenges for using automatic human affect analysis in consumer research. *Frontiers in Neuroscience, 14,* 400.

Kayser, J. (2003). Polygraphic Recording Data Exchange-PolyRex [http://psychophysiology. cpmc. columbia. edu/PolyRex. htm]. Department of Biopsychology. *New York State Psychiatric Institute*.

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., . . . Yee, C. M. (2014). Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology, 51*(1), 1-21.

Knapp, M. L., & Daly, J. A. (2002). *Handbook of interpersonal communication*: Sage.

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition, 4*(3), 265-273.

Korb, S., Grandjean, D., & Scherer, K. (2008). Motor commands of facial expressions: the bereitschaftspotential of posed smiles. *Brain Topography, 20*(4), 232-238.

Korb, S., Grandjean, D., & Scherer, K. R. (2010). Timing and voluntary suppression of facial mimicry to smiling faces in a Go/NoGo task—An EMG study. *Biological Psychology*, *85*(2), 347-349.

Krumhuber, E., & Kappas, A. (2005). Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior, 29*(1), 3-24.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods, 49*(2), 457-470.

Lang, P., & Bradley, M. M. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. *Handbook of Emotion Elicitation and Assessment, 29,* 70-73.

Leonhart, R., & Lichtenberg, S. (2009). *Lehrbuch Statistik*: H. Huber.

Luck, S. J., & Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components*: Oxford university press.

Manstead, A. S., & Fischer, A. H. (2001). Social appraisal. *Appraisal Processes in Emotion: Theory, Methods, Research*, 221-232.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion, 14*(3), 195-214.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annu. Rev. Psychol., 59*, 507-536.

Mehrabian, A. (2017). *Nonverbal communication*: Routledge.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion, 7*(2), 447.

Morecraft, R. J., Stilwell–Morecraft, K. S., & Rossing, W. R. (2004). The motor cortex and facial expression:: New insights from neuroscience. *The Neurologist, 10*(5), 235-249.

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes-Ker, A. H. (2000). Emotional state and the detection of change in facial expression of emotion. *European Journal of Social Psychology, 30*(2), 211-222.

Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2004). Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiology, 41*(1), 157-160.

Olderbak, S., Hildebrandt, A., Pinkpank, T., Sommer, W., & Wilhelm, O. (2014). Psychometric challenges and proposed solutions when scoring facial emotion expression codes. *Behavior Research Methods, 46*(4), 992-1006.

Pessoa, L., Padmala, S., Kenzer, A., & Bauer, A. (2012). Interactions between cognition and emotion during response inhibition. *Emotion*, *12*, 192-197.

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2010). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience, 6*(4), 486-494.

Postma, E. O., & Postma-Nilsenová, M. (2016). *Measuring the Causal Dynamics of Facial Interaction with Convergent Cross Mapping.* Paper presented at the CogSci.

Questback. (2016). EFS Survey, Version Fall 2016. . *Köln: Questback GmbH.*

Recio, G., Shmuilovich, O., & Sommer, W. (2014). Should I smile or should I frown? An ERP study on the voluntary control of emotion‐related facial expressions. *Psychophysiology, 51*(8), 789-799.

Recio, G., & Sommer, W. (2018). Copycat of dynamic facial expressions: Superior volitional motor control for expressions of disgust. *Neuropsychologia, 119,* 512-523.

Risko, E. F., Laidlaw, K. E., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6*, 143.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161.

Russell, J. A., & Fernández-Dols, J. M. (1997). What does a facial expression mean. *The Psychology of Facial Expression,* 3-30.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11*(3), 273-294.

Scherer, K. R. (1988). On the symbolic functions of vocal affect expression. *Journal of Language and Social Psychology, 7*(2), 79-100.

Scherer, K. R. (1992). *What does facial expression express?* Paper presented at the Parts of the argument in this chapter have been presented at the 36th congress of the German Society of Psychology in Berlin, 1988.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23*(7), 1307-1351.

Schmidt-Atzert, L., Stemmler, G., & Peper, M. (2014). *Emotionspsychologie: ein Lehrbuch*: Kohlhammer Verlag.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 116*(S33), 3-24.

Schulz, K. P., Fan, J., Magidina, O., Marks, D. J., Hahn, B., & Halperin, J. M. (2007). Does the emotional go/no-go task really measure behavioral inhibition? Convergence with measures on a non-emotional analog. *Archives of Clinical Neuropsychology, 22*(2), 151-160.

Smith, J. L., Johnstone, S. J., & Barry, R. J. (2008). Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. *Clinical neurophysiology, 119(3),* 704-714.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology, 38*(2), 343-358.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research, 168*(3), 242-249.

Valentine, T. (1988). Upside - down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*(4), 471-491.

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion, 11*(4), 907-920. doi:10.1037/a0023853

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current directions in psychological science*, *18*(3), 184-188.

Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., & Fu, X. (2013). How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior, 37*(4), 217-230.

2.3 Publication 3: Beringer, M., Wacker, J., & Recio, G. (under review). Interferences between facial expressions of joy and disgust depending on facial attractiveness. An ERP Study. *Journal of Psychophysiology.*

I will now present the third publication of our research project. In this publication we explore how facial expressions of joy and disgust interfere with each other and how facial attractiveness influences this interaction. This study addresses several key objectives of our project. Specifically, Objective 2 investigates the interplay between emotion and executive control over facial expressions. Objective 3 examines the automaticity of facial expressions, while Objective 4 evaluates how social context variables affect the control of facial expressions, specifically whether social motives can enhance or hinder this control.

Interferences between facial expressions of joy and disgust depending on facial attractiveness. An ERP Study

M. Beringer[a], J. Wacker[a], G. Recio[b]

[a] Universität Hamburg Von-Melle-Park 5, D-20146 Hamburg, Germany

[b] Departament de Psicologia Clínica i Psicobiologia, Institut de Neurociències, Universitat de Barcelona, Ps de la Vall d'Hebron, 171, 08035 Barcelona, Spain

m.q.beringer@gmail.com,

jan.wacker@uni-hamburg.de, guillermo.recio@ub.edu

Funding: This work was supported by the Deutsche Forschungsgemeinschaft [grant number RE 3721/2-1 to GR].

Corresponding author:
Matthias Beringer
Dillstraße 1
D-20146 Hamburg, Germany
m.q.beringer@gmail.com
Tel: +49 (0)176 303 116 77

Abstract

The display of facial expressions requires sophisticated control processes, depending on the intentions of the actor and demands of the social situation. *Simulating an online dating situation, we asked participants to rate the attractiveness of pictures of women, to study possible moderation effects of perceived attractiveness on producing facial expressions.* Reaction times (RTs) showed facilitation effects for both deliberate expressions of joy and disgust by congruent expressions in the pictures; however, interference effects due to incongruent emotional expression only occurred in RTs for expressions of joy. Analyses of the moderating effect of facial attractiveness revealed significantly lower accuracy when participants expressed disgust to a smiling stimulus perceived as attractive, while no effect was found for RTs. Analyses of ERP data showed an early posterior negativity (EPN), indicating better sensory processing for expressions of joy rather than neutral facial expressions. ERP data also revealed a more negative N2 associated with conflict detection emerging from the mismatch between participants' deliberate facial expression and the one displayed by the stimulus face. Our findings reflect automatic tendencies to imitate facial expressions of joy and disgust and its modulation by facial attractiveness.

Keywords: facial expression, emotional expression interference, Stroop task, automatic facial expression recognition, facial mimicry

Interferences between facial expressions of joy and disgust depending on facial attractiveness. An ERP Study

In our daily life we use facial expressions of emotion for communication in social interactions. Smiles often express joy and affiliative intent, whereas expressions of disgust may show aversion. But can we deliberately control our facial expressions if we are in a situation that pushes us to smile, e.g., when someone smiles at us, but we do not want to smile back in order to not show affiliation? In situations like this, emotion and cognitive control systems are integrated and mutually modulate the supporting neural mechanisms, e.g., by facilitating the whole system into an appropriate control state matching the demands of a given situation (e.g., Gray, 2004). However, many questions remain unclear, in particular, how the integration process of emotion and cognition occurs. I.e., for example, which are the moderating factors that facilitate and hamper the production and control of facial expressions?

With the current study we investigated whether facial mimicry, elicited by facial expressions of women in a simulated dating situation, facilitates or interferes with the control over deliberate facial expressions of joy and disgust in heterosexual men. We used machine vision to measure facial expressions, combined with EEG to provide evidence concerning neural correlates, which we explain in more detail below.

Facial mimicry is defined as the tendency to mimic the specific facial movements of a facial expression (e.g., Chartrand & van Baaren, 2009). This process is automatic, that is, it occurs without deliberate control. It also is unconscious, appears for subliminal stimulus presentation and has been observed in laboratory and naturalistic settings even when participants are told not to react (e.g., Chartrand & Bargh, 1999; Heyes, 2011; Dimberg, Thunberg, & Elmehed, 2000; Dimberg, Thunberg, & Grunedal, 2002). Facial mimicry takes place shortly after stimulus presentation and is robust over several successive trials, even when stimulus expressions are task irrelevant (e.g., Harrison, Morgan, & Critchley, 2010).

Prior research has shown that social context influences goal-directed mimicry of emotions through cognitive control processes associated with frontal areas of the brain (e.g., Bourgeois & Hess, 2008; Lee, Joseph, Dolan & Critchley, 2006). Hence, the mere observation of a smile evokes a tendency to smile back, which needs to be inhibited when someone exerts deliberate control not to do so (Cracco et al., 2018). Inhibition is a key element for deliberate control of inappropriate expressions and one aspect of the framework of executive functions, conceptualized as a wide range of cognitive top-down processes (e.g., Diamond, 2013).

The neural basis of facial mimicry is suggested to involve mirror-neuron circuits, located in the right inferior frontal cortex and activated by observed actions that are part of the observer's own motor repertoire, for instance, during imitation of specific positive and negative emotional expressions (Rizzolatti & Craighero, 2004; Lee et al., 2006). Although, in humans the activation of mirror neuron circuits in mimicry can only be indirectly observed (e.g., Rizzolatti & Craighero, 2004), the mirror-neuron system is presumably involved in all imitative behavior including facial mimicry and held to be the neural substrate of the link between perception and action (e.g., Heyes, 2011). From a neurobiological perspective, in overcoming mimicry participants need to handle different processes: first, diverting attention from irrelevant stimulus expressions as much as possible and, second, inhibiting automated mimicry processes that likely occur nonetheless to some degree. To be precise, the targeted processes here are attentional control, understood as resistance to distractor interference, and the inhibition of automatic responses (see e.g., Friedman & Miyake, 2004).

Cognitive processes like resistance to distractor interference can be reflected in the amplitudes and latencies of event-related brain potentials (ERPs). The N2 component of the ERP has been associated with conflict detection (e.g., Larson, Clayson, & Clawson, 2014). Hence, the greater need of top-down control (resistance to distractor interference) from the frontal cortex in incongruent conditions should be reflected in more negative N2 amplitudes, as they come with a stronger conflict (e.g. frowning rather than smiling back at a smiling face). Furthermore, the early posterior negativity component (EPN), typically elicited by emotional content after 200 to 300 ms at occipital sites, is interpreted as enhanced sensory processing, associated with reflexive attention for emotional stimuli (Hajcak, Weinberg, MacNamara, & Foti, 2011). Taken together, the N2 marks cognitive conflict when a distractor needs to be inhibited and the EPN marks whether this is due to the emotional content or simply an artifact of visual perception.

To investigate the deliberate control of facial expressions of emotion, previous studies used paradigms such as response priming (e.g., Recio, Shmuilovich, & Sommer, 2014; Beringer, et al., 2019), go/no-go tasks (e.g., Beringer, Wacker, & Recio, 2022; Korb, Grandjean, & Scherer, 2010) and Stroop-like tasks (e.g., Recio et al., 2022). In this study we used the *emotional expression interference task* (Recio et al., 2022), a modification of the Stroop task (Stroop, 1935). In congruent trials participants were asked to produce deliberate expressions of joy while a smiling stimulus face was seen or to produce deliberate expressions of disgust while a disgust showing face was seen. In incongruent trials, participants were instructed to show deliberate expressions of disgust to smiling stimuli and deliberate expressions of joy to disgust-expressing faces. By presenting participants with pictures of faces showing emotional expressions, we intended to elicit automatic facial mimicry, which should facilitate a congruent response in congruent trials, but should interfere with performance in incongruent trials (e.g., Recio et al., 2022).

In the present study, we also investigated whether facial attractiveness impacts the deliberate control of facial expressions. Social factors (i.e., affiliative intent) also impact mimicry in a complex way (Bourgeois, & Hess, 2008; Künecke, Wilhelm & Sommer, 2017). We therefore expected an impact of attractiveness on mimicry. Although facial attractiveness seems to be important for both men and women when choosing dating partners (e.g., Asendorpf et al., 2011), some studies showed that attractiveness seems somewhat more important for men than for women (e.g., Buss, 2006). Aiming to maximize potential effects in this initial study, we investigated resistance to distractor interference in heterosexual men responding to female faces, and probed effects of subjective facial attractiveness as perceived by every participant.

Bringing facial expression of emotion and facial attractiveness together, recent research supports the idea that high facial attractiveness seems to be an approach signal related to perceived fitness and healthy genes (e.g., Tatarunaite et al., 2005) and is frequently responded to with expressions of joy. However, low facial attractiveness seems to be related with pathogens and activates the levator labii superioris, a muscle involved in expressions of disgust (e.g., Principe & Langlois, 2011). This behavior seems to be part of nonverbal communication and intuitively makes sense, especially for dating when first sight initiates approach or avoidance. Assuming that social rejection represents a special form of avoidance, a facial expression of disgust would be a congruent response in the absence of affiliative intentions. To the authors knowledge, this idea has not been tested so far and the current study is the first to investigate effects of perceived attractiveness on the inhibition of facial mimicry.

In the present study, in some conditions we created an interference and in others a facilitation between an observed facial expression (i.e., joy and disgust) in a set of pictures and the participant's deliberate facial expressions of joy and disgust. As Bourgeois and Hess (2008) point out, that emotional facial expressions not only signal emotional states, but also affiliative intent. Participants who show expressions of joy are perceived as highly affiliative. Thus, the social signal value of emotional facial expressions interacts with the main function of mimicry, to create affiliation, in a way e.g., hand signals would not. It is impossible to study emotional mimicry without considering the specific emotion expression at hand. This implies that one would not expect all emotion expressions to be mimicked equally and independent of social context (Bourgeois & Hess, 2008). High attractiveness should create affiliative intentions in the observer (e.g., Lakin & Chartrand, 2003). Therefore, high attractiveness should facilitate congruent smiles and increase interference for incongruent smiles as compared to low attractive faces. Conversely, low attractiveness should facilitate congruent expressions of disgust and increase interference for incongruent disgust expressions as compared to high attractive faces. Hence, induced mimicry should result in more interference (i.e., longer RTs and lower accuracy) when not smiling back and producing a deliberate expression of disgust to the picture of a woman perceived as attractive, compared to producing an expression of disgust to a smiling stimulus perceived as less attractive. Besides, induced mimicry should result in more facilitation (i.e., shorter RTs and higher accuracy) when smiling back to the picture of a woman perceived as attractive, compared to smiling back to a stimulus perceived as less attractive.

As incongruent conditions involve more top-down control, resulting in longer RTs, the conflict between observed and produced deliberate facial expressions should elicit more negative N2 amplitudes than congruent reactions. We expected stronger conflicts, reflected in more negative N2 amplitudes, when participants had to display a deliberate expression of disgust to attractive than to less attractive faces displaying an expression of joy. Similarly, there should be more conflict (larger N2) for deliberate expression of disgust to faces expressing joy than to faces expressing disgust. We expected larger EPN for conditions with stimuli showing expressions of joy or disgust compared to neutral expressions as an emotion specific arousal/mimicry needs to be inhibited in addition to a purely cognitive conflict.

To sum up, we expected (1) *emotional expression facilitation effects*, that is, shorter RTs and more accurate deliberate expressions for our all-male participants in congruent conditions e.g., when smiling to a picture of a smiling woman perceived as attractive or when showing disgust to a picture of a woman expressing disgust and perceived as less attractive, compared to neutral or scrambled stimuli. Additionally, we expected (2) *emotional expression interference effects*, that is longer RTs, less accurate deliberate expressions, greater N2 amplitudes and EPN components reflecting a cognitive conflict in incongruent conditions (e.g., when showing disgust to a picture of a smiling woman perceived as attractive).

Methods

Participants

A priori G*Power analysis (Faul, Erdfelder, Buchner, & Lang, 2009) suggested a total sample size of 20 participants for a repeated measurement design and an effect size of $f = 0.3$, estimated based on a previous study using a similar task (Korb et al., 2010), an alpha error probability of 0.05, desired power of 0.80, one group, three measurements, a correlation among repeated measurements of .70, and nonsphericity correction of 0.5. To account for typical loss of data due to poor task performance and to increase power for testing interactions, we recruited a sample of 30 men. All participants but one reported to be “exclusively heterosexual” and one participant reported being “predominantly heterosexual/only incidentally homosexual” on the Kinsey Scale of Sexual Orientation (Kinsey, Pomeroy, Martin, & Sloan, 1948). In order to make sure that participants' deliberate facial expressions could be correctly measured without any visual barriers for automated assessment tools of facial expressions, we followed the same inclusion and exclusion criteria as in a previous study on the reliability and validity of machine vision to measure facial expressions (Beringer et al., 2019), excluding participants with full beard or glasses with more than ±1.00 diopters (contact lenses were acceptable). All participants provided written informed consent. Psychology students (13%) received course credits and all others 20 Euro (87%) for their participation. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the German Psychological Society (protocol number GR 112016_amd_082014).

To retain good data quality we excluded four participants who had more false positive responses than 2 *SDs* above the mean (Leonhart & Lichtenberg, 2009). These participants were excluded because they showed very poor performance accuracy posing facial expressions (i.e., less than 2 standard deviations below the average participant, < 62.3 % hits). Subsequently, we identified and excluded all trials with RTs shorter or longer than 2 SDs below or above the individual average, respectively, in a given condition. The final test sample consisted of 26 participants (M_{age} = 24.11 years, SD_{age} = 3.97; relationship status single = 65.4%, long-term relationship = 30.8%, open relationship = 3.8%).

Procedure and Apparatus

After signing consent forms, participants were seated on a fixed chair in a quiet and electromagnetically shielded chamber. Subsequently, they read instructions on a 21-inch LCD display (75 Hz refresh rate), approximately 80 cm from their eyes. Indirect light from three LED strips (40-60cm length) homogeneously illuminated participants' faces. For video recordings we used a Logitech HD Pro Webcam C920 with a sample rate of 30 frames per second, fixed at the bottom of the monitor. We used the software FACET (version 6.1.2667.3, iMotions, 2016) to analyze videos of participants' facial expressions. The software provides results of the video analyses frame by frame as *evidence scores* for seven facial expressions (joy, surprise, anger, disgust, sadness, contempt, fear) on a decadic logarithmic scale (e.g., an evidence score of zero for joy indicates that in this frame it is equally likely, that the targeted face shows joy or no joy). Evidence scores can be transformed for each expression to the probability (*P*) using the formula: $P = 1/1+10^{\text{-evidence score}}$. Previous research demonstrated satisfactory psychometric quality for these measures of facial expressions (Beringer et al., 2019).

Before starting the emotional expression interference task proper, participants completed 15 practice trials where they posed facial expressions of joy and disgust and received visual feedback from FACET evidence scores. In the practice trial, they were asked to produce and hold the specific facial expression with maximal intensity before return to a neutral face in several consecutive trials. This allowed the experimenters to ascertain correct scoring of the facial expressions by the software and provided some practice to participants in controlling their expression intensity.

Two experimenters (one female) prepared EEG recordings after which participants remained alone in the shielded chamber. EEG and video recordings were monitored from outside. Instructions on the monitor emphasized not to move the head, to show all expressions as intensely as possible and to return to a neutral face immediately after each trial.

In the emotional expression interference task, participants were shown face portraits of women that either expressed joy, disgust, showed a neutral face, or scrambled versions of the neutral face. Participants responded to the letters “W” and “M” (randomly assigned between participants) placed as small symbols on the noses of the models' pictures by showing either a smile or an expression of disgust. Before the test phase we trained participants with 16 practice trials of the expression interference task. The temporal sequence of the task is displayed in Figure 1.

We used two different control conditions: (1) Stimuli with neutral expressions to minimize mimicry effects (Bruce, 2017) and (2) scrambled faces to minimize facilitation and interference effects that even neutral expressions of faces might create. Scrambled faces were deconstructed pictures of faces with neutral expressions, cut into 240 squares and reorganized randomly within the oval shape used for the face stimuli. Thus, scrambled faces present the same visual information of brightness and color as the original pictures of faces without any possibility of holistic face perception. Thus, we used the scrambled faces as control conditions for behavioral effects, but only neutral faces were used as reference condition for the ERPs, as scrambled faces initiate different neuronal processing mechanisms (faces vs. non-faces; e.g., Haxby, Hoffman, & Gobbini, 2002) and therefore are no appropriate reference for EEG. To sum up, combinations of deliberate expression and stimulus expression, result in eight different conditions:
(1) *congruent deliberate expressions of joy to stimulus expressions of joy*,
(2) *incongruent deliberate expressions of joy to stimulus expressions of disgust,*
(3) *deliberate expressions of joy to neutral stimulus expressions*,
(4) *deliberate expressions of joy to scrambled stimuli,*
(5) *congruent deliberate expressions of disgust to stimulus expressions of disgust*,
(6) *incongruent deliberate expressions of disgust to stimulus expressions of joy*,
(7) *deliberate expressions of disgust to neutral stimulus expressions*,
(8) *deliberate expressions of disgust to scrambled stimuli.*

Altogether, we presented 640 trials (i.e., 240 congruent, 160 incongruent, 160 neutral, and 80 scrambled trials, 50% each with deliberate smiles and expressions of disgust as responses.

The ratio of congruent to incongruent trials was shifted to approximately 65:35, making congruent trials more likely than incongruent ones, because participants typically show greater interference effects in Stroop tasks when the proportion of congruent trials is high (e.g., Crump et al., 2006). We presented all 640 trials in a fully randomized order, which took approximately 40 minutes plus five self-paced breaks.

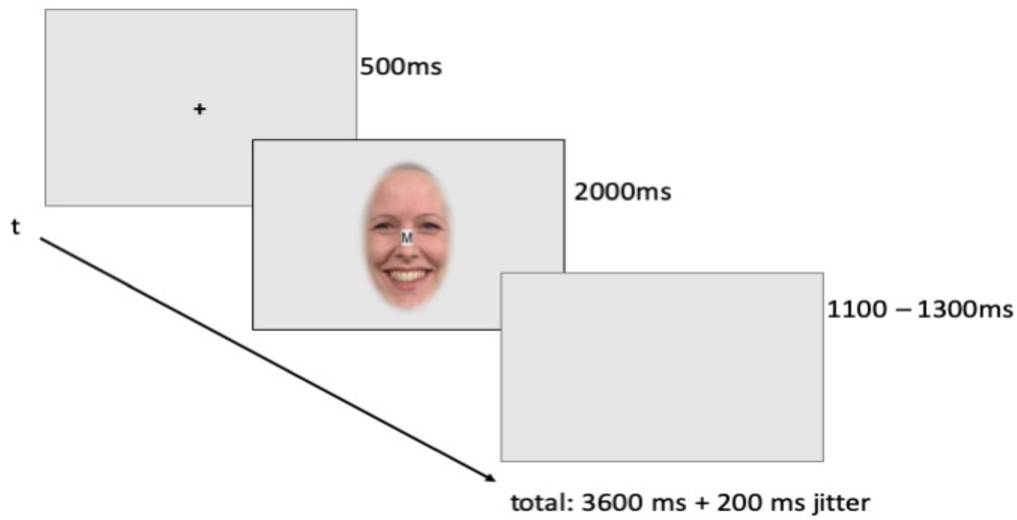


Figure 1. Trial scheme of the emotional expression interference task. Instructions before the task specified whether "W" or "M" (placed on the nose) was the signal for a deliberate expression of joy or disgust (randomly assigned). (+) Fixation cross.

Stimuli and facial attractiveness rating

Caucasian female models were taken from the Radboud database (Langner et al., 2010) (N = 19) and from the FACES database (Ebner, Riediger, & Lindenberger, 2010) (N = 21). We used Gimp (version 2.8.18) to fit the models' faces into a standardized vertical ellipse of 250x360 pixels, generating pictures of 400x600 pixel size with a resolution of 72 dpi. Scrambled face stimuli were generated with a Python-based program (version 2.7.12, Python Software Foundation) that separated the faces into blocks of 35 pixels and reassigned them randomly. That way, we produced 40 standardized stimuli, each one depicting a Caucasian female showing either disgust, joy, or a neutral expression that is reliably recognized by others (Ebner, Riediger, & Lindenberger, 2010; Langner et al., 2010).

Participants were to imagine that the female faces were viewed in the context of a dating platform. To obtain individual attractiveness ratings before the experiment proper we presented all three pictures (showing joy, disgust and neutral expression) of a given model (models' identity in randomized order) simultaneously and side by side for 1 second, with left, middle and right position randomly assigned (e.g., Golle, Mast, & Lobmaier, 2014). After one round of passively viewing all stimuli and imagining the online dating scenario, we presented the pictures again in the same way and asked participants after each stimulus presentation to rate the perceived facial attractiveness of each model on two visual analogue scales from 0-100 directly (first "How attractive do you find this woman?", 0 = not at all, 33 = somewhat, 66 = quite, 100 = very; second "Would you like to go for a date with this woman?", 0 = no, 33 = rather not, 66 = rather yes, 100 = yes).

Data Analyses

Video-data processing

In our previous work we defined standards of data preprocessing for automated software analyses and preregistered our analysis pipeline on the Open Science Framework (Beringer, 2018, e.g., https://osf.io/wh6rx/wiki/home/) before analyzing data (Beringer, 2018). We employed the same analysis pipeline as in a former study (Beringer, Wacker, & Recio, 2022). Exported data with ms timestamps and stimuli events was preprocessed with MATLAB (R2016a, The MathWorks, 2016). We provide a short description of how the code works in Appendix A of the supplemental material (a more detailed MATLAB example of how the code deals with an experimental condition can be found online: https://osf.io/wh6rx/wiki/home/). Here, we describe the two major steps in video data processing.

(1) Classification: We classified all trials that showed an onset of the target expression but not in other expressions as *hits* (90.4% of data), trials with onsets in any other expression but not in the target expression as *false positives* (9.4% of data) and trials without any data (*NaNs*) as *omissions*. Omissions also included trials with expression onsets within the first 7 frames after stimulus onset (*too early reactions*, because participants must have launched neural motor response before they have seen the stimulus) (0.1% of the data).

(2) RTs: we calculated RTs for hits as the interval between stimulus onset and participants' deliberate expression onset. Based on RTs, 2.5 % of all trials were classified as too early and 6.2% of the trials as too slow as described in the Participants section.

Due to the subject level dependencies inherent in our within-subject design, we used linear mixed effect models (LMM) for statistical analysis of RTs and generalized linear mixed models (GLMM) to analyze the accuracy data. Both models were calculated with the R package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) using R version 3.5.1 (R Core Team, 2018) while *p* values were obtained using the *Sattertwhaite approximation* for degrees of freedom (LMM), as implemented in the *lmerTest* R package (Kuznetsova, Brockhoff, & Christensen, 2017), and the asymptotic *Wald test* (GLMM) as implemented in the *lme4* package. All fixed effects were included as random effects in the LMM, as suggested by Barr et al. (2013), while we included deliberate expression as random effect in the GLMM due to convergence problems with higher order random structures. We used dummy coding (described as treatment coding in R) for all analyses with scrambled stimuli and deliberate joy expressions as reference condition. All analyses were conducted with a significance level of $\alpha = .05$.

Electrophysiological signals and signal processing

For EEG recordings we used 64 active electrodes (10-20 system) with a BioSemi ActiveTwo Mk2 amplifier, sampled at 2048 Hz with DG-2kHz bandwidth. Raw EEG data was down-sampled to 512 Hz with PolyRex (Kayser, 2003) and processed using Brain Vision Analyzer (version 2.1, Brain Products GmbH, Munich, Germany). Offline, continuous signals were low-pass filtered (12 Hz), high-pass filtered (0.03 Hz) and then segmented into 1 sec epochs starting 200 ms before stimulus onset. We considered only hits and false positives by exporting behavioral performance data from video recordings. After reimporting adjusted event markers to Brain Vision Analyzer we removed blinks and muscle artifacts from segmented data by means of independent component analyses. Thereafter, we interpolated faulty electrodes or those with high noise due to muscular activity using spherical spline functions. In 14 out of 26 cases (53.8%) we interpolated either the T7 or T8 electrode due to excessive artifacts during deliberate expressions of joy resulting from activation of the nearby zygomaticus and temporalis muscles. Because artifacts in frontal and temporal electrodes (above forehead and ears) were considerably larger (more than 2 SDs within participants), the following electrodes were excluded from further analyses prior to the process of artifact rejection: Fp1, AF7, AF8, T7, T8, TP7, TP8, Fpz. Thereafter, we computed an average reference for all remaining electrodes, followed by a baseline correction for the mean activity in the 200ms before stimulus onset. We marked segments with voltage steps larger than 100 µV/ms, amplitude shifts of 200 µV within a period of 200 ms, or amplitudes exceeding ± 200 µV as artifacts (6.1% of the data). Before conducting statistical analyses, we visually inspected grand averages of the N2 and EPN components across conditions in the current data to identify time windows in which the components were maximal over frontal sites (N2, based on Fz electrode, 180-220 ms after stimulus onset) and occipital sites (EPN, based on O1 electrode, 180-220 ms after stimulus onset, e.g., Hajcak et al., 2011), respectively.

For attractiveness ratings, a tercile split was calculated, thereby dividing the face pictures into three groups based on their perceived attractiveness by a given participant. Stimuli belonging to a participant's upper tercile of the distribution were classified as "attractive" while those in the lower tercile were classed as "less attractive". All trials scored as a hit were then sorted into 14 conditions for each participant (see Table 1). After the segmentation process, we applied another baseline correction and averaged all trials within conditions.

Table 1

Mean trial numbers for all conditions (interference – expression – attractiveness)

Congruent	**#Trials**	**Incongruent**	**#Trials**
joy – attractive	23.38	joy – attractive	36.27
joy – less attractive	28.50	joy – less attractive	43.88
disgust – attractive	20.96	disgust – attractive	34.77
disgust – less attractive	27.62	disgust – less attractive	42.15

Neutral	**#Trials**	**Scrambled**	**Trials**
joy – attractive	23.77	joy	38.04
joy – less attractive	29.38		
disgust – attractive	22.19	disgust	36.04
disgust – less attractive	26.96		

Note: Upper part indicates whether the stimulus expression was congruent or incongruent with the deliberate expression separately for stimuli rated attractive and less attractive (eight conditions). Left side of the lower part shows conditions with neutral stimuli and deliberate expressions of joy and disgust separately for attractive and less attractive stimuli (four conditions). Right side of the lower part shows conditions with scrambled faces as stimuli and deliberate expressions of joy and disgust (two conditions). #Trials = mean number of trials per condition after all data removal.

A priori, we planned the statistical analysis with two steps. (1) We first performed analyses of variance for repeated measures (rmANOVA; using R Studio, version 1.0.143, R Core Team, 2018) for EPN with two factors *type of stimulus* (expression of *joy, disgust, neutral*) and *response* (expression of *joy, disgust*) and for N2 with three factors *interference* (*congruent, incongruent, neutral*), *response* (expression of *joy, disgust*) and *perceived attractiveness* (*attractive, less attractive*). This step has the highest statistical power to indicate global main effects and their interactions. Whenever sphericity assumptions were violated, we report corrected Greenhouse-Geiser *p*-values. (2) If the first step revealed significance, we performed post-hoc *t*-tests to clarify the main effect. Pairwise comparisons between conditions with Bonferroni corrected *p*-values tested for significant differences in N2 and EPN.

Results

Behavioral performance

Reaction times and accuracy. Figure 2 depicts the average RT and accuracy in each condition. As shown in Tables 2 and 3, trials with deliberate expressions of joy were generally faster and more accurate than those with deliberate expressions of disgust. Moreover, congruent responses were faster and more accurate than responses to all other stimulus expressions regardless of the deliberate expression. There appeared to be little difference between the three other stimulus expressions, especially both reference conditions. Although, response expressions of joy in incongruent trials were slower than for all other stimulus expressions and less accurate than response expressions of disgust (see Table 2; diagnostic plots graphically revealing no violation of the statistical assumptions of the LMM are depicted in Appendix B of the supplemental material).

As depicted in Table 2, when responses were expressions of joy, congruent and neutral stimulus expressions were responded to significantly faster than scrambled stimuli. We found no interference effect for response expressions of joy as incongruent stimulus expressions were not responded to more slowly than scrambled stimuli. Moreover, deliberate expressions of joy were significantly faster than deliberate expressions of disgust. Interference effects of deliberate expressions of joy and disgust did not differ in the incongruent condition. In the incongruent conditions interference effects of deliberate expressions of joy were significantly bigger than interference effects of deliberate expressions of disgust. Interference effects of deliberate expressions of joy were significantly larger compared to deliberate expressions of joy in the condition with scrambled stimuli.

Interactions between RTs in conditions with deliberate expressions of disgust and congruent stimulus expressions were non-significant. RTs in conditions with deliberate expressions of disgust and neutral stimulus expressions did not differ, showing that the facilitation effect and the difference between neutral stimulus expressions and scrambled stimuli was also apparent with deliberate expressions of disgust.

The results of the accuracy analysis are depicted in the Appendix C of the supplemental material. None of the coefficients estimated by the GLMM reached significance, therefore we did not observe interference or facilitation effects in the accuracy data.

Figure 2

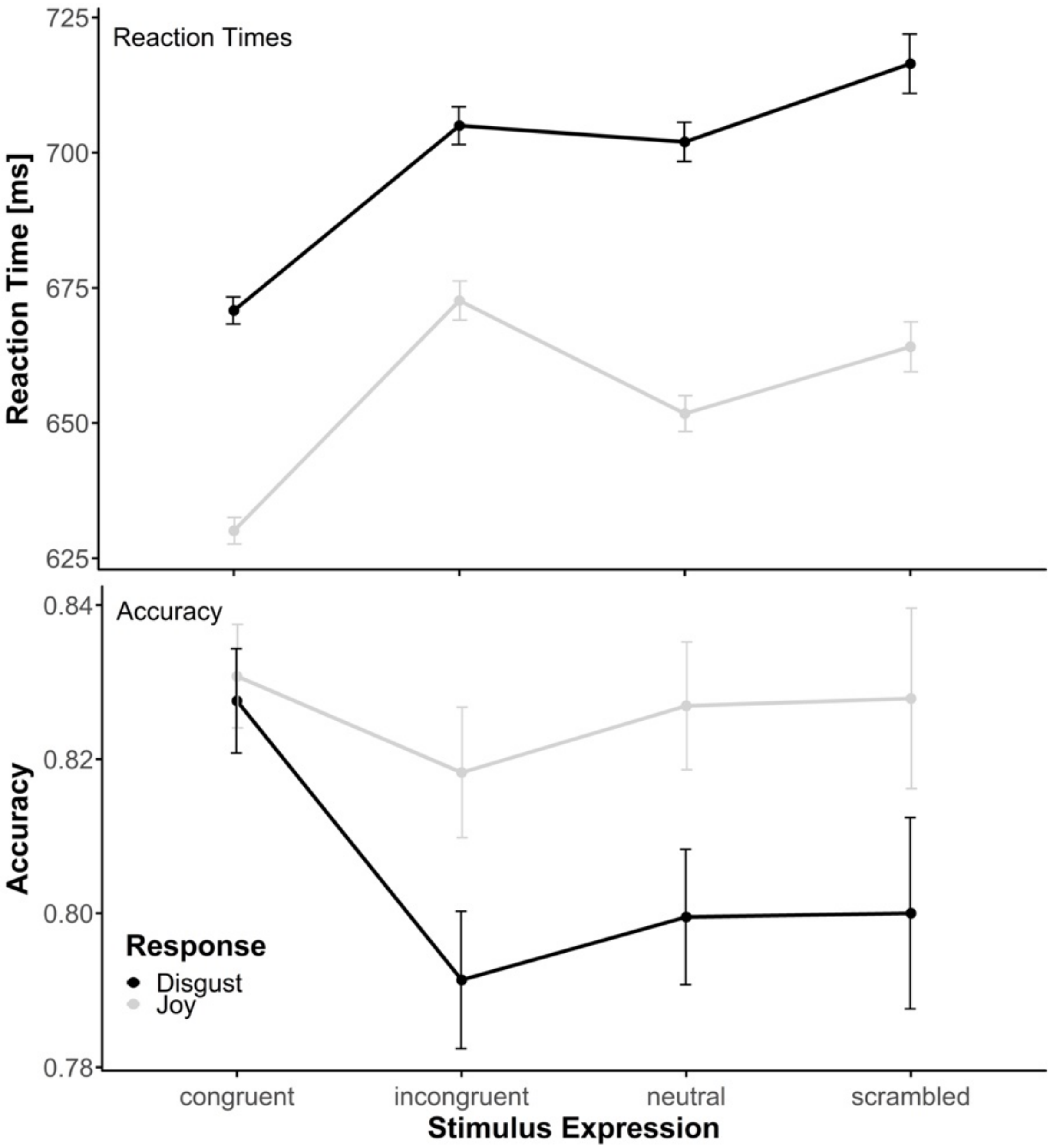


Figure 2. Mean response times (top) and mean accuracies (bottom) and their corresponding standard errors (provided by LMM/GLM) of each condition.

Table 2.

Fixed effects of linear mixed model of the reaction times.

Predictors	***b***	***ß***	**df**	***t* value**	***p* value**
Intercept (deliberate Joy, scrambled stimulus)	667.56 (17.98)		23.03	37.13	< .001*
Congruent-Scrambled	-34.82 (5.66)	-.12 (.02)	23.67	-6.16	< .001*
Incongruent-Scrambled	9.14 (6.24)	.03 (.02)	23.82	1.47	.160
Neutral-Scrambled	-13.51 (6.46)	-.04 (.02)	22.59	-2.09	.048*
Disgust-Joy	50.29 (12.34)	.18 (.02)	22.51	4.08	.001*
Congruent-Scrambled x Disgust-Joy	-11.2 (8.31)	-.03 (.02)	22.86	-1.35	.191
Incongruent-Scrambled x Disgust-Joy	-20.59 (9.31)	-.05 (.02)	23.25	-2.21	.037*
Neutral-Scrambled x Disgust-Joy	-1.48 (8.81)	.004 (.02)	22.23	-0.17	.868

Note. N = 26, df = Degrees of freedom, Standard errors are depicted in parenthesis, * p < .05

Facial attractiveness. To analyze the interaction of facial attractiveness with deliberate expression and the three-way interaction (facial attractiveness x deliberate expression x stimulus expression) we fit another LMM for RT analyses and another GLM for analyses of accuracy adding these effects as fixed effects and as random slopes for subjects (scrambled faces were excluded as they cannot contribute to any effects of facial attractiveness), attractiveness ratings were grand mean-centered, the neutral stimulus condition was the reference category for stimulus expression and joy responses were the reference category for deliberate expression. The results of LMM and GLM relevant to the influence of facial attractiveness on RTs and accuracy show that none of the interactions predicted RTs and accuracy significantly, except for the three-way interaction of deliberate disgust expressions in incongruent trials with perceived attractiveness for accuracy (see Table 3). There were no significant effects for RTs.

Table 3

Fixed effects of generalized linear mixed effect model predicting response accuracy by interactions of facial attractiveness, deliberate expression, and stimulus expression.

Predictors	**Estimate**	**OR**	***z* value**	***p* value**
AT x deliberate Joy	0.01 (0.01)	1.01	2.30	.022
AT x deliberate Disgust	-0.01 (0.004)	1.00	-1.17	.242
AT x deliberate Joy x stimulus Joy	0.01 (0.01)	1.00	0.74	.461
AT x deliberate Joy x stimulus Disgust	0.01 (0.01)	1.01	1.40	.162
AT x deliberate Disgust x stimulus Disgust	0.004 (0.01)	1.00	0.62	.537
AT x deliberate Disgust x stimulus Joy	-0.01 (0.01)	0.99	-2.73	.006*

Note. N = 26. AT = Attractiveness, OR = Odds Ratio. Standard errors are in parenthesis. Estimates and standard errors are given in log odds. *significant at a Bonferroni-corrected alpha level of α = .008 (0.05/6).

Event related potentials

EPN – effects of emotional expressions. The rmANOVAs of the EPN measured at the O1 electrode with the factors *stimulus type* (*stimulus expressions of joy*, *disgust* and *neutral expression*) and *deliberate expression* (of *joy* and *disgust*), revealed a significant main effect for the factor stimulus type, $F(2,50) = 6.62$, $p = .00067$, $\eta_p^2 = 0.007$. Pairwise two-sided *t*-tests between all three stimulus types showed significant differences between stimulus expressions of disgust and both other stimulus types, whereas smiling and neutral stimulus expressions did not significantly differ from each other in the associated EPN amplitudes (see Table 4). Neither the main effect of deliberate expression nor the interaction effect between the two factors was significant. Figure 3 shows the scalp distributions of the effects between 180-219 ms.

Table 4

Mean amplitudes (and SDs) of the EPN and p-values of pairwise two-sided t-tests adjusted according to the Holm method (Holm, 1979).

stimulus type	**mean activity (SD)**
expression of joy	-2.37 (4.44)
expression of disgust	-2.94 (3.96)
neutral expression	-2.13 (3.96)

Contrast between stimulus types	***p*-value**
expressions of joy vs. expressions of disgust	0.0021
expressions of joy vs. neutral expressions	0.24
expressions of disgust vs. neutral expressions	< 0.001

Figure 3

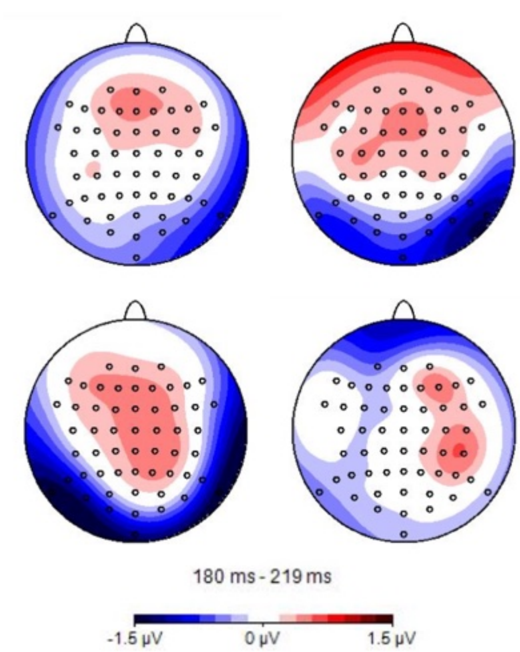


Figure 3. Scalp distributions 180-219 ms after stimulus onset for difference scores of *congruent expressions of joy* (upper left), *incongruent expressions of joy* (upper right), *congruent expres-*

sions of disgust (bottom left) and *incongruent expressions of disgust* (bottom right) and the respective neutral condition. Only incongruent expressions of joy (upper right) show a typical EPN topography.

N2 – effects of emotional expression interference. The rmANOVAs of N2 amplitudes at Fz with the three factors *stimulus congruency* (*congruent, incongruent, neutral*), *deliberate expression* (of *joy* and *disgust*) and *facial attractiveness* (*attractive, less attractive*), revealed a significant main effect for the factor stimulus congruency, $F(2,50) = 4.10$, $p = .0225$, $\eta_p^2 = .005$, as well as a significant interaction between the factors stimulus type and deliberate expression, $F(2,50) = 4.14$, $p = .0359$, $\eta_p^2 = .004$. The main effects of deliberate expression, facial attractiveness, and their interactions were not significant ($p > .05$ for all). Pairwise post-hoc *t*-tests showed that the main effect of the factor stimulus type was driven by a significant difference between congruent and neutral conditions ($p = .00024$). Examination of the interaction between stimulus type and deliberate expression indicated that the congruent disgust condition significantly differed from the incongruent disgust condition ($p = .0042$). No other pairwise comparisons revealed significant effects. Figure 4 shows the scalp distribution of the ERP in the N2 time window.

Figure 4

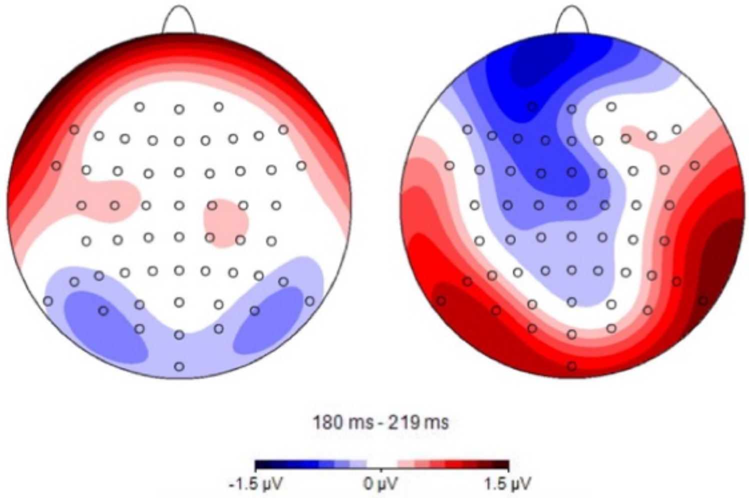


Figure 4. Scalp distributions 180-219 ms after stimulus onset for difference scores of *incongruent deliberate expressions of joy* minus *congruent deliberate expressions of joy* (left) and *incongruent deliberate expressions of disgust* minus *congruent deliberate expressions of disgust* (right). In line with inferential statistics, disgust shows a stronger negativity in the frontal electrodes and therefore a stronger emotional expression interference component.

Discussion

Using a variant of the Stroop task, we investigated how male participants control their deliberate facial expressions of joy and disgust *when seeing (task-irrelevant) female faces that show facial expressions, which were congruent, incongruent or unrelated (neutral) to the participants' expression.* We also investigated whether perceived attractiveness moderated behavioral performance. Analysis with machine vision software revealed facilitation effects for both deliberate expressions of joy and disgust in terms of shorter RTs in congruent conditions relative to the scrambled condition. Moreover, interference effects in RTs (incongruent relative to neutral expressions and scrambled faces) only occurred for deliberate expressions of joy. Furthermore, ERP data showed an EPN component in incongruent trials with deliberate expressions of joy and a main effect for stimulus type, indicating enhanced sensory processing of these conditions.

Whereas we did not observe a main effect of deliberate expressions on the N2 component, a significant modulation of the N2 associated with congruency seems to reflect a conflict emerging from the mismatch between the deliberate facial expression of disgust and stimulus expression of joy. Faces perceived as attractive revealed significantly lower accuracy when participants expressed disgust to a smiling stimulus. No effects of perceived attractiveness were observed for either RTs or ERPs.

Emotional expression facilitation and interference effects

First, facilitation effects for both deliberate expressions of joy and disgust reduced RTs in the congruent conditions and support our hypothesis about stimulus expressions as a task irrelevant information facilitating congruent deliberate expression. This can be interpreted as an effect of mimicry, therefore, as an automatic process occurring without deliberate control (e.g., Chartrand & van Baaren, 2009; Otte, Jost, Habel, Koch, 2011; Recio et al., 2022). Second, expression interference effects lead to longer RTs in incongruent conditions only for expressions of joy. It seems that, while RTs of expressions of joy can be modulated by mimicry in congruent conditions and distractors in incongruent conditions, RTs of expressions of disgust can be facilitated in congruent conditions only. From an evolutionary perspective this makes sense as RTs of facial expressions of disgust should be accelerated to avoid any intake of disgusting material, e.g., rotten food and therefore, are mimicked as fast as possible within a group. Whereas, facial expressions of joy can be useful in many different situations, therefore, RTs can be accelerated, e.g., to smile together and show affiliation and decelerated if an expression of joy is less appropriate.

Third, completing perspectives on behavioral effects, we need to discuss why there was no effect on accuracy. One could argue, to distinguish between two types of quite distinct expressions of emotion probably was no challenge for our participants. Participants also could have used strategies to avoid mistakes as the task description somehow suggests that the successful handling of the task is to avoid mistakes. However, thinking from a more psychological perspective of emotions for both theories of basic emotions (e.g., Ekman, 1971; Ekman et al., 1987) and dimensional circumplex models (e.g., Russell, 1980) it is crucial to not confuse expressions as different as joy and disgust. Both theories explain big differences e.g., of social functions of expressions joy and disgust (e.g., affiliation vs. aversion). This could explain why participants have been really carefully to not mix them up. Indeed, there is a meaning in facial expressions of emotion that probably is harder to confuse than reading the word blue in a green color but calling it green or clicking left or right as it is used in a lot of tasks in cognitive research. This point needs further elaboration especially important for formation of new hypotheses in psychological research of emotion.

Neurocognitive mechanisms of deliberate expressions

So far, the EPN component has been said to reflect visual processing of emotional stimuli (Hajcak et al., 2011), e.g., found as a negative deflection of difference scores between emotional and neutral pictures as well as between stimuli of facial expression of emotion and neutral facial expressions (Pourtois, Schettino, & Vuilleumier, 2013; Schupp, Flaisch, Stockburger, & Junghöfer, 2006), which is why we expected greater EPN components in incongruent conditions compared to congruent and neutral conditions. In former studies both facial expressions of joy and disgust elicited an EPN (e.g., Recio, Schacht, & Sommer, 2014; Xu, Sommer, & Recio, 2023). Our data do not confirm these findings in total, as we found a typical EPN scalp distribution only for the condition when participants responded with expressions of joy to stimulus expressions of disgust. One interpretation could be that the combination of our stimuli and the task was simply not emotionally evocative enough (e.g., stimulus expressions of joy are unrelated to threat or danger for the organism) and motivational salience and goal relevance are likely to lie behind the increased visual processing of emotional faces (Pourtois, et al., 2013). In other words, EPN reflects a processing of emotional stimuli relative to neutral stimuli as some attention to the emotional content. As EPN only revealed level of significance for stimulus expressions of disgust the remaining question is, how the better sensory, reflexive processing for stimulus expressions of disgust but not joy helps to explain the behavioral findings. Again, the effect of visual processing of emotional stimuli in the EPN, suggesting that observing a picture of a person showing disgust produces more cognitive processing. From an evolutionary perspective, it makes sense to be highly sensitive to another facial expression of disgust, e.g., as it could indicate that a member of the group shows a strong aversive reaction to someone, while it could be less crucial to be sensitive for neutral and expressions of joy.

Incongruent deliberate expressions of disgust elicited larger N2 amplitudes at fronto-central sites than congruent deliberate expressions of disgust. Previous research indicates that the N2 reflects conflict detection (e.g., Larson, Clayson, & Clawson, 2014) and executive cognitive control functions such as executive inhibition process (e.g., Kieffaber, & Hetrick, 2005). As hypothesized, the strongest conflict was likely associated with responding with an expression of disgust to a smiling stimulus in this task. Interestingly neither RTs nor accuracy stood out for this condition. The question here is why N2 indicates a conflict detection for disgust expressions while participants went on with prompt and accurate responses of disgust expressions. Firstly, RTs show an interference tendency for deliberate expressions of disgust in incongruent conditions similar to the significant interference effect of deliberate expressions of joy in incongruent conditions (see Figure 2). From that perspective, the effect of congruency in the N2 interacted with facial expressions, suggesting that smiling while observing a picture of a person showing disgust produces greater conflict. Ignoring facial expressions of disgust (during smiling) could demand more top-down attentional control when they are interfering with task goals. A further explanation for the relatively small N2 (and, thus, presumably conflict) in incongruent as compared to congruent deliberate expressions of joy is that smiling as a signal of affiliation is frequently used to de-escalate difficult situations and tense atmosphere, e.g., a disgust expression during a dating scene (e.g., Burgeois & Hess, 2008; Niedenthal, Mermillod, Maringer, & Hess, 2010). This way smiling to someone who shows an expression of disgust e.g., to cool the situation down would be an adaptive tendency to cope with the situation, which is surely part of our normal everyday behavioral repertoire.

A further point of discussion in the understanding of our results is that we observed the N2 in an earlier time interval (around 180- 220ms after stimulus onset) than it is usually observed (200-300ms, e.g., Hajcak et al., 2011), potentially due to the holistic and fast processing of faces compared to other stimuli (e.g., Richler et al., 2009). Up until this point, however, conflicts that can occur beyond laboratory settings have not been investigated extensively with N2, although social exclusionary events seem to be promising (Themanson, Khatcherian, Ball, & Rosen, 2013). Regarding the imaginary dating scenario, we could classify expressions of disgust of pictures of a woman as social exclusionary events, that makes the activation of the N2 very plausible. This far-reaching interpretation requires more attention of future research.

Facial attractiveness

When our male participants were to respond with expressions of disgust to a picture of a smiling woman perceived as attractive, this likely was the condition with the highest chance for emotional expression interference moderated by facial attractiveness. More specifically, in this condition participants were confronted with two sources of information making it more likely to respond with a smile. Firstly, the picture of a female face perceived as attractive supports a deliberate expression of joy rather than a deliberate expression of disgust, especially in the context of an online dating situation. Secondly, the smiling stimulus produces mimicry to smile back. In line with the idea of former research, mimicry probably is stronger in conditions with affiliative intent (e.g., Bourgeois, & Hess, 2008). The three-way interaction, increased errors in incongruent conditions with deliberate expressions of disgust are presumably due to the necessary deployment of cognitive processes required to resolve the detected conflict and to inhibit a seen expression of joy (e.g., Katembu et al., 2022).

Limitations and future perspectives

The present study has demonstrated that a combination of EEG and a machine vision software to classify facial expressions of emotion derives rich information about how deliberate control of facial expressions of emotion can be understood from a cognitive perspective. However, several limitations of the present study deserve to be mentioned. First, due to the all-male sample the results cannot be generalized to females and the social context in our laboratory setting was limited compared to a naturalistic use of online dating apps. Furthermore, our findings are restricted to posed facial stimuli and deliberate expressions, which are part of emotions but cannot stand for the whole complexity of the experience of emotion. In particular, we focused on facial expressions of joy and disgust, so we cannot generalize to other expressions nor to a higher level of abstraction, e.g., negative vs. positive emotional expressions, even though, the literature shows similar effects for expressions of anger compared to those of joy (e.g., Lee et al., 2008). Following, a categorial perspective on basic emotions and their facial expressions findings cannot be generalized on spontaneous emotional facial expressions which are probably more common in daily life.

Another problem was the limited variance in ratings of perceived facial attractiveness (see Appendix D in the supplemental material) in part probably due to our own stimulus standardization procedure which also lacks external validity for an online dating scenario. Wearing an EEG cap might have diminished participants' evaluation of their own attractiveness (e.g., Lee, Loewenstein, Ariely & Young, 2008) and might have affected the affiliative component of facial attractiveness.

Furthermore, future studies should elaborate emotional interference effects of different emotional expressions and close the gap between posed and spontaneous stimulus and deliberate expressions. This way we could investigate the link from expressions to the experience and functioning of emotions. Therefore, naturalistic study designs including affective arousal and expressions of emotion are necessary. Modern measurement techniques as machine vision software could help to analyze larger data sets than human raters could handle in more complex social settings. Therefore, standards and guidelines of pre-processing procedures must be developed to promote comparability between studies using machine vision software.

Conclusion

In general, the present investigation can improve understanding of cognitive control and deliberate facial expression of emotion. It seems, facilitation effects evoked by facial mimicry accelerates facial expressions for both joy and disgust making it plausible that cognitive functions modulate facial expressions. In particular, interference effects in RTs shown for deliberate expressions of joy and accuracy effects for deliberate expressions of disgust in incongruent conditions with stimuli perceived as highly attractive depict cognitive inhibition processes (e.g., Diamond, 2013). Here, we go beyond behavioral description of cognitive processes as EPN reflects bottom-up reflexive attention to the emotional content of disgust and causes higher inhibition costs when participants respond with an expression of joy underpinning behavioral effects of prolonged RTs. Furthermore, larger N2 amplitudes in incongruent deliberate expressions of disgust compared to congruent deliberate expressions of disgust indicate a conflict detection. Finally, and as hypothesized, the three-way interaction resulting from participants being less accurate in their responses with expressions of disgust to smiling stimuli perceived as attractive, probably indicates the strongest conflict in the emotional expression interference task. Thus, facial attractiveness seems to be a modulating factor for facial expressions of disgust. Taken all together, the presented study connects theories of expressions of emotion (e.g., Ekman, 1971; Ekman et al., 1987) and theories of executive functions (e.g., Diamond, 2013) and provides enriched information of behavioral effects by the combination of machine vision software and ERP correlates.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft [grant number RE 3721/2-1 to GR]. The authors thank Kristina Bien and Ronja Heintzsch for their help with data collection and analyses. We verified license agreements of iMotions before publishing the present study.

Competing interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplemental Material

Appendix A

Baseline: we computed individual baselines to control for individual differences in emotionality during neutral state due to individual differences in facial morphology, using a 6 sec interval without any head and face movements before the training phase (Olderbak, Hildebrandt, Pinkpank, Sommer, & Wilhelm, 2014). Additionally, we used the mean of the 7 frames (210 ms) before target stimulus was presented and subtracted it from every following evidence score of a given trial (sliding average). This procedure was done for every expression of all trials separately (Beringer et al., 2019) with the aim to correct intra-individual changes (e.g., mood shift) in facial expressions across the experimental session as it is a common mistake of measuring emotions without controlling for mood (Schmidt-Atzert, Stemmler, & Peper, 2014).

For classification we defined the expression onset using a threshold of evidence scores greater or equal 1, as recommended by software developers to measure clear expressions, for at least 7 frames (210 ms, lower limit for brief expression; Yan, Wu, Liang, Chen, & Fu, 2013) in a trial for all expression in the same way. See https://osf.io/wh6rx/wiki/home/ for a detailed example.

Appendix B

Diagnostic Plots

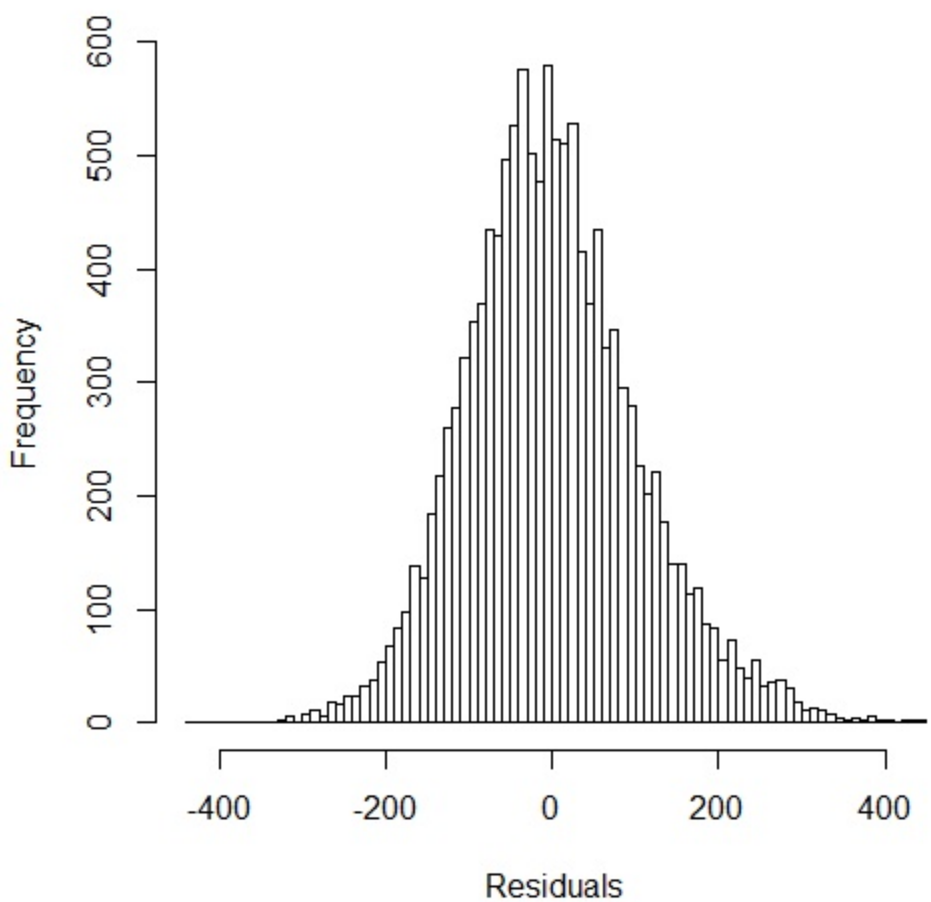


Figure A1: Histogram of the residuals of the linear mixed model analyzing the reaction times. No violation to the normality assumption of the residuals is assumed.

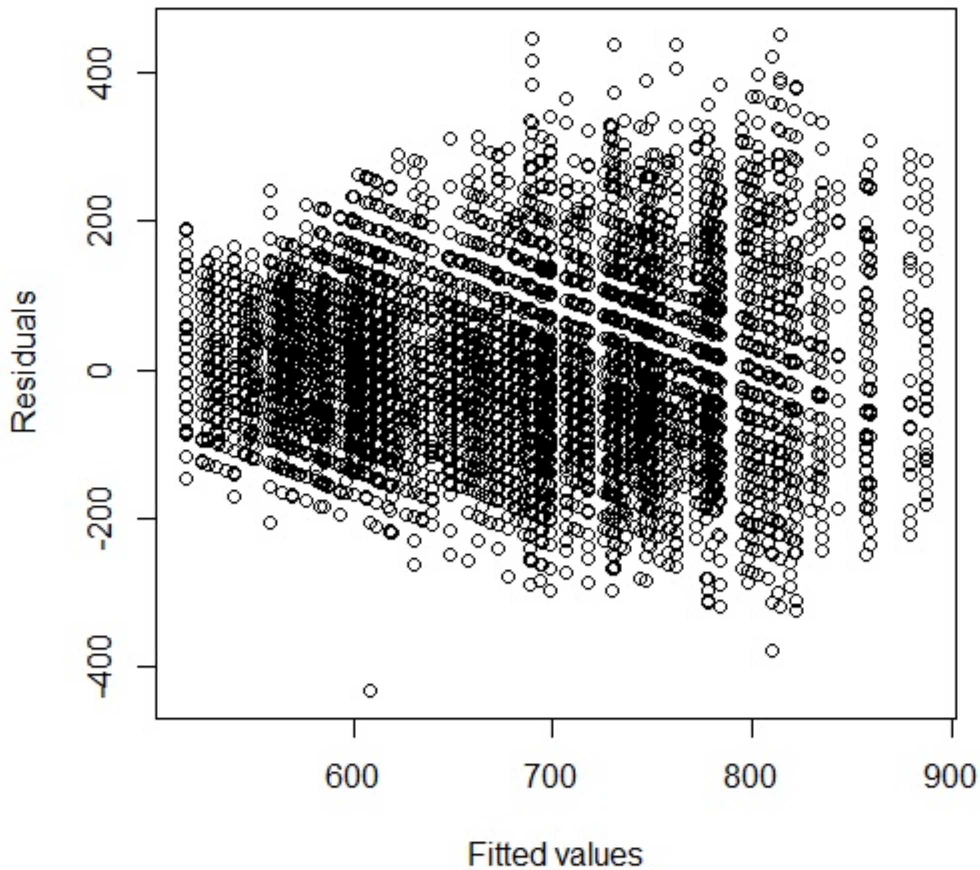


Figure A2. Plot of residuals (y-axis) and fitted values by the linear mixed model analyzing the reaction times. As there is no clear increase or decrease of residuals with increasing values, it shows no violation against the assumption of independence of residuals and predictors.

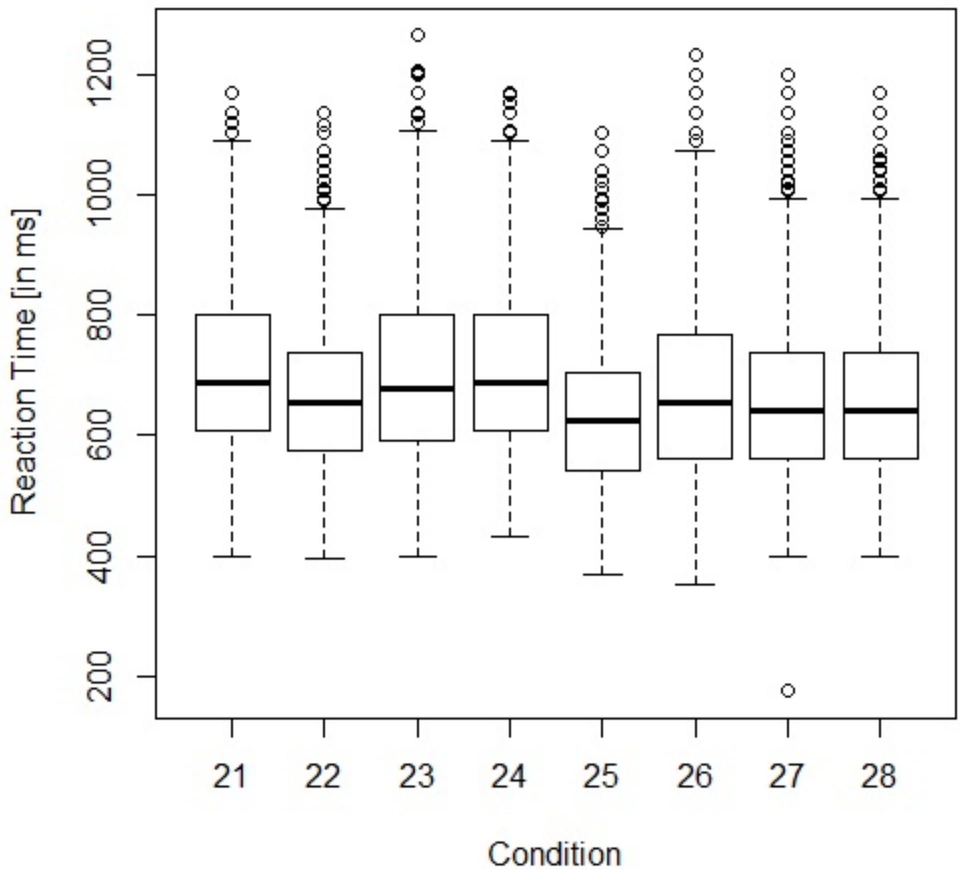


Figure A3: Boxplot of the reaction times in every condition. The conditions, depicted as deliberate expression (stimulus expression) are as follows: 21: Disgust (incongruent), 22: Disgust (congruent), 23: Disgust (neutral), 24: Disgust (scrambled), 25: Joy (congruent), 26: Joy (incongruent), 27: Joy (neutral), 28: Joy (scrambled). As they all appear to be roughly the same, no violation against the assumption of equal variances across conditions is to be seen.

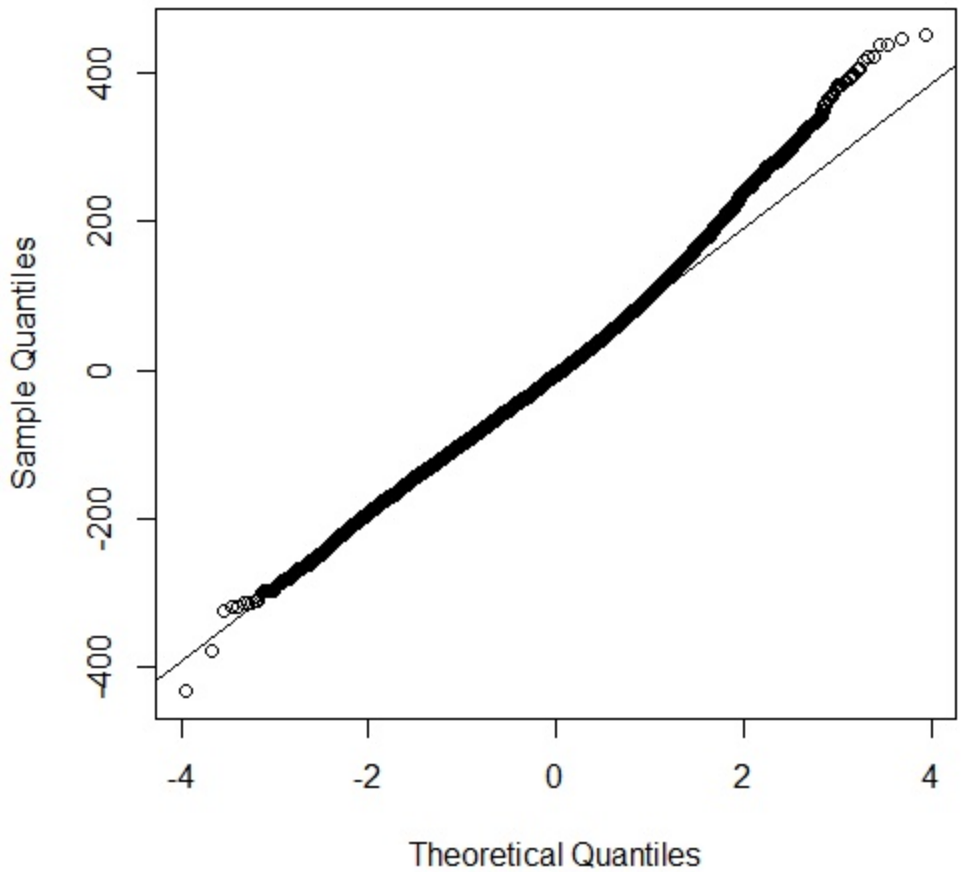


Figure A4: Normal Q-Q plot of the linear mixed model analyzing the reaction times. The sample quantiles fit the theoretical quantiles well until they reach the *z* values of about 2 and divert with higher values. As all other plots show no violations, we neglect this slight violation against the assumption of the normality of residuals.

Appendix C

Table A5. Fixed effects of generalized linear mixed model for response accuracy.

Predictors	**log Odds**	**Odds**	***z* value**	***p* value**
Intercept	1.81 (.13)	6.11	14.09	< .001*
Congruent	-.02 (.11)	0.98	-0.16	.087
Incongruent	-.13 (.11)	0.88	-1.17	.243
Neutral	-.02 (.11)	0.98	-0.19	.852
Disgust	-.27 (.16)	0.76	-1.72	.085
Congruent x Disgust	.19 (.14)	1.21	1.34	.18
Incongruent x Disgust	.1 (.15)	1.1	0.65	.518
Neutral x Disgust	.04 (.15)	1.04	0.25	.799

Note. N = 26, df = Degrees of freedom, Standard errors are depicted in parenthesis, * $p < .05$

Appendix D

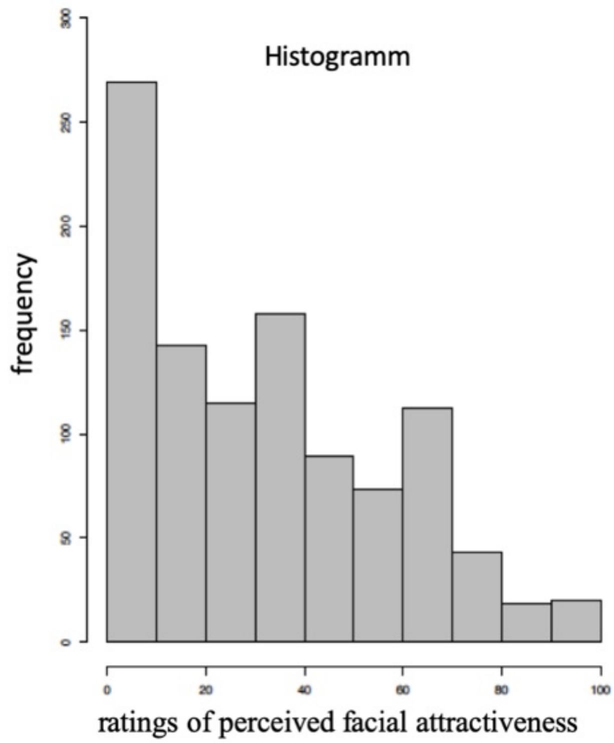


Figure A6: Histogram of attractiveness ratings.

References

Asendorpf, J. B., Penke, L., & Back, M. D. (2011). From dating to mating and relating: Predictors of initial and long–term outcomes of speed–dating in a community sample. *European Journal of Personality, 25*(1), 16-30.

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

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. doi:10.1016/j.jml.2012.11.001.

Beringer, M. (2018, March 14). Go-NoGo Facial Expression. Retrieved from osf.io/wh6rx.

Beringer, M., Wacker, J., & Recio, G. (2022). Deliberate control of facial expressions in a go/nogo task: An ERP study. *Acta Psychologica*, 230, 103773.

Beringer, M., Spohn, F., Hildebrandt, A., Wacker, J., & Recio, G. (2019). Reliability and validity of machine vision for the assessment of facial expressions. *Cognitive Systems Research*, *56*, 119-132.

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological psychology,* 77(3), 343-352.

Bruce, V. (2017). *Recognising faces*: Routledge.

Buss, D.M. (2006). Strategies of human mating. *Psychological Topics,* 15(2), 239-260.

Chartrand, T.L., & Bargh, J.A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology,* 76(6), 893-910. doi: 10.1037//0022-3514.76.6.893.

Chartrand, T.L., & van Baaren, R. (2009). Human mimicry. *Advances in Experimental Social Psychology,* 41, 219-274. doi: 10.1016/S0065-2601(08)00405-X.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., ... & Brass, M. (2018). Automatic imitation: A meta-analysis. *Psychological Bulletin*, 144(5), 453.

Crump, M. J., Gong, Z., & Milliken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic bulletin & review*, 13(2), 316-321.

Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*, 135-168.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological science, 11*(1), 86-89.

Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition & Emotion, 16*(4), 449-471.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods*, 42(1), 351-362.

Ekman, P. (1971). *Universals and cultural differences in facial expressions of emotion.* Paper presented at the Nebraska symposium on motivation.

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General,* 133(1), 101–135. doi: 10.1037/0096-3445.133.1.101.

Golle, J., Mast, F. W., & Lobmaier, J. S. (2014). Something to smile about: The interrelationship between attractiveness and emotional expression. *Cognition & emotion*, 28(2), 298-310.

Gray, J. R. (2004). Integration of emotion and cognitive control. *Current directions in psychological science*, 13(2), 46-48.

Hajcak, G., Weinberg, A., MacNamara, A., & Foti, D. (2011). ERPs and the study of emotion. *Oxford Handbooks Online*. doi: 10.1093/oxfordhb/9780195374148.013.0222.

Harrison, N. A., Morgan, R., & Critchley, H. D. (2010). From facial mimicry to emotional empathy: A role for norepinephrine? *Social Neuroscience, 5*(4), 393-400.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological psychiatry,* 51(1), 59-67.

Heyes, C. (2011). Automatic Imitation. *Psychological Bulletin,* 137(3), 463-483. doi: 10.1037/a0022288.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65-70.

Katembu, S., Xu, Q., Rostami, H. N., Recio, G., & Sommer, W. (2022). Effects of Social Context on Deliberate Facial Expressions: Evidence from a Stroop-like Task. *Journal of Nonverbal Behavior*, 1-21.

Kayser, J. (2003). Polygraphic Recording Data Exchange-PolyRex. Department of Biopsychology. *New York State Psychiatric Institute*.

Kieffaber, P. D., & Hetrick, W. P. (2005). Event-related potential correlates of task switching and switch costs. *Psychophysiology,* 42(1), 56-71.

Kinsey, A.C., Pomeroy, W.B., & Martin, C.E. (1948). *Sexual Behavior in the Human Male* (pp. 636–659). Philadelphia: W.B. Saunders; Bloomington: Indiana University Press.

Korb, S., Grandjean, D., & Scherer, K. R. (2010). Timing and voluntary suppression of facial mimicry to smiling faces in a Go/NoGo task—An EMG study. *Biological psychology,* 85(2), 347-349.

Künecke, J., Wilhelm, O., & Sommer, W. (2017). Emotion Recognition in Nonverbal Face-to-Face Communication. *Journal of Nonverbal Behavior*, 41, 221–238. https://doi.org/10.1007/s10919-017-0255-2

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82(13).* doi:10.18637/jss.v082.i13.

Lakin, J.L., & Chartrand, T.L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science,* 14(4), 334-339. doi: 10.1111/1467-9280.14481.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and emotion*, 24(8), 1377-1388.

Larson, M.J., Clayson, P.E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology,* 93(3), 283-297. doi: 10.1016/j.ijpsycho.2014.06.007.

Lee, L., Loewenstein, G., Ariely, D., Hong, J., & Young, J. (2008). If I'm not hot, are you hot or not? Physical-attractiveness evaluations and dating preferences as a function of one's own attractiveness. *Psychological Science,* 19(7), 669-677.

Lee, T. W., Dolan, R. J., & Critchley, H. D. (2008). Controlling emotional expression: behavioral and neural correlates of nonimitative emotional responses. *Cerebral Cortex,* 18(1), 104-113.

Lee, T.-W., Josephs, O., Dolan, R.J., & Critchley, H.D. (2006). Imitating expressions: emotion-specific neural substrates in facial mimicry. *SCAN,* 1, 122-135. doi: 10.1093/scan/nsl012.

Leonhart, R., & Lichtenberg, S. (2009). *Lehrbuch Statistik*: H. Huber.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and brain sciences,* 33(6), 417.

Olderbak, S., Hildebrandt, A., Pinkpank, T., Sommer, W., & Wilhelm, O. (2014). Psychometric challenges and proposed solutions when scoring facial emotion expression codes. *Behavior research methods, 46*(4), 992-1006.

Otte, E., Jost, K., Habel, U., & Koch, I. (2011). Exploring cross-task compatibility in perceiving and producing facial expressions using electromyography. *Acta psychologica,* 138(1), 187-192.

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.

Principe, C. P., & Langlois, J. H. (2011). Faces differing in attractiveness elicit corresponding affective responses. *Cognition and Emotion,* 25(1), 140-148.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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.

Recio, G., Shmuilovich, O., & Sommer, W. (2014). Should I smile or should I frown? An ERP study on the voluntary control of emotion - related facial expressions. *Psychophysiology, 51*(8), 789-799.

Recio, G., Surdzhiyska, Y., Bagherzadeh-Azbari, S., Hilpert, P., Nowparast Rostami, H., Xu, Q., & Sommer, W. (2022). Deliberate control over facial expressions in motherhood. Evidence from a Stroop-like task. *Acta Psychologica*, 228, 103652.

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision research,* 49(23), 2856-2861.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-192. doi: 10.1146/annurev.neuro.27.070203.144230.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161.

Schmidt-Atzert, L., Stemmler, G., & Peper, M. (2014). *Emotionspsychologie: ein Lehrbuch*: Kohlhammer Verlag.

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Progress in brain research*, 156, 31-51.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology,* 18(6), 643.

Tatarunaite, E., Playle, R., Hood, K., Shaw, W., & Richmond, S. (2005). Facial attractiveness: a longitudinal study. *American Journal of Orthodontics and Dentofacial Orthopedics,* 127(6), 676-682.

Themanson, J. R., Khatcherian, S. M., Ball, A. B., & Rosen, P. J. (2013). An event-related examination of neural activity during social interactions. *Social Cognitive and Affective Neuroscience*, 8(6), 727-733.

Xu, Q., Sommer, W., & Recio, G. (2023). Control over emotional facial expressions: Evidence from facial EMG and ERPs in a Stroop-like task. *Biological Psychology,* 108611.

Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., & Fu, X. (2013). How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of nonverbal behavior, 37*(4), 217-230.

3 Discussion

The presented project explored facial expressions of emotion through the lens of motor control, utilizing experimental tasks such as response priming and go/no-go tasks that traditionally involve pressing keys e.g., on a keyboard. For these types of responses, the calculation of hits, false positives, or omissions is straightforward, based on whether the participant pressed the correct key or not. In this project, similar tasks were adapted to require facial movements from the participants instead of pressing keys. In the first part of the discussion, I will briefly summarize the key findings of the studies presented above. Then, I give overall perspectives and a critically appreciative point of view before this section ends with a conclusion and an outlook into the future of research of facial expressions of emotion.

Publication 1 assessed (a) the accuracy of the automated assessment (i.e., FACET, iMotions, 2016) with two standardized databases of facial stimuli and whether recognition would be compromised by variations in angle, contrast, resolution, and size of the input video; and (b) the feasibility of using automated software analyses to measure experimental effects (i.e., differences in RT between two experimental conditions) using facial expressions as responses, and its correlation with EMG. The results demonstrated satisfactory reliability and validity of facial expression measurements with automated software analyses using FACET (Beringer et al., 2019). We observed similar validity effects in RTs with both FACET and EMG, indicating the suitability of FACET for scoring experimental effects (Objective 1). These findings support the use of this technique to investigate the motor control of facial expressions.

Publication 2 examined (a) the behavioral and ERP correlates of inhibition in producing smiles, fear, and disgust expressions using a go/no-go task; and (b) the moderation effects due to mimicry. A group of 37 participants either smiled when they saw a smiling face from the ADFES database (van der Schalk, Hawk, Fischer, & Doosje, 2011), showed disgust when they saw a disgusted face, or showed fear when they saw a fearful face (go trials). They inhibited the prepared expression (no-go trials) when indicated by a letter (W/M) randomly assigned on the models' noses. The same task was done with inverted faces and emotion-inducing pictures from the OASIS database (Kurdi, Lozano, & Banaji, 2017) without faces. For classifying FACET's raw data, a new multi-class algorithm was developed and preregistered by me (Beringer, 2018) on the Open Science Framework (Go-NoGo Facial Expression, e.g., https://osf.io/wh6rx/). Contrary to our expectations that participants would react faster and make more errors when seeing a face compared to inverted faces and OASIS pictures, no such effect was observed. Additionally, there was no higher P3 for harder-to-inhibit expressions, such as inhibiting a smile while viewing a smiling person compared to inverted faces and pictures without faces. Exploratory analyses revealed a strong effect of emotion, with more errors occurring in smiling conditions than in conditions of negative emotions. These findings reflect social norms of smiling back at smiling people and inhibiting negative emotional states (Objective 2 and 4). Nevertheless, these results must be interpreted cautiously, as there are still open questions about whether FACET's evidence score for smiling is on the same scale as other expressions.

The second Part of Publication 2 examined (a) the behavioral correlates of inhibition in the production of smiles and disgust expressions using a go/no-go task in a real-life interaction between a female confederate and a female participant; and (b) the moderation effects due to eye contact. As this experiment was designed as a pilot, a relatively small group of 16 participants either smiled back through a voltage-sensitive LC glass shutter fixed on a table (see Hietanen, Leppänen, Peltola, Linnaaho, & Ruuhiala, 2008 for a similar procedure), responded with a disgust expression (go trials), or inhibited their prepared expression (no-go trials), indicated by a red light signal randomly glowing on the nose of the model. Meanwhile, in half of the trials, the model was instructed to make eye contact with the participant, and in the other half, to avoid eye contact by focusing on the right or left frame of the glass shutter. RTs were consistent with our hypotheses, showing the fastest RTs for expressions of joy and eye contact. Error rates revealed a significant interaction between expression type and eye contact. Specifically, there were more errors in no-go trials involving eye contact only for expressions of joy. This suggests that mimicry might be greater for expressions with high affiliative intent, such as expressions of joy, and reduced for negative expressions. These findings support the idea of emotion-specific effects of facial expressions and the specific role of eye contact in face-to-face interactions (Objective 4).

Publication 3 examined (a) the behavioral and ERP correlates of facilitation and interference in producing smiles versus prototypical expressions of disgust using a modified Stroop task; and (b) the moderation effects due to attractiveness. A group of 26 men first observed a set of portraits of women from two standardized databases, indicating whether they found the women attractive and if they would be willing to go on a date with them if they were on an online dating platform. In the second part of the experiment, participants were instructed by two letters (M, W) to either smile or show facial expressions of disgust. The letters were displayed superimposed on pictures of women showing different facial expressions. The facial expressions requested in the instruction and the one shown in the picture could be congruent (e.g., a smile requested as a response while a task-irrelevant picture of a woman smiling is seen in the background) or incongruent (e.g., disgust requested as a response while a task-irrelevant picture of a woman smiling is seen in the background). Neutral and scrambled faces were used as baseline conditions. As expected, results revealed performance facilitation in the congruent conditions (faster and more accurate responses) and interference in the incongruent conditions (slower and less accurate control) compared with performance when neutral and scrambled faces were shown in the background (objective 1). These findings reflect the automatic tendency to imitate the facial expressions of joy and disgust, which must be inhibited to solve the task (objective 3). ERP data showed that interference was larger for joy than disgust. Smiling at a picture of a person showing disgust interfered more with performance than showing disgust to a person who is smiling, presumably because negative expressions are more difficult to ignore (objective 2). Since smiles typically reflect affiliative motives, we expected perceived attractiveness to modulate the congruency effect. Linear mixed models (LMMs) revealed that attractiveness facilitated smiles in congruent trials (objective 4). ERP data revealed a modulation of the N2 component associated with congruency, possibly reflecting a conflict emerging from the mismatch between the facial expression requested in the task and the one shown in the stimulus as a distractor (objective 1). Mirroring the behavioral data, the effect of congruency in the N2 interacted with facial expressions, suggesting that smiling while observing a picture of a person showing disgust produces greater conflict (objective 2). Ignoring facial expressions of disgust seems to demand more top-down attentional control when they interfere with task goals.

That concludes the summary of the key findings; a detailed discussion of these findings can be found in the respective publications and will not be repeated here. Furthermore, this dissertation highlights three global key perspectives, which I will describe before presenting the conclusion and future directions. The first perspective relates to the methodological development required to integrate traditional EMG research with our new measurement technique of automated assessment of facial expressions of emotion. The second, derived perspective is theoretical, focusing on the underlying emotion theories. This theoretical perspective represents the core psychological aspect of the research project and invites a critical evaluation within the context of the employed investigative methods. The third and final perspective is a meta-perspective, which allows for an overarching view of the entire research project, from planning and execution to analysis and publication. Of course, this differentiation into three distinct perspectives is somewhat artificial, as these perspectives interact and overlap.

3.1 Methodological perspective

The technical implementation of automated assessment techniques, which proved to be far more complex than initially anticipated and comparable to EEG data, involved numerous challenges and relatively advanced statistical analyses, making communication and publication particularly difficult. Nevertheless, I consider the technical emancipation of this work to be its greatest and most significant contribution to the current state of research. Across all three publications presented here, it has become evident that the new automated assessment technique for facial expressions of emotion offers significant technical emancipatory potential. This technology allows, for the first time, the rapid conversion of diverse datasets from an empirical to a numerical framework, enabling measurements of material that, unlike EMG techniques, was not specifically produced for this purpose. In principle, it can analyze all image and video data where the Viola-Jones algorithm used by FACET detects a face. This capability enabled us not only to re-evaluate established and standardized research databases with the new technique, as detailed in the first part of the first publication, but also to convert our own video recordings of participants into a comparable and objective data structure without the need for extensive, costly, and time-consuming training in coding Action Units. While this new automated measurement technique overcomes the "black box" nature of EMFACS by Friesen and Ekman (1983), which requires training to use effectively, it also introduces a new "black box" that needs further investigation.

The first publication addresses this question, but the development of an appropriate code for baselining and classification of FACET data was a part of the analysis throughout all the studies described here. The initial studies were essential for assessing the feasibility of automated analyses for investigating facial expression control. Study 1 proposed that a value of 1 serves as a good threshold to indicate the presence or absence of fully intense facial expressions. This threshold appears to be valid across different facial expressions for most participants. Analyses using controlled stimulus material showed that, similar to human raters, the automated assessment software (i.e., FACET) tends to classify expressions of happiness more accurately. This difference is particularly significant for weak and blended expressions (Calvo, Avero, Fernández-Martín, & Recio, 2016; Del Líbano, Calvo, Fernández-Martín, & Recio, 2018). Additionally, these analyses demonstrated that FACET evidence scores are viable for quantifying expression intensity. An open question remains whether the evidence scores for expressions of joy are on the same scale as those for other expressions or if they differ. This is one of the reasons we have avoided statistically comparing indicators derived from FACET data, such as reaction times and amplitudes, across different emotional qualities up to this point. It remains unclear whether any observed differences are truly attributable to variations in emotional facial expressions or if they might instead result from technical artifacts, such as differences in data scaling.

Throughout the project, we focused intensely on developing an algorithm to classify participants' responses as hits, omissions, and false positives. We built upon the algorithm developed by Beringer et al. (2019), which compared data from automated assessment and EMG by using EMG onset for target and non-target muscles to classify hits and errors. The new algorithm faced several challenges, such as ensuring consistent performance across all trials and participants, regardless of individual differences in performance and speed, and managing unexpected artifacts in facial expression measurements, such as co-activation of similar expressions or blinks. Our developed algorithm computes reaction times (RTs), hits, false positives, and omissions for all participants across three different tasks: response-priming, Stroop, and go/no-go. It simultaneously considers all emotional expressions for classification, rather than just distinguishing between two categories (e.g., target vs. non-target). This approach provides qualitative information on whether a participant's expression was "pure" or mixed with other expressions. Additionally, the algorithm accounts for both the onset and offset of expressions, allowing for the quantification of expression duration and intensity. In summary, the novel aspects and advantages of the developed algorithm are:

a) Reliable classification of all trials across multiple participants and tasks

b) Handling of artifacts in the data without relying on Loess smoothing

c) Consideration of all expression categories, distinguishing between pure and mixed expressions

d) Quantification of onset, offset, overall duration, and intensity of expressions.

It should be noted, however, that we encountered a relatively high dropout rate due to issues with FACET. These problems included difficulties in scoring facial expressions for participants with beards or those not displaying prototypical facial expressions (i.e., expressions that do not match Ekman's descriptions, which the program classifies most effectively). Additionally, trials where the software indicates the presence of two facial expressions simultaneously are more likely when the expressions involve the same facial areas. For instance, both disgust and anger engage the m. corrugator supercilii, while both disgust and smiles involve mouth opening. In summary, the specificity of the raw data provided by the software is limited in certain cases.

Ultimately, it became clear that the complexity of data generated by automated assessment of facial expressions of emotion is comparable to that of EMG and EEG data, and therefore requires careful data preparation. Such data preparation entails several challenges. Decisions frequently had to be made, such as which threshold to use, whether to fix the threshold or determine an empirically optimal threshold for each experiment, participant group, or even at the individual level. Although these decisions were carefully considered, they were also somewhat arbitrary and required clear communication, which complicated the methods and analysis sections of our publications. Consequently, the technical details often ended up in the appendices of the publications, likely leading to limited discussion.

Beyond that, it remains unclear how the decisions made during data preparation affect the analysis of psychologically relevant effects, whether found or not. This issue is common in empirical research with multiple possible analytical approaches, often receiving little attention in the scientific community. However, due to limited resources, we were unable to thoroughly investigate all possible analytical paths down to the last technical detail (e.g., in the context of a forking path analysis, see Wacker, 2017). This highly technical but equally significant and interesting aspect of the dissertation could have been developed into a separate paper in collaboration with a methodologically focused research group. As it stands, our work was a balancing act between producing a technically and methodologically rigorous paper and addressing interests in emotion theories. Ultimately, we reached a compromise by developing and testing the analysis algorithm on a small sample of two to three participants from each experiment, and then preregistering it online on a scientific platform to meet the newer standards of the empirical research community. However, it is foreseeable that gold standards for the development of such analysis protocols will need to be established, similar to those now available for EEG, MRI, and fMRI data preparation. This would require much greater collaboration among research groups and laboratories.

Another challenge was the presence of artifacts in the EEG data resulting from facial muscle activation. To address this issue, we removed artifacts using independent component analysis (ICA) and concentrated on the ERP components occurring approximately 400 ms before the onset of facial expressions. Additionally, we focused on central electrode sites, where artifacts were less likely to be present. There is a knowledge gap in systematic research regarding the compatibility and interactions of different measurement methods, such as EEG and automated assessment of facial expressions. Moving forward, future research needs to employ more advanced methods for data cleaning, which have proven effective in removing muscle artifacts from speech production (e.g., Ouyang, Sommer, Zhou, Aristei, Pinkpank, & Abdel Rahman, 2016).

3.2 Theoretical perspective

I will now evaluate the results in a straightforward manner, as they were mainly published, before moving on to a critical perspective. One of the significant contributions of this work to current research on emotions and their expression within psychology is its connection to well-known and established emotion theories. This includes Darwin's theory, who observed typical facial movements for discrete emotional states, often referred to as "basic emotions" (Darwin, 1998), and the later more formalized theories of Ekman and Oster (e.g., Ekman & Oster, 1979), which have proven valid for assessing facial expressions with considerable cross-cultural commonalities. Additionally, the work references Scherer's theory, who attempted to link emotion psychological concepts with cognitive psychological concepts and developed a theory of internal push factors (e.g., Scherer, 1992). Competing models, such as those by Russell and Mehrabian, which describe emotions in a dimensional rather than discrete categorical manner (e.g., Russell & Mehrabian, 1977), are also considered, although this work does not aim to resolve the historical debate between categorical and dimensional models. Furthermore, we have attempted to connect these emotion theories with theories of cognitive executive functions. This includes Diamond's theory, which describes various cognitive top-down processes for adapting and controlling behavior, often divided into updating and switching (e.g., Diamond, 2013), as well as the theory by Miyake and colleagues, which describes the voluntary control of a prepotent or automatic motor response (e.g., Miyake et al., 2000).

The present investigation can significantly enhance our understanding of the control over facial expressions, an area with limited research. Control of facial expressions is crucial for successful social interactions and emotion regulation. The project also demonstrated that automated analyses of facial expressions using computer software could be a highly effective method for further exploring facial expressivity in controlled settings, leveraging tasks with a long tradition in experimental psychology that use facial expressions as dependent variables. A particularly interesting application of this method would be in psychotherapeutic settings. For instance, using the software to investigate facial expressions of emotion could help determine how psychotherapy assists patients in connecting with difficult feelings over time. It would be fascinating to observe whether event-related facial expressions (e.g., when a patient discusses a fearful situation or a traumatic experience) change over time, such as the intensity of emotional facial expressions becoming more appropriate to the recounted situation, or less/more intense.

Future research could explore additional tasks using facial expressions as responses, increasing task difficulty, and measuring variance in speed and accuracy performance. This approach would provide new insights into individual differences in controlling facial expressivity. Another valuable avenue for future research is examining the use and control of facial expressions in face-to-face communication settings, employing similar methodologies as used in the last experiment. This would deepen our understanding of the dynamics of facial expressions in real-world social interactions. Overall, the potential applications and follow-up research outlined here can offer valuable contributions to both theoretical understanding and practical applications in the study of facial expressions and emotion regulation.

So much for a progressive, more favorable than unfavorable but perhaps somewhat unrealistic perspective on the present work. However, I do not wish to overlook the growing criticism within the research community regarding emotional facial expressions. Retrospectively, and with all the experience we have gained throughout the research process up to now, I would say that the present work lacks something very crucial: a process theory of emotions that could be empirically examined and falsified, similar to cognitive process theories. While the theories of Darwin, Ekman, and Russell are interesting and important for the work presented here, they are less suited to describe temporally fine-grained questions such as the inhibition of joy or anger. There are process theories in emotion psychology, such as Schachter and Singer's two-factor theory (e.g., Schachter & Singer, 1962), James's theory (e.g., James, 1884), or Gross and Thompson's emotion regulation model (2007), but the experimental paradigms used in this work are not suitable for examining these theories. In hindsight, it seems that an attempt was made here to apply clear, well-known, and relatively simple ideas from cognitive process research to emotion psychology. This research project draws on theories of cognitive control, such as those by Diamond and Miyake et al. (e.g., Diamond, 2013; e.g., Miyake et al., 2000). However, these theories are investigated with research paradigms that traditionally come from fields like intelligence research and, like classical test theory, were developed for a group of psychological constructs closely related to cognitive performance, or what is now presumably called "g". In the context in which they were developed, it might make sense to use very similar, almost identical items to improve reliability, or to span a broader variance and thus differentiate a construct more effectively within a studied population, or, as often used in EEG research, to repeat dozens or even hundreds of slightly varied items and measurements to filter a signal from substantial noise. Unfortunately, this approach in emotion psychology leads to the paradox that the repeated measurement of an emotional phenomenon results in the disappearance of the emotion in the studied subject, often replaced by other psychological phenomena such as fatigue, anger, frustration, or boredom, where there was once joy at the sight of an e.g., IAP slide with a small dog or a cute cat. While e.g., IAP slides, or similar images may be well-researched for their emotion-inducing effects, it is retrospectively apparent that our participants did not experience and express the induced emotion continuously over 90 minutes. Instead, they produced an emotional facial expression that matched the affective quality of an image because we instructed them to do so. Clearly, emotional states cannot be induced indefinitely and thus evade the possibility of reliable and valid measurement, as can be achieved in many areas of cognitive psychology.

One might refer to this as "emotion research without emotions," and Barrett and her colleagues provide a detailed and impressive account of the deadlock we currently face with this type of research (e.g., Barrett et al., 2019). They propose that the extensive literature on emotional facial expressions, which has existed for many decades, should be reinterpreted as a study of facial expressions in a contemporary scientific understanding, removing the term "emotional."

Barrett and her colleagues demonstrate that most publications in our research field significantly lack the actual induction and/or measurement of emotions. This critique is comprehensive and addresses the discussion about the knowledge gained from such studies, including those presented here. In our case, what remains is a lengthy learning process regarding the handling of software for the automated measurement of facial expressions “associated with emotional states”, but unfortunately, no gain in understanding the functioning and differences between emotional experiences and their facial expressions.

One of the most consistent effects we observed during all laboratory assessments – and which Prof. Dr. Werner Sommer, who has decades of experience in the study of emotions and emotional facial expressions, spontaneously shared at a conference between our labs – is that participants during these kinds of experiments suffer from fatigue, their mood becomes bad and even their eyes start drooping between trials. Although they continue to produce facial expressions rhythmically, they are certainly not experiencing joy after 60 to 90 minutes when they see the thirtieth slightly varied cute pet on an IAP slide “valid” to induce joy. We accounted for this effect through our baselining procedures during data preparation, incorporated breaks for participant recovery during measurement, and carefully monitored participants from the observation room to avoid massive data loss. However, we never systematically investigated, published, or made this habituation effect available for discussion to the scientific community. The question remains as to what causes the effects we found. I suspect that most of the effects can be explained by cognitive processes, and that emotional involvement is little to none.

Unfortunately, the theoretically robust connection to the neuropsychological underpinnings in the presented studies is also rather weak, and the analysis of ERP components is more aligned with theories of cognitive psychology rather than emotional psychology. Here too, formulating a process theory of emotions would be a priori necessary to derive insights from empirical experiments. In the future, it will be crucial to avoid such eclectic effect research and invest much more effort into developing robust or at least promising theories if emotional psychology is to have a future. Viewed generously, this dissertation reflects the longstanding problems of emotional psychology, which has been a neglected research field within psychology for decades and has developed relatively little compared to e.g., research on intelligence. In this sense, the present work can be seen as another attempt to address emotional constructs through empirical research, and it is important that research endeavors are also allowed to fail.

3.3 Meta-perspective

In addition to the fundamental critique of this dissertation mentioned above, this work still represents a serious attempt to establish a link between the necessary methodological work and emotional psychology, and thus was well situated within the research group of Differential Psychology and Psychological Diagnostics, even though no differential hypotheses were examined. The work occupies a potential intersection between at least three “research universes” that surprisingly have little overlap. Worldwide, there is a large research community focused on developing statistical methods for analyzing complex data, which seems to have more in common with mathematics than with psychology. On the other hand, there is another research community focused on developing and programming software that recognizes faces and extracts information from them. This research group has grown significantly in recent decades and is more rooted in computer science than in psychology.

The third major research group publishing on the topics described here consists of emotion psychologists, whose work is divided into somewhat disparate cognitive theoretical frameworks attempting to bridge with emotional psychology, genuine emotional psychological theories that stand somewhat isolated from cognitive psychology, and a broad field of psychoanalytic emotion theories that have been entirely omitted in this work. These psychoanalytic theories are often explained through narrative forms rather than reduced models and frequently address unconscious emotions (e.g., Benecke & Brauner, 2017), but are well-integrated within their own tradition of theories.

In this dissertation, it was a significant challenge to navigate, read, empirically work, and write across these three “research universes”, with the constant risk of veering too much into technical-methodological details, particularly as the new technical challenges often dominated the work. At the same time, our publications are aimed at a psychological research audience that often does not deeply understand the necessary programming, such as for a new multiclass classifier algorithm. On one hand, there were no other research groups from psychology with whom one could exchange e.g., questions about movement artifacts of a specific facial muscle during a facial expression of disgust. On the other hand, our own understanding of FACET is also limited, e.g., as my mathematical understanding of vector transformations and deep learning algorithms is quite restricted.

Additionally, it was strikingly impressive in this research endeavor that shortly before we published our initial studies validating FACET, Apple Inc. acquired the FACET algorithm from iMOTIONS Inc. Since then, FACET has been classified as confidential and is now the property of Apple Inc., no longer available to the scientific community. We were fortunate to continue using an offline version of the FACET algorithm, but understandably, the scientific interest and long-term usefulness of our validation studies have been limited. iMOTIONS Inc. subsequently offered us another software that, according to the software engineers, functions “similarly well”. However, a new software would not have been scientifically usable without a renewed evaluation of the main quality criteria. Today, open-source tools are of increasing interest, though they again require validation studies to establish their psychological validity criteria. Moreover, it becomes clear that significant involvement from the research community is needed to ensure that good measurement instruments can be sustainably used in scientific settings, such as universities.

Ultimately, the "don't open Pandora's box" mentality among colleagues, including at conferences within the emotion psychology research community – particularly in response to critique by Barrett and others – was not particularly helpful in addressing the challenge of conducting emotion psychology without accurately capturing emotions. This issue of lacking construct validity also affects neighboring disciplines, e.g., research on creativity but the only explanation I can offer is that many colleagues write their scientific qualification theses on this topic and fear that failure could prevent them from publishing. Therefore, it is important, at least at this point in a scientific work that is free from publication bias, to be able to formulate such criticism and acknowledge a certain level of failure.

3.4 Conclusion and future perspectives

With the criticism I have mentioned, I do not mean to suggest that the present work has not yielded important insights. As mentioned above, we have found a reliable and valid way to measure emotional facial expressions using automated assessment techniques and have gained substantial experience with this relatively new method. What we have not achieved is finding a way to truly induce emotions so that we can examine them quantitatively and validly. This issue pertains to the construct validity of facial expressions of emotion, bringing me back to the beginning of this dissertation, namely its title, which I registered at the beginning of my research in 2017. Critically speaking, I would no longer agree that we have studied facial expressions of emotion here. More generously, we have contributed by testing automated assessment of facial expression techniques in the laboratory with familiar experimental paradigms.

At this point, I would like to attempt to integrate the praise and criticism described in this discussion and re-evaluate the goals and questions outlined in the introduction. Objective 1 addressed the question of whether automated assessment of facial expressions of emotion enables reliable and valid measurement. I am confident that we have achieved this goal and gained significant insights, particularly into the complexity of real data from real participants. We have found a transparent and attractively complex solution for data handling with the new Multiclass Classifying algorithm, although many further developments are still to be expected, similar to the development of gold standards for EEG data. At this point in the research process, I would assume that measurement with FACET meets the criterion of construct validity. The lack of construct validity in our studies is attributed to the experimental designs.

Objective 2 concerned the potential interactions between emotion and executive control over facial expressions. I believe the studies presented have demonstrated that facial expressions, much like hand movements, require executive control, such as inhibition. However, whether this inhibition is specific to emotional states and their expressions cannot be conclusively determined from the current work, according to the critiques mentioned earlier. Thus, we have only partially achieved Objective 2, although I consider the insight that emotional arousal is indeed necessary for this investigation to be less trivial than it may initially seem.

Objective 3 focused on the automaticity of facial expressions. Publication 3 has demonstrated that emotional facial expressions also elicit automatic response tendencies, such as mimicry, which may need to be inhibited. In particular, I find the pilot study involving a real person as the counterpart promising in terms of construct validity for emotional facial expressions, as it is more likely that many of the evaluated trials actually induced emotional states that we were able to measure. For the experimental Stroop design in Publication 3, however, my previously expressed concerns still apply. Therefore, I would conclude that we partially achieved Objective 3. In retrospect, it is particularly regrettable that we did not ensure that Studies 1 and 2 in Publication 3 generated data that could be statistically compared, as such an analysis would have been at least partially suitable for clarifying the issue of construct validity.

Objective 4 pertained to evaluating how social context variables influence the control of facial expressions. Unfortunately, we did not succeed in inducing a significant sense of attractiveness in the participants in Publication 2. Informal feedback from the participants indicated that this aspect was challenging to assess based on pictures showing only the face, with the hair and neck obscured by an oval shape. Fischer and van Kleef (2010) argued that pictures are merely an abstraction of real social stimuli, and face-to-face interaction is much more powerful as a social stimulus than static images of faces. Here, too, the pilot study in Publication 3 is likely to be more informative, as it involves a real person as the interaction partner.

What I would do if I were to start again would be:

a) To invest time and effort in developing an emotional process theory that could be formulated a priori and experimentally supported. b) To invest significantly more time and effort into developing an experimental paradigm where participants exhibit genuine, spontaneous emotional facial expressions. c) To continue to engage intensively with data handling. d) To postpone the combination of EEG methods and automated assessment of facial expressions of emotion techniques to a later stage, and to devote more time and effort to foundational work. e) To design such a project in collaboration with other laboratories, each specializing in technical or theoretical aspects, but also working together on programming solutions for data handling and analysis strategies.

Considering these reflections and the insights gained, it becomes evident that while significant progress has been made, there is still much to be achieved in the field of facial expression research. The challenges encountered underscore the need for continued innovation and refinement in both – theoretical frameworks and experimental methodologies. Future research should focus on developing more nuanced emotional process theories and experimental paradigms that can better capture the complexity of genuine emotional experiences. This includes designing studies that not only induce and measure authentic emotional responses but also address the limitations of static or abstracted stimuli. Collaborative efforts with specialized laboratories, leveraging diverse expertise in technical and theoretical domains, will be crucial in overcoming these limitations.

Furthermore, integrating advanced data analysis techniques and exploring the interplay between emotional expressions and various contextual factors – such as social interactions, environmental settings, and individual differences – will be essential for enhancing the construct validity of our findings. As we move forward, it will also be important to establish and validate new benchmarks for automated assessment tools and ensure that these tools are tested across a wide range of real-world scenarios. By addressing these challenges and building upon the foundational work laid out in this dissertation, we can pave the way for more accurate and meaningful assessments of facial expressions, ultimately advancing our understanding of emotion and its expression in diverse social contexts and everyday interactions. This approach will not only refine the techniques used in emotion research but also contribute to a deeper understanding of how facial expressions are influenced by and interact with the broader social and psychological environment.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft [grant number RE 3721/2-1 to GR]. The author thanks Prof. Dr. Jan Wacker and Dr. Guillermo Recio for their supervision. The author thanks Espen Henken, Kristina Bien, Jeanne Desbuleux, My Nguyen, Elena Hackmann, Lilly Kramer, Christopher Dannheim, Daron Rother and Ronja Heintzsch for their help with data collection and analyses. We verified license agreements of iMotions before publishing the present studies.

Competing interest statement

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this work.

References

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews, 1*(1), 21-62.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest, 20*(1), 1-68.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition, 104*(3), 644-653.

Benecke, C., & Brauner, F. (2017). *Motivation und Emotion: psychologische und psychoanalytische Perspektiven.* Kohlhammer Verlag.

Beringer, M. (2018). Preregistration on https://osf.io/wh6rx/wiki/home/

Beringer, M., Spohn, F., Hildebrandt, A., Wacker, J., & Recio, G. (2019). Reliability and validity of machine vision for the assessment of facial expressions. *Cognitive Systems Research*, *56*, 119-132.

Beringer, M., Wacker, J., & Recio, G. (2022). Deliberate control of facial expressions in a go/no-go task: An ERP study. *Acta Psychologica*, 230, 103773.

Beringer, M., Wacker, J., & Recio, G. (under review). Interferences between facial expressions of joy and disgust depending on facial attractiveness. An ERP Study. *Journal of Psycho-physiology.*

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., . . . Zhao, W. (2015). *Automatic detection of learning-centered affective states in the wild.* Paper presented at the Proceedings of the 20th international conference on intelligent user interfaces.

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion, 16*(8), 1186.

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion, 30*(6), 1081-1106.

Cohn, J.F., & De la Torre, F. (2015). Automated face analysis for affective computing. In R.A. Calvo, S. DiMello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 131-150). New York: Oxford University Press

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE transactions on pattern analysis and machine intelligence, 23*(6), 681-685.

Cottrell, G. W., & Hsiao, J. H. (2011). Neurocomputational models of face processing. *The Oxford handbook of face perception*, 402-423.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of cognitive neuroscience, 14*(8), 1158-1173.

Darwin, C. (1998). *The expression of the emotions in man and animals*: Oxford University Press, USA.

Del Líbano, M., Calvo, M. G., Fernández-Martín, A., & Recio, G. (2018). Discrimination between smiling faces: Human observers vs. automated face analysis. *Acta psychologica, 187*, 19-29.

Dente, P., Küster, D., Skora, L., & Krumhuber, E. (2017). *Measures and metrics for automatic emotion classification via FACET.* Paper presented at the Proceedings of the Conference on the Study of Artificial Intelligence and Simulation of Behaviour (AISB).

Deriso, D., Susskind, J., Tanaka, J., Winkielman, P., Herrington, J., Schultz, R., & Bartlett, M. (2012). *Exploring the facial expression perception-production link using real-time automated facial expression recognition.* Paper presented at the Computer Vision–ECCV 2012. Workshops and demonstrations.

Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*, 135-168.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*(6), 643-647.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological science, 11*(1), 86-89.

Ekman, P. (1972). Universal and cultural differences in facial expression of emotion. In J. R. Cole (Ed.), *Nebraska Symposium on Motivation, 1971,* pp. 207-283. Lincoln, NE: Nebraska University Press

Ekman, P. (2009). Lie catching and microexpressions. *The philosophy of deception*, 118-133.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*(1), 88-106.

Ekman, P., & Friesen, W. (1978). *Manual for the facial action coding system*: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of nonverbal behavior, 6*(4), 238-252.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *J Pers Soc Psychol, 53*(4), 712-717.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology, 30*(1), 527-554.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin, 128*(2), 203.

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of personality and social psychology, 64*(1), 83.

Frank, R. H. (1988). *Passions within reason: the strategic role of the emotions*: WW Norton & Co.

Fridlund, A. J. (2014). *Human facial expression: An evolutionary view*: Academic Press.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*(2), 201.

Friesen, W. V., & Ekman, P. (1983). EMFACS-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco, 2*(36), 1.

Gray, J. R. (2004). Integration of emotion and cognitive control. *Current Directions in Psychological Science*, 13, 46–48.

Gross, J. J., & Thompson, R. A. (2007). Emotional regulation - conceptual foundations. In J. J. Gross (Ed.), *Handbook of Emotion Regulations*: New York: Guildford Press.

Hajcak, G., Weinberg, A., MacNamara, A., & Foti, D. (2011). ERPs and the study of emotion. *Oxford Handbooks Online*. doi: 10.1093/oxfordhb/9780195374148.013.0222.

Hess, U., Kappas, A., McHugo, G. J., Kleck, R. E., & Lanzetta, J. T. (1989). An analysis of the encoding and decoding of spontaneous and posed smiles: The use of facial electromyography. *Journal of nonverbal behavior, 13*(2), 121-137.

Hietanen, J.K., Leppänen, J.M., Peltola, M.J., Linnaaho, K., & Ruuhiala, H.J. (2008). Seeing direct and averted gaze activates the approach-avoidance motivational brain systems. *Neuropsychologia*, 46(9), 2423–30.

Hooker, C. I., & Knight, R. T. (2006). in the inhibitory control of emotion. *The orbitofrontal cortex*, 307.

James, W.: „What is an emotion?“. *Mind 19* (1884), 188–204.

Kappas, A., Bherer, F., & Thériault, M. (2000). Inhibiting facial expressions: Limitations to the voluntary control of facial expressions of emotion. *Motivation and Emotion, 24*(4), 259-270.

Korb, S., Grandjean, D., & Scherer, K. R. (2010). Timing and voluntary suppression of facial mimicry to smiling faces in a Go/NoGo task—An EMG study. *Biological psychology, 85*(2), 347-349.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior research methods, 49*(2), 457-470.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and brain sciences, 35*(3), 121-143.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). *The computer expression recognition toolbox (CERT).* Paper presented at the Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.

Malmir, M., Forster, D., Youngstrom, K., Morrison, L., & Movellan, J. (2013). *Home alone: Social robots for digital ethnography of toddler behavior.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision Workshops.

Masaki, H., & Sommer, W. (2012). Cognitive neuroscience of motor learning and motor control. *The Journal of Physical Fitness and Sports Medicine, 1*(3), 369-380.

Matsumoto, D., & Hwang, H. C. (2014). Judgments of subtle facial expressions of emotion. *Emotion, 14*(2), 349.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annu. Rev. Psychol., 59*, 507-536.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

Morecraft, R. J., Stilwell–Morecraft, K. S., & Rossing, W. R. (2004). The motor cortex and facial expression:: New insights from neuroscience. *The neurologist, 10*(5), 235-249.

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes-Ker, A. H. (2000). Emotional state and the detection of change in facial expression of emotion. *European Journal of Social Psychology, 30*(2), 211-222.

Ouyang, G., Sommer, W., Zhou, C., Aristei, S., Pinkpank, T., & Abdel Rahman, R. (2016). Articulation Artifacts During Overt Language Production in Event-Related Brain Potentials: Description and Correction. Brain Topography, 29(6), 791-813. doi:10.1007/s10548-016-0515-1

Pessoa, L., Padmala, S., Kenzer, A., & Bauer, A. (2012). Interactions between cognition and emotion during response inhibition. *Emotion*, *12*, 192–197.

Recio, G., Shmuilovich, O., & Sommer, W. (2014). Should I smile or should I frown? An ERP study on the voluntary control of emotion - related facial expressions. *Psychophysiology, 51*(8), 789-799.

Rinn, W. E. (1991). Neuropsychology of facial expression.

Rosenbaum, D. A. (1980). Human movement initiation: specification of arm, direction, and extent. *Journal of Experimental Psychology: General, 109*(4), 444.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality, 11*(3), 273-294.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review, 69*(5), 379.

Scherer, K. R. (1992). *What does facial expression express?* Paper presented at the Parts of the argument in this chapter have been presented at the 36th congress of the German Society of Psychology in Berlin, 1988.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23*(7), 1307-1351.

Schmidt, K. L., Cohn, J. F., & Tian, Y. (2003). Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles. *Biological psychology, 65*(1), 49-66.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.

Susskind, J., Littlewort, G., Bartlett, M., Movellan, J., & Anderson, A. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia, 45*(1), 152-162.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature neuroscience, 11*(7), 843-850.

Swick, D., Ashley, V., & Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *Neuroimage, 56*(3), 1655-1665.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research, 168*(3), 242-249.

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion, 11*(4), 907-920. doi:10.1037/a0023853

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current directions in psychological science, 18*(3), 184-188.

Vogel, D. L., Heimerdinger-Edwards, S. R., Hammer, J. H., & Hubbard, A. (2011). “Boys don't cry”: Examination of the links between endorsement of masculine norms, self-stigma, and help-seeking attitudes for men from diverse backgrounds. *Journal of Counseling Psychology, 58*(3), 368.

Wacker, J. (2017). Increasing the reproducibility of science through close cooperation and forking path analysis. *Frontiers in Psychology, 8,* 1332.

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Transactions on Affective Computing, 5*(1), 86-98.

Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in clinical neuroscience, 17*(4), 457.

Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., & Fu, X. (2013). How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of nonverbal behavior, 37*(4), 217-230.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Erklärung gemäß *(bitte Zutreffendes ankreuzen)*

- ☐ **§ 4 (1c) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**
- ☒ **§ 5 (4d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich,

Matthias Beringer (Vorname, Nachname),

dass ich mich an einer anderen Universität oder Fakultät noch keiner Doktorprüfung unterzogen oder mich um Zulassung zu einer Doktorprüfung bemüht habe.

Hamburg, 21.10.24
Ort, Datum

M. Beringer
Unterschrift

Studien- und Prüfungsbüro Bewegungswissenschaft • Fakultät PB • Universität Hamburg • Mollerstraße 10 • 20148 Hamburg
Studien- und Prüfungsbüro Psychologie • Fakultät PB • Universität Hamburg • Von-Melle-Park 5 • 20146 Hamburg
www.pb.uni-hamburg.de

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Eidesstattliche Erklärung nach *(bitte Zutreffendes ankreuzen)*

☐ **§ 7 (4) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**

☒ **§ 9 (1c und 1d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich an Eides statt,

1. dass die von mir vorgelegte Dissertation nicht Gegenstand eines anderen Prüfungsverfahrens gewesen oder in einem solchen Verfahren als ungenügend beurteilt worden ist.
2. dass ich die von mir vorgelegte Dissertation selbst verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die wörtlich oder inhaltlich übernommenen Stellen habe ich als solche kenntlich gemacht.

Hamburg, 21.10.24
Ort, Datum

H. Bellinger
Unterschrift

Studien- und Prüfungsbüro Bewegungswissenschaft • Fakultät PB • Universität Hamburg • Mollerstraße 10 • 20148 Hamburg
Studien- und Prüfungsbüro Psychologie • Fakultät PB • Universität Hamburg • Von-Melle-Park 5 • 20146 Hamburg

· www.pb.uni-hamburg.de