DEPARTMENT of PSYCHOLOGY

Influence of State Empathic Concern and Personal Distress on Ratings of Dynamical Facial Expressions

Lisa Holzer

Master's Thesis (30 hp)
Spring 2021

Supervisor: Åse Innes-Ker

Author Note

First and foremost, I would like to thank Åse for contributing her expertise and helpful feedback to this thesis. Special thanks to Lola for her time and valuable comments on the experiment and to Blake for his academic and personal support in all respects.

Abstract

Inferring emotions from other people's facial expressions is essential for interpersonal functioning, but not everyone is equally capable. Facets of trait empathy (personal distress and empathic concern) influence emotion perception into opposing directions. Whether situational empathic states are similarly influential has not yet been studied and was the focus of this paper. A 3 (induction) x 2 (intensity) within-subjects online experiment was created, in which 82 participants read vignettes paired with emotional pictures to induce empathic concern, personal distress, and a neutral baseline condition. The order of the inductions was randomized and presented to each participant. Afterward, dynamical facial expressions were presented in two different intensities. Participants had to match each video with a label from a list of basic and complex emotions. Then they indicated their confidence in the rating. Afterward, they rated how positive or negative the expression was (valence) and how intense (arousal). Differences in categorical emotion ratings were found for expressive intensities, where high-intensity expressions matched the predefined emotion labels 17 % more often than low-intensity expressions. For empathic states, no significant differences in categorical emotion perception ability were found. Confidence ratings were higher in high-intensity expressions and the personal distress induction condition. Arousal ratings increased with expressive intensity. This study was the first to examine the connection between empathic states and emotion perception experimentally. It incorporated categorical and dimensional ratings to combine different approaches within the field of emotion research. Thereby novel insights into emotion perception were gained, which can guide future research.

Keywords: empathic state, empathic concern, personal distress, emotion perception, emotion labeling, core affect

Introduction

How is it possible to infer how others feel just from looking at them? And why do those judgments vary between situations and people? The presence of facial expressions in daily life is frequent, as are impairments to their correct identification. Understanding why and when some people have more difficulties recognizing another's emotional state can improve interventions and interpersonal skills. The current study examined if different empathic states influence how dynamical facial expressions are perceived.

Internal feeling states can be expressed via facial movements, and observers can infer emotions based on those expressions. Especially when expressions are ambiguous, intra- and interindividual differences feed into variations in the perception of expressions.

One variable tightly linked to emotion perception is *empathy*; it is a complex multidimensional construct with *affective* and *cognitive* components. Researchers have operationalized empathy in various domains but acquired mixed results. The realization that facets of empathy commonly lumped together predict a wide range of outcomes into opposing directions was essential to understand those contrary findings. For example, one paper that analyzed all 10 DSM-5 personality disorders found them to be positively related to *personal distress* but negatively related to *empathic concern*, two subscales of the affective component of empathy (Hengartner et al., 2014). Authors recently found a similar pattern regarding the ability to infer emotions from another person’s facial expressions (Israelashvili et al., 2020a; Mayukha et al., 2020). Empathic concern and personal distress were shown to diverge in their influence on how facial expressions were judged. While personal distress decreased the ability to categorize facial expressions, empathic concern increased this ability. Those findings were based on correlational designs, focusing on relatively stable empathic traits. However, as Fabi et al. (2019) emphasized, situational factors might be more critical for the experience of empathy, hence also for the judgment of facial expressions.

The current study experimentally induced states of empathic concern and personal distress and measured subsequent emotion perception; it thereby extended past correlational results of trait empathy and emotion perception and bridged the existing gap to empathic states. Because empathy and emotion perception are highly relevant for interpersonal functioning, every attempt to increase the understanding of those complex concepts and their interactions is valuable. Considering many decades of research and sustained ambiguity within the field, significant and non-significant results can be equally important. In what follows, the conceptualization and operationalization of empathy are introduced. Then different theories of emotion and their link to facial expressions are discussed. Lastly, the connection between empathy and emotion perception is elucidated, and variables that can impact this relationship are specified before detailing the research at hand.

Empathy

Essentially “empathy is to understand, feel, and share what someone else feels, with self-other differentiation” (Håkansson Eklund & Summer Meranius, 2021, p. 306). While *feeling* and *sharing* relate to empathy's emotional or affective component, *understanding*, and *self-other differentiation* can be linked to the cognitive component. Both affective and cognitive empathy can be distinguished on a behavioral and neural level. Behaviorally, autism spectrum disorder (ASD) and psychopathy are disorders linked to impaired social cognition. However, individuals with ASD commonly show impairments in cognitive empathy, whereas individuals with psychopathic traits show deficits in affective empathy (Lockwood et al., 2013). Neuronally, affective empathy shows activity in motor and somatosensory cortices, temporal poles, anterior cingulate cortices, amygdala, hypothalamus, and thalamus. In contrast, tasks that require cognitive empathy activate areas in the prefrontal cortex, superior temporal gyrus and junction, inferior parietal lobule, and fusiform gyrus (de Waal & Preston, 2017). Affective empathy is activated by automatic, bottom-up responses, while cognitive empathy evolves from more deliberate, top-down processes (de Waal & Preston, 2017).

The multidimensionality of empathy makes it difficult to develop reliable and valid instruments. One of the most commonly used instruments is Davis’ Interpersonal Reactivity Index (IRI; Davis, 1980, 1983). It consists of four subscales: personal distress, empathic concern, perspective taking, and fantasy. However, these have frequently been combined into two scales, measuring affective (personal distress, empathic concern) and cognitive (perspective taking, fantasy) components of empathy. This reduction likely contributed to inconsistent results and replication problems because the conflated scales operate in opposing directions. An analysis of the IRI's factor structure and psychometric properties also emphasized the importance of distinguishing between subscales (Chrysikou & Thompson, 2016; Jordan et al., 2016). Nevertheless, the concepts are connected and occur in succession so that affective empathic processes precede cognitive empathic processes (Israelashvili & Karniol, 2018). Therefore, the current study focused on the affective components of empathy and discriminated between the two affective facets, empathic concern, and personal distress. While empathic concern was conceptualized as experiencing "'other-oriented feelings of sympathy and concern for unfortunate others", personal distress was conceptualized as experiencing "'self-oriented feelings of personal anxiety and unease in tense interpersonal settings" (Davis, 1983, p. 114).

In addition to the difficult conceptualization, empathic features can be measured on a characteristic *trait* level and a situational *state* level. Importantly state and trait empathy do not have to point in the same direction, as often indirectly implied (Fabi et al., 2019). Today 80 % of studies measured trait empathy via self-reports, mainly with the IRI (34 %) (J. A. Hall & Schwartz, 2019). Instead, this study focuses on the different influences of state personal distress and empathic concern. It operationalizes both states with induction procedures (see *Method*) independently identified by two research groups (Fabi et al., 2019; Stellar et al., 2019) before measuring their influence on emotion perception.

Emotional Experience, Expression, and Perception

The interest to measure emotion perception requires a clear understanding of what emotions are. The American Psychological Association defines emotions as a “complex reaction pattern, involving experiential, behavioral, and physiological elements […]” (*APA Dictionary of Psychology*, 2006). This complexity leaves researchers until today with divergent understandings about the exact nature of emotions. *Basic emotion theory* (BET) and the theory of *core affect* are influential for emotion research and the base of emotional understanding in this paper.

Within BET, a limited set of distinct emotions are conceptualized as universally experienced and expressed due to their adaptiveness in evolution (e.g., *anger, disgust, fear, sadness*; for a review, see Tracy & Randles, 2011). “Look at the crying man; he must be sad” – who has not heard sentences like this before? Labels for specific emotions in language, frequently used in interactions, imply clear cuts between different emotional states (e.g., sadness) and distinct connections to their observable behaviors (e.g., crying). Paul Ekman contributed substantially to this line of research and emphasized the automaticity and universality of human emotions and their facial expressions (Ekman, 1992). Following BET, recognizing those facial constellations should be readily possible cross-culturally when emotional experiences are innate and expressed universally in peoples' faces (Ekman et al., 1972). Many lines of research assumed precisely such a link, and operationalizations to measure emotions and emotion perception are often based on those assumptions. To distinguish between different facial expressions, Ekman and his colleague developed the facial action coding system (FACS; Ekman & Friesen, 1978). Within FACS, separate facial muscular movements are identified as action units (AU). Multiple AUs combined represent specific expressions that are assumed to be connected to internal emotional states and rooted in evolution. When people are asked to identify which emotions are expressed in stimuli based on the FACS, their recognition rates are relatively high. However, those rates can vary with the intensity of the expressions so that high-intensity, compared to low-intensity stimuli, are relatively easy to identify (Hess et al., 1997; Wingenbach et al., 2016, 2018). Moreover, the number of answer choices given can alter emotion perception, with more options decreasing the overall recognition rate (Phillips et al., 2008). The current study chose the term *emotion perception* consciously over *emotion recognition* to emphasize the process by which a person personally perceives a face rather than recognizes a predefined label.

Contrary to the clearly distinguishable discrete emotions and facial expressions suggested by BET, dimensional theories are favored by some emotion researchers instead. One prominent opponent is Russell (1994). He criticized the methodology in Ekman and colleagues' studies that supported the BET view. Ekman used static pictures of posed expressions and forced answer choices of highly stereotypical expressions, thereby reducing ecological validity and distorting results as well as inferences drawn from them. Russell instead suggested the concept of *core affect*, a neurophysiological state that involves variations in valence and arousal (Russell, 2003). Within this theory, any state can be placed within a circumplex model based on the axis of pleasantness-unpleasantness (valence) and activation-deactivation (arousal). Emotional experiences can thus be expressed on a simpler core level and do not need an explicit emotion category assigned. *Anger*, for example, could be indicated on the dimension of pleasantness as relatively low and on the dimension of arousal as high without claiming the existence of a universally experienced, expressed, and recognized emotion *anger*, as claimed by BET.

Independently of the specific theory, internal emotional experiences, external expressions, and the perception of those expressions in others are linked. Behaviorally *facial mimicry*, an automatic process by which observers imitate the expression of others' facial expression (Hess & Fischer, 2013), triggers the same feeling within their own body, thereby increasing the understanding of what another person feels and expresses. Neuronally, so-called *mirror neurons* fire when performing an action but also when just observing it (Schmidt et al., 2021). Consequently, watching emotional expressions in someone else can trigger emotional experiences in oneself and supports the understanding of other's affective states and intentions (de Waal & Preston, 2017; Gallese, 2007).

Even though the exact nature and manifestation of emotional experiences in facial expressions are still debated, the importance of emotion perception (be it discrete or dimensional) in others' faces for interpersonal functioning is undeniable. A better understanding of factors that influence emotion perception from facial expressions is essential for clinical samples and the general population alike.

Empathy and Emotion Perception

The influence that empathy has on emotion perception was well described by Chakrabarti & Baron-Cohen (2011) as "[…] a lens through which we perceive emotions" (p. 5). Like an actual lens that improves the view for one person but not for the other, different empathic states might improve or distort emotion perception. Besides the easily labeled clear-cut expression stimuli, usually shown in full-blown intensity, natural expressions are far less intense and allow room for variations between individual ratings on the same expression. Different modes of thought, for example, showed variations in the perception of expressions (Ma-Kellams & Lerner, 2016). A systematic way of thinking improved, and an intuitive style decreased emotion perception. Moreover, whether individuals focus more on the eye or the mouth region can also influence how emotional expressions are perceived (Abbruzzese et al., 2019). Importantly, trait empathy has also been linked to emotion perception, with empathic concern being beneficial and personal distress detrimental (Israelashvili et al., 2020a; Mayukha et al., 2020). This link was also examined experimentally using reaction time measures, eye-tracking, event-related potentials (ERPs), and electromyography (EMG). Those higher on trait empathy (measured with the balanced emotional empathy scale; Mehrabian & Epstein, 1972) reacted faster to emotional stimuli, fixated their gaze more on critical facial features, showed higher cortical reactivity (N200), and had more intense facial muscle movements coherent with the presented emotional expression (Balconi & Canavesio, 2016). In another study, Gery et al. (2009) compared sex offenders with non-sex offenders. They found elevated personal distress and low empathic concern in sex offenders, and only those facets differentiated them from non-sex offenders in their ability to recognize emotions. All those results indicate a link between trait empathy and emotion perception. Whether state empathy is similarly linked to emotion perception remained uncertain.

Remember the crying man from the previous section? Does seeing him make observers sad, too? If so, then most likely affective empathy is at its play. The specific reaction eventually shown towards the man depends on trait-level individual differences (including trait empathy) and state-level situational factors, possibly also state empathy. While one person might be overwhelmed and distressed by the situation, another might feel concerned and ready to help. The question if those states differently affect subsequent emotion perception in peoples' faces was the focus of this study. It was assumed that empathic concern would improve, and personal distress would decrease emotion perception, similar to studies by Israelashvili et al. (2020a) and Mayukha et al. (2020). The first leads to outward-directed attention to comfort the other person and the latter to inward-directed attention to settle the feelings of distress in oneself (Zaki, 2014), hence the difference in emotion perception accuracy.

Other Influential Variables

The integral role that empathy plays in emotion perception from facial expressions was discussed before. This section introduces other variables that were measured because they can influence empathy and emotion perception.

Age and *gender* were frequently reported as influential for emotion perception (Abbruzzese et al., 2019; Dores et al., 2020). Olderbak et al. (2019) found that between the age of 15 – 30 years, the general ability of emotion perception across all discrete emotions constantly increases and declines afterward. For gender Wingenbach, Ashwin, et al. (2018) found a general female advantage in emotion perception independent of the specific emotion categories. Attentional focus to eye or mouth region also varied with sex (J. K. Hall et al., 2010; Sullivan et al., 2017) and age (Sullivan et al., 2007) and influenced emotion perception.

Likewise, *affective states* can influence empathic reactions (Nezlek et al., 2001) and emotion perception (Schmid et al., 2011). The induction of happy, neutral, and sad states biased the response tendencies of ambiguous faces towards the mood-congruent end (Trilla et al., 2021). Similarly, Harris et al. (2016) found that more positive states biased female faces' perception in a more positive direction.

Emotion regulation styles can also impact empathic reactions and emotion perception. Overall, individuals show strong dispositions to either reappraise (positive) or suppress (negative), and both styles are opposingly correlated (as indicated in brackets) to variables like affective functioning, depression, self-esteem, or life-satisfaction (Gross & John, 2003). Additionally, different regulation styles were found to moderate the relationship between empathy and affective distress (Powell, 2018) as well as prosocial behavior (Lockwood et al., 2014).

The Current Research

The current work examined whether empathic states influence how dynamic facial expressions that varied in intensity (high, low) and complexity (e.g., happiness, pride) are perceived. Based on previously described studies that assessed facets of trait empathy and emotion perception accuracy, I hypothesized that state empathic concern would improve emotion perception accuracy. In contrast, state personal distress would decrease it. A neutral induction served as a baseline condition. Furthermore, I assumed that more subtle, low-intensity emotional expressions would result in higher ambiguity and, therefore, lower emotion perception than high-intensity expressions. The experimental sample consisted of healthy adults rather than a clinical sample with impaired social cognition abilities, therefore high-intensity emotions were assumed to result in equally well perception among induction conditions. In contrast, low-intensity expressions were assumed to show differences in emotion perception between state empathic concern and state personal distress as previously described.

Empathic traits were measured to replicate previous findings. Consistent with prior results, trait personal distress was assumed to correlate negatively and trait empathic concern positively with emotion perception. The relationship between those measures was assessed exploratorily because it remains unclear whether trait and state empathy are correlated (Fabi et al., 2019; Mayukha et al., 2020)**.**

Additional analyses were conducted for ratings of confidence, valence, and arousal. A study where participants first had to decide if sentences convey an emotional message and then rate the perceived intensity and confidence found a positive correlation between those variables (Troiano et al., 2021). Therefore, I assumed that confidence and arousal ratings would increase with high-intensity expressions and subsequently also influence emotion perception. Variations between induction conditions (state empathy) and dimensional ratings (confidence, valence, arousal) were explored without prior hypotheses.

Method

Design

The experiment comprised a 3 (induction) x 2 (intensity) repeated measures design where each participant was exposed to three experimental conditions in randomized order. The conditions consisted of inductions of different empathic states (empathic concern, personal distress) and a neutral control condition. For the induction, it was essential to use stimuli that elicit empathic reactions without prior facial expression identification to measure the emotion perception accuracy independently after the empathic inductions. Physical pain, emotional situations, and emotional faces can evoke empathic reactions (Ding et al., 2020). Utilizing the opposing effects physical and mental pain conditions have on individuals (Fabi et al., 2019; Stellar et al., 2019), empathic concern and personal distress were induced without any need for the ability to identify facial expressions. Hence, general emotion perception abilities should not have affected the induction success, making an independent measure of emotion perception possible. After each induction, participants rated a series of videos depicting dynamic facial expressions in two different intensities (high, low). The experiment was designed with plugins from the open-access library *jsPsych* (de Leeuw, 2015). Data collection was implemented online via *Cognition*, a platform to conduct experimental research online based on Javascript code and the jsPsych library. It took three weeks between March and April 2021 to collect data from the required sample size of 80 participants. The overall hit rate (oH) includes the ratio of correct ratings to total ratings accumulated for all emotions and represents the study's dependent variable. The significance level was set to $\alpha = 0.05$.

Participants

Before data collection, the sample size was planned with a simulation-based power analysis in R (Caldwell & Lakens, 2019; Lakens & Caldwell, 2019). Values for the simulation were taken from the unbiased hit rates in the study by Wingenbach et al. (2016). The mean value of the oH was set to 63 % for all conditions (empathic concern, personal distress, neutral) because inductions were predicted not to influence the perception of high-intensity emotional expressions. For the low-intensity condition, the mean value of 43 % was only taken for the neutral condition. Empathic concern was hypothesized to increase (*M* = 46 %) and personal distress to decrease (*M* = 40 %) the oH. Wingenbach et al. (2016) reported a standard deviation of *SD* = 11.51 % for low-intensity and *SD* = 12.54 % for high-intensity facial expressions, which were adopted for the simulation. The correlation between conditions was set to r = .50. Plotting this design revealed a needed sample size of N = 80 to find an effect with 80 % expected power. With N = 80 added to the design, 10,000 simulations with an α-level of 5 % were performed, and post hoc multiple comparisons were corrected with Holm's method. The simulation resulted in an expected power of 100 % for the intensity (high, low), 78.95 % for the condition (empathic concern, personal distress, neutral), and 78.20 % for the interaction term. Post hoc analysis compared induction conditions for the low-intensity expressions and found an expected power of 97.40 % and an average effect size of Cohen's f = 0.53.

The experiment was advertised on social media (Facebook, Reddit, WhatsApp), where participants were encouraged to redistribute the study, thereby creating a diverse sample. Criteria for participation were fluency in English and aged 18 or above; no other exclusion criteria were placed. Participants were instructed to complete the experiment on a computer or laptop and not on their phones for compatibility reasons. The total sample consisted of 82 participants, out of which n = 50 identified as female, n = 30 as male, and n = 2 as non-binary/third gender. The sample's mean age was 29.11 years (*SD* = 9.64), with a minimum age of 18 and a maximum age of 68.

Ethical Considerations

The experimental design and implementation was guided by the Declaration of Helsinki (WMA Declaration of Helsinki, 2013). Ethical approval for the study from the internal review board was not obtained because the empathic inductions contained short scenario descriptions and low-intensity pictures, unlikely to upset or disturb participants. Therefore, the influence on participants was expected to be short-term, superficial, and not harmful. All participants received detailed information about the study and their rights and gave their consent before the experiment.

Empathy Induction

Empathic concern and personal distress were induced with the method tested by Fabi et al. (2019). Participants read a descriptive text of a situation followed by pictures related to that story. The texts described a mental pain scenario (neighbor crying because his/her mum died suddenly) to induce empathic concern and a physical pain scenario (neighbor suddenly complaining about violent pain in his/her chest) to induce personal distress. Additionally, a neutral condition was presented (saying 'hi' to the neighbor) as a baseline measure. After each vignette, eight pictures were presented to reinforce the empathic response. In the following sections, induction conditions will directly refer to the empathic states (empathic concern, personal distress, neutral) rather than the induction method (mental pain, physical pain, neutral). Participants completed a short empathic response scale consisting of three adjectives for empathic concern and three for personal distress as a method to control the induction's success. The procedure was previously applied and validated in a German sample by Fabi et al. (2019).

Emotion Perception

After scanning through the KAPODI database (Diconne et al., 2021), a comprehensive collection of emotional stimuli, the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV), was selected as the best suitable stimuli for the aims of the current research because it includes dynamic expressions in varying intensities. The ADFES-BIV consists of 370 posed video clips, each with a length of 1400 milliseconds. It shows nine different emotional facial expressions (anger, contempt, disgust, fear, embarrassment, joy, pride, sadness, surprise) and a neutral condition in three different intensities (low, medium, high). All video clips start with a neutral expression and progress to the intended expression (Wingenbach et al., 2016). Five actresses and seven actors performed each emotion in three different intensities. For this study, only high and low-intensity videos of each facial expression and two neutral videos (to keep the ratio equal) were included per actor and actress, resulting in a total number of 240 video clips. Additionally, ten practice videos with a separate actress were provided to familiarize participants with the task. In the experiment, the participants selected an emotion label for each video and rated their confidence in this selection. Requesting *categorizations* from multiple emotion labels and asking for *confidence* in those selections widened the scope of otherwise limited labeling tasks. As in Wingenbach et al. (2016), hit rates were calculated separately for each induction condition by expressive intensity. Each expression category was rated correctly above chance level (> 10 %), ranging between 16 % (for pride, neutral*low) and 96 % (for surprise, personal distress*high). Additionally, they rated each facial expression for perceived valence and arousal. The application of *dimensional ratings* was beneficial because they do not rely on previously agreed-upon emotion labels. These considerations enhanced prior methodological constraints and allowed a detailed examination of emotion perception from facial expressions.

Questionnaires

The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) is a 20 item self-report questionnaire measuring the momentary affective states of participants. This study used the shortened version with only ten items (I-PANAS-SF; Thompson, 2007). Each item was rated on a 5-point Likert scale, and the anchors were adopted to those used in the original PANAS, ranging from 1 (very slightly or not at all) to 5 (extremely). The short form showed acceptable validity of .59 on the positive affect subscale and .65 on the negative affect subscale compared to the original 20 item scale. Internal consistency indicated by Cronbach's alpha was acceptable in the I-PANAS-SF with .78 for positive and .76 for negative affect (Thompson, 2007b), with slightly decreased reliability in the current study, .65 and .70 respectively.

As a manipulation check, the shortened Empathic Response Scale (ERS) applied and validated in German by Fabi et al. (2019) was used in its English translation. It was applied three times throughout the experiment, after each induction condition. Personal distress was assessed with items 'alarmed', 'distressed' and 'compassionate' and empathic concern was assessed with items 'worried', 'moved', and 'tender' on a 5-point Likert scale ranging from 1 (not at all) to 5 (very much). Internal consistency indicated by Cronbach's alpha was acceptable for both empathy inductions, the mental pain condition (empathic concern .84; personal distress .79) and the physical pain condition (empathic concern .75; personal distress .84) but reduced for the neutral condition (empathic concern .65; personal distress .73).

The Emotion Regulation Questionnaire (ERQ; Gross & John, 2003) is a 10-item self-report questionnaire measuring two emotion regulation strategies (cognitive reappraisal, expressive suppression) on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Cronbach's alpha was acceptable for both subscales with .79 for reappraisal and .67 for suppression.

The Interpersonal Reactivity Index (IRI; Davis, 1980, 1983) is the most widely used self-report measure (for a review see: Hall & Schwartz, 2019) for dispositional empathy. It consists of 28 items and includes four scales (perspective taking, fantasy, empathic concern, personal distress). In this study, a shortened version with 16 items was used (B-IRI; Ingoglia et al., 2016). It showed the same component structure as the original scale, demonstrating its validity and proving reliability for all subscales (empathic concern .68; personal distress .72; perspective-taking .68; fantasy .79). Participants rated on a 5-point Likert scale ranging from 1 (does not describe me at all) to 5 (describes me very well). In the current study, Cronbach's alpha showed questionable values of .58 for *empathic concern* and .55 for *personal distress*, which could not be improved by deleting any item in the respective subscales. The other two subscales had an acceptable internal consistency of .77 for *fantasy* and .74 for *perspective-taking*. The implementation of this questionnaire allowed for the replication of prior correlational findings between empathic traits and emotion perception.

The application of the ERQ and the B-IRI within the experiment had a two-fold function. Primarily they were utilized to measure emotion regulation styles and empathic traits. Additionally, they provided the opportunity to downregulate previously triggered empathic states to a baseline level. This was inevitable to guarantee independent induction effects despite the repeated-measures design.

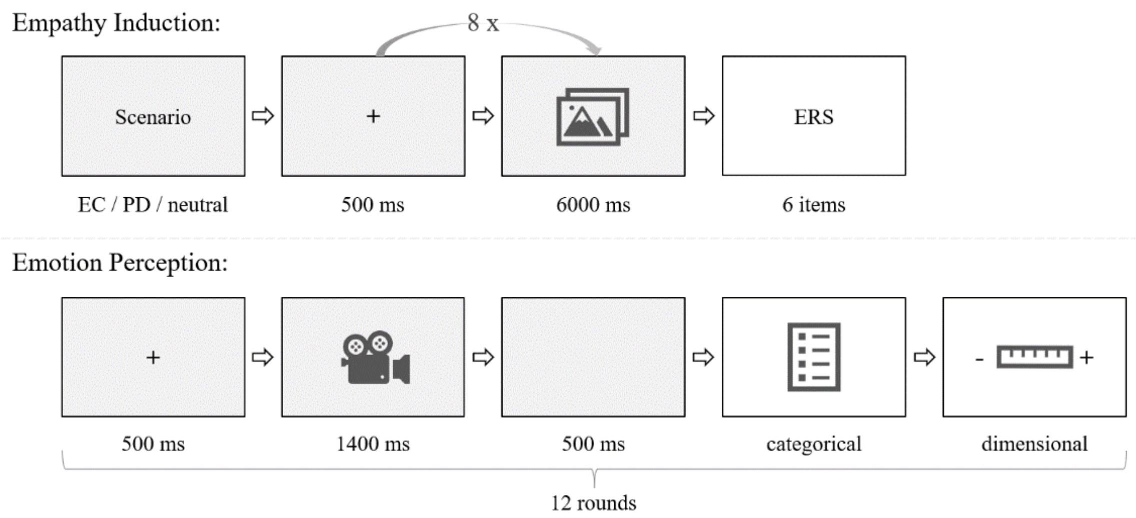
Procedure

First, participants were informed about the content of the study and that their emotional state could be temporarily altered by the materials they were about to see. Additionally, they received information about their rights and contact details of the researchers before giving their consent to participate. A welcome page emphasized that the experiment should be executed on a computer or laptop in a quiet environment. Then participants adjusted a square so that the stimuli of the study were customized for their screen size and filled in age and gender information. To check their current affective states, the I-PANAS-SF was answered before the experiment started.

For the main part of the experiment, which consisted of three blocks, a start screen with all possible emotion labels and definitions was provided. Those who were uncertain about answer options had the opportunity to improve their understanding. Furthermore, each participant performed two practice rounds of facial expression ratings (bottom of *Figure 2*) to clarify the task. Those videos were randomly picked from the ten practice videos provided in the ADFES-BIV, thereby avoiding increased familiarity with any specific actor or actress later in the experiment. The first block randomly picked an induction condition (empathic concern, personal distress, neutral) and presented the respective written scenario vignette, followed by eight presentations of a fixation cross (500 ms) coupled with a picture (6000 ms) corresponding to the vignette. Afterward, participants answered the ERS to check for induction success before displaying 12 consecutive facial expression videos and rating scales. The videos were randomized for the sex of the actors and actresses (six females, six males) and their intensity (three males*low, three males*high, three females*low, three females*high). Each video was preceded by a fixation cross (500 ms) and followed by a white screen (500 ms). After each video, participants had to pick an emotion label from the provided list: anger, contempt, disgust, embarrassment, fear, joy, neutral, pride, sadness, surprise, to indicate the emotion that they perceived from the facial expression. Alternatively, they could indicate “I do not know” (IDK) if they could not identify the expression. Then they were asked to indicate the confidence in their rating from 1 (not confident at all) to 5 (very confident). Afterward, participants also rated how positive or negative they perceived the expression from 1 (very negative) to 5 (very positive) and how intense from 1 (not intense at all) to 5 (very intense). The graphical depiction of the experimental procedure of empathy induction and emotion perception is presented in *Figure 2*.

Figure 2.

Graphical depiction of the empathic induction procedure (top) and the emotion perception procedure (bottom).



Note. After each induction, 12 videos were presented and judged on a categorical list including nine emotion labels and IDK and on three dimensional scales for confidence, valence, and arousal.

After the first block, one randomly selected questionnaire (ERQ, B-IRI) was presented to downregulate the emotional state prevalent from the first induction condition. Then the second block started by randomly presenting one of the two remaining induction conditions and showing 12 randomized video clips with the subsequent rating scales.

After the second block, the other questionnaire (ERQ, B-IRI) was presented, and the third block started by repeating the same procedure as described above.

In the end, participants had the option to write in a free text entry field if they were distracted or if any irregularities happened during the experimental procedure. This option was implemented to account for unclean data and clarify if technical problems arose during the experimental procedure. Lastly, participants were debriefed, thanked for their participation, and provided contact details if they had any comments or questions regarding the experiment and the research study.

Analysis

The collected data was prepared in R (Team, 2021) and analyzed with R and IBM SPSS Statistics (Version 26). Several peculiarities emerged that need to be considered in the analysis and are therefore described in the following section.

During the experiment, participants saw expressions balanced for intensity and sex of the actors and actresses but not balanced for specific emotions. Overall, each expression was presented approximately 10 % of the time, achieving desirable randomization. However, the same balance was not given for individuals because each participant saw only 36 emotional expressions, meaning some expressions were seen more often and others never by a single participant. Therefore available ratios were sampled to facilitate follow-up analyses on specific emotional expressions. Another post-hoc analysis compared the emotion perception of participants with strong and weak induction successes.

Dependent variable. All emotion label ratings were coded 0 for a response that did not match the presented emotional expression and 1 for matches. An additional group indicating IDK was also created because it cannot be considered right or wrong. Afterward, for each facial expression, a ratio of $\frac{\textbf{\textit{number emotional expression rated correctly}}}{\textbf{\textit{number emotional expression presented}}}$ was calculated, consequently referred to as the hit rate, and accumulated for all emotional expressions, referred to as the overall hit rate (oH).

Induction. Repeated measures ANOVAs were calculated to assess induction success. Higher values in personal distress than empathic concern in the physical pain induction condition and the opposite pattern for the mental pain induction condition indicated success.

Intensity. High- and low-intensity expressions were compared in the analyses, therefore hit rates were calculated separately for high and low-intensity expressions. The neutral expressions constitute an exception, as, by definition, they do not have varying intensities. Those were excluded and analyzed separately.

Main analysis. A 3 x 2 repeated measures ANOVA was calculated, with inductions (empathic concern, personal distress, neutral) and intensities (high, low) as independent variables and oH as the dependent variable.

Follow-up analyses. Follow-up analyses for the hit rates of individual emotional expressions were calculated because nine emotional expression groups were condensed in the primary analysis. The experiment balanced randomization for sex and intensity but not for specific emotional expressions, therefore a subset of the data was taken per emotional expression accounting for varying inductions and expressive intensities (n = 29) to run ANOVAs without missing cells. Even though the sampling procedure did not allow repeated measures analyses, the oH showed comparable results to the previous analysis accounting for repeated measures. Therefore, two-way ANOVAs were conducted for each emotional expression separately.

Additionally, participants were divided into two groups based on responses on the empathic response scale (ERS, measuring empathic states). In the personal distress condition, participants were divided at the mean of the personal distress ratings on the ERS. In the empathic concern condition, participants were divided at the mean of their empathic concern ratings on the ERS.

Secondary and exploratory analyses. Additionally, participants indicated their confidence in the emotional expression rating on a 5-point Likert scale. Two two-way ANOVAs were performed separately for correct and incorrect responses to assess if those confidence ratings varied with different induction conditions and different intensities of the facial expressions. Emotional expressions rated with the IDK option were excluded from this analysis because no logical confidence ratings can be given. As an alternative to the categorical selection of specific emotions, a dimensional *valence* and *arousal* rating was given for each expression on a 5-point Likert scale. To see if those ratings varied with different induction conditions and expressive intensities, two two-way ANOVAs were conducted. This time IDK ratings were included in the analysis because dimensional ratings do not require agreement in a predetermined emotion label for emotional expressions. That means, even when the emotional expression presented in the video was not recognized, the participant can still judge on the personally perceived valence and arousal of the respective expression.

Results

Questionnaires

The data collected with questionnaires was examined for irregularities. After assessing distributions, means, and standard deviations, no influential outliers were found. Generally, participants rated their affective state as more positive (M = 2.94, SD = 0.70) than negative (M = 1.55, SD = 0.61) and emotion regulation scores were higher for reappraisal than for suppression (see *Table 1*). Trait empathic concern was rated higher than trait personal distress. However, internal consistencies for those two subscales measured with Cronbach's α were low. All items of the B-IRI were positively correlated (exception perspective-taking and personal distress), as were perspective-taking (B-IRI) and reappraisal (ERQ). Empathic concern (B-IRI) and suppression (ERQ) were negatively correlated.

Table 1.

Correlation matrix, with means, standard deviations, and Cronbach's alphas.

	M	*SD*	1	2	3	4	5	6	7
1 ERQ_RA	4.63	1.20	(.79)						
2 ERQ_SP	3.18	1.16	.13	(.67)					
3 B-IRI_PD	2.87	0.88	-.03	-.13	(.55)				
4 B-IRI_EC	3.98	0.69	.07	-.28*	.32**	(.58)			
5 B-IRI_FS	3.41	1.03	.09	-.05	.22*	.49**	(.77)		
6 B-IRI_PT	3.63	0.83	.29**	-.03	.05	.33**	.28*	(.74)	
7 age	29.19	9.75	.04	-.02	.18	.19[a]	.31**	.21[a]	-

Note. ERQ = emotion regulation questionnaire (RA = reappraisal, SP = suppression); B-IRI = brief interpersonal reactivity index, in this context referring to the brief version (PD = personal distress, EC = empathic concern, FS = fantasy, PT = perspective taking). In brackets presented are Cronbach's α. * indicates p < .05, ** p < .01, [a] marginally significant.

Gender Differences

T-tests were run to compare differences between genders on the applied questionnaires. Only for trait *personal distress* male participants had significantly higher ratings (n = 30; M = 3.21; SD = 0.68) than female participants (n = 50; M = 2.68; SD = 0.70), $t(78) = 2.72$; $p = .008$, $d = .77$. However, when comparing those ratings to the overall state empathy ratings (not controlling for induction conditions), females consistently scored higher in empathic concern (female: M = 2.13; SD = 1.16; male: M = 1.93; SD = 1.03) and also higher in personal distress (female: M = 1.60; SD = 1.30; male: M = 1.35; SD = 1.22) but those differences did not reach significance. When analyses were calculated separately for induction conditions females only scored significantly higher in state empathic concern in the physical pain condition $t(78) = 2.13$; $p = .036$; $d = .49$. T-tests for emotion perception accuracy did not show any gender differences. The participants' ages were not sufficiently spread to conduct meaningful analyses and were disregarded.

Induction Success

The different empathic states were successfully induced. Two repeated-measures ANOVAs determined that personal distress and empathic concern ratings differed significantly among the induction conditions, for personal distress $F(1.84, 148.83) = 86.84$, $p < .001$, $\eta^2_p = .52$, and for empathic concern $F(1.73, 139.70) = 70.45$, $p < .001$, $\eta^2_p = .47$. Bonferroni corrected post-hoc test revealed a significant difference between all conditions, with higher personal distress ratings in the physical pain condition and higher empathic concern ratings in the mental pain condition, indicating successful empathy inductions. The neutral condition resulted in the lowest results in both ratings (see *Table 2*).

Table 2.

Estimated means and standard errors for empathic concern and personal distress ratings per induction condition.

	EC rating			PD rating		
Induction	*Mean*	*SE*	*95% CI*	*Mean*	*SE*	*95% CI*
1 (EC)	3.70	0.11	[3.47, 3.92]	2.55	0.13	[2.30, 2.79]
2 (PD)	3.17	0.11	[2.95, 3.39]	3.29	0.14	[3.02, 3.57]
3 (neutral)	2.28	0.09	[2.10, 2.45]	1.63	0.09	[1.46, 1.81]

Note. Presented in rows are the three induction conditions: empathic concern for the mental pain condition, personal distress for the physical pain condition, and the neutral condition.

Influence of Expression Intensity and Induction Condition on Emotion Perception

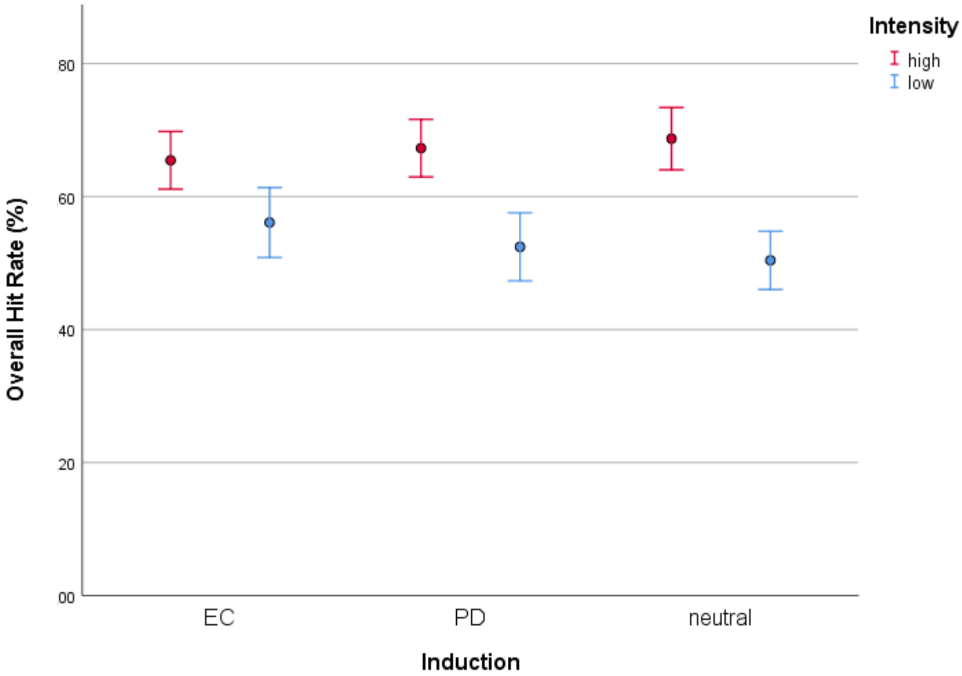
A repeated-measures ANOVA for the oH showed a significant main effect for intensity $F(1, 81) = 62.71$, $p < .001$, $\eta^2_p = .44$, but not for induction condition. The interaction between induction and intensity showed a small significant effect $F(2, 162) = 3.94$, $p = .021$, $\eta^2_p = .05$. Post-hoc analysis revealed an on average 17 % higher emotion perception accuracy for high compared to low-intensity expressions $p < .001$. This also applies to the interaction effects, where all inductions paired with low-intensity expressions had lower accuracy than all inductions paired with high-intensity expressions (p's $< .05$). For the comparisons of different induction conditions with the same intensity, no significant differences were found. *Figure 3* shows oH separately for induction condition and expressive intensity.

Because nine different emotional expression groups were accumulated in the above analysis, an exploratory follow-up analysis with each emotional expression was conducted separately. The analysis showed significant results for the main effect of intensity (p's $< .05$

for all emotions but *disgust*, *sadness*, *surprise*) but no main effect for induction condition nor an interaction effect for any of the emotions.

Figure 3.

Graphical depiction of correct ratings in % per empathic induction condition and expressive intensity.



Note. Red markers show the overall hit rates of high-intensity expressions, and blue markers the overall hit rates of low-intensity expressions. Induction conditions on the x-axis are empathic concern (EC), personal distress (PD), and neutral. Error bars represent 95 % confidence intervals.

Lastly, exploratory post-hoc analyses were conducted to compare overall hit rates for participants that indicated higher or lower than average empathic reactions. Groups were split at the mean of the respective scale. One independent t-test assessed differences in the oH based on personal distress ratings in the physical pain condition, which was supposed to induce personal distress. For high-intensity facial expressions the oH did not differ much between the high (*M* = 67.00 %, *SD* = 22.11 %) and the low (*M* = 66.18 %, *SD* = 21.56 %) personal distress groups. For low-intensity facial expression, the oH was lower for the high personal distress group (*M* = 49.11 %, *SD* = 24.18 %) than the low personal distress group (*M* = 52.37 %, *SD* = 23.62). Another t-test assessed differences in the oH based on empathic concern ratings in the mental pain condition, which was supposed to induce empathic concern. The high-intensity expressions were not differing between high (*M* = 65.82 %, *SD* = 22.62 %) and low (*M* = 65.16 %, *SD* = 20.44 %) empathic concern groups. However, for the low-intensity expressions, the oH was higher for the high empathic concern group (*M* = 57.10 %, *SD* = 27.59 %) and lower for the low empathic concern group (*M* = 50.75 %, *SD* = 21.76

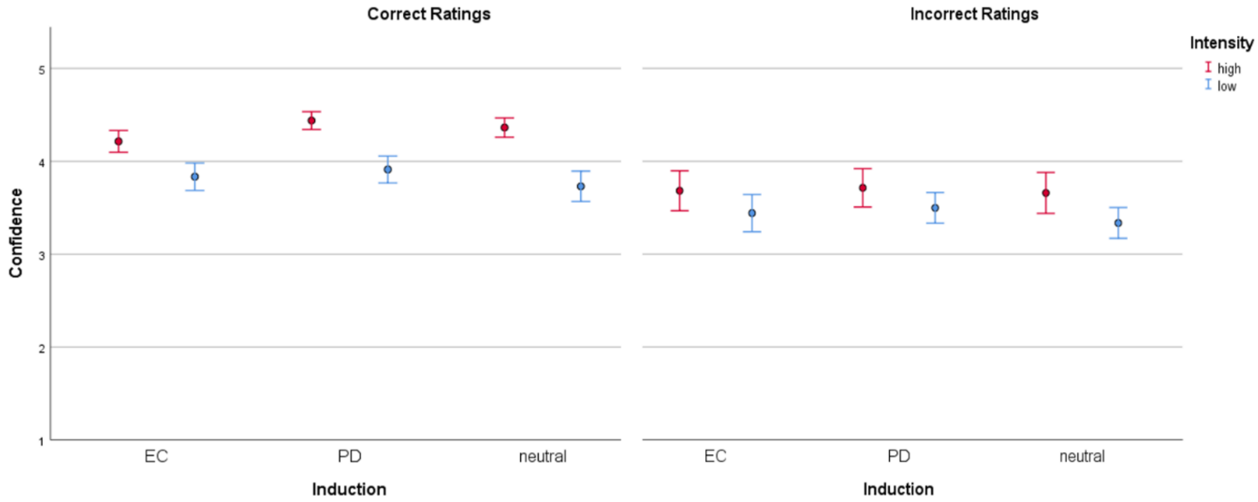
%). All values were non-significant but showed an interesting trend in the hypothesized direction.

Confidence in Selected Expression Labels and Ratings of Valence and Arousal

To assess how confidence, valence, and arousal ratings were linked to different expressive intensities and different induction conditions, several two-way ANOVAs were run. Because confidence ratings were significantly higher for correct (M = 4.12, SD = 1.07) relative to false ratings (M = 3.49, SD = 1.21), $t(2765) = 14{,}15$, $p < .001$, $d = .55$ two separate ANOVAs were calculated. The analysis for correct ratings (see *Figure 4 left*) showed significant main effects for both, intensity $F(1, 1544) = 94.62$; $p < .001$, $\eta^2_p = .11$ and induction condition $F(2, 1544) = 3.37$, $p = .035$, $\eta^2_p = .002$. Tukey's post-hoc tests showed that on average high-intensity expressions were rated 0.51 points higher for confidence than low-intensity expressions $p < .001$. Confidence ratings in the personal distress induction condition were 0.16 higher relative to the empathic concern condition $p = .033$. The analysis for incorrect ratings (see *Figure 4 right*) showed a significant main effect only for expressive intensity $F(1, 910) = 15.19$, $p = .001$, $\eta^2_p = .02$. A post-hoc test indicated that high-intensity expressions were rated 0.26 points higher on confidence than low-intensity expressions.

Figure 4

Graphical depiction of confidence ratings per empathic induction condition and expressive intensity. On the left side for correct ratings and on the right side for incorrect ratings.



Note. Red markers show the average confidence rates of high-intensity expressions, and blue markers the average confidence rates of low-intensity expressions (ratings were made on a scale from 1-5). Induction conditions on the x-axis are empathic concern (EC), personal distress (PD), and neutral. Error bars represent 95 % confidence intervals.

For valence and arousal analyses were calculated independent of the correctness of the labeling task. For valence there were no significant effects, while for arousal only intensity had a significant effect $F(1, 2623) = 427.98$, $p < .001$, $\eta^2_p = .14$. Figures of the valence and arousal ratings can be found in Appendix A (*A1* for arousal and *A2* for valence). Tukey's post-hoc test showed that high-intensity expressions were rated 0.95 points higher for arousal than low-intensity expressions p < .001.

Discrete Emotion Perception

Confusion matrices were created for the primary 3 (induction) x 2 (intensity) hypothesis and are displayed in *Tables 3* (high-intensity) and *4* (low-intensity).

Table 3.

Confusion matrix for high-intensity emotional expression ratings (values in %).

	Perceived emotion											
Presented emotion	**Ang**	**Con**	**Dis**	**Emb**	**Fea**	**Joy**	**(Neu)**	**Pri**	**Sad**	**Sur**	**IDK**	**Ind**
Ang	**86,89**	4,92	1,64	0	0	0	0	0	1,64	3,28	1,64	*N*
	79,25	5,66	9,43	1,89	1,89	0	0	0	0	1,89	0	*EC*
	80,49	4,88	4,88	2,44	2,44	0	0	0	0	4,88	0	*PD*
Con	7,69	**36,54**	5,77	1,92	0	9,62	9,62	3,85	9,62	3,85	11,54	*N*
	0	**47,06**	5,88	7,84	0	5,88	3,92	1,96	9,80	5,88	11,76	*EC*
	1,59	**42,86**	0	7,94	0	7,94	11,11	3,17	3,17	4,76	17,46	*PD*
Dis	16,22	5,41	**72,97**	0	0	0	0	0	0	0	5,41	*N*
	25,00	12,50	**57,50**	0	0	0	0	0	0	0	5,00	*EC*
	8,00	4,00	**86,00**	2,00	0	0	0	0	0	0	0	*PD*
Emb	2,50	12,50	2,50	**60,00**	2,50	0	2,50	0	7,50	0	10,00	*N*
	1,79	16,07	0	**69,64**	0	1,79	1,79	0	8,93	0	0	*EC*
	2,08	12,50	0	**56,25**	0	2,08	2,08	0	10,42	0	14,58	*PD*
Fea	0	0	21,28	4,26	**61,70**	0	0	0	0	12,77	0	*N*
	2,04	0	8,16	0	**59,18**	0	0	0	0	28,57	2,04	*EC*
	0	0	12,20	4,88	**58,54**	0	0	0	0	24,39	0	*PD*
Joy	0	0	0	0	0	**94,83**	0	5,17	0	0	0	*N*
	0	0	0	0	0	**86,54**	1,92	7,69	0	1,92	1,92	*EC*
	0	0	0	2,27	0	**95,45**	0	2,27	0	0	0	*PD*
Pri	0	10,00	0	0	0	40,00	4,00	**40,00**	0	4,00	2,00	*N*
	0	2,08	0	0	0	43,75	0	**45,83**	2,08	0	6,25	*EC*
	0	3,77	0	0	0	54,72	0	**35,85**	0	3,77	1,89	*PD*
Sad	2,38	9,52	14,29	0	2,38	0	0	0	**66,67**	2,38	2,38	*N*
	0	7,50	15,00	0	10,00	0	5,00	0	**60,00**	0	2,50	*EC*
	2,38	7,14	11,90	0	9,52	0	2,38	0	**61,90**	0	4,76	*PD*
Sur	0	0	0	0	5,56	0	0	0	0	**94,44**	0	*N*
	0	0	2,13	0	6,38	2,13	0	0	0	**89,36**	0	*EC*
	0	0	0	0	1,96	0	0	0	1,96	**96,08**	0	*PD*

Note. Rows present the emotional expressions shown in the videos. Columns show the respective ratings averaged for each induction condition (Ind) from top to bottom: the neutral condition (N), empathic concern (EC) for the mental pain condition, and personal distress (PD) for the physical pain condition. Bold values present the raw hit rates, and red values indicate confusions above 11%. IDK = "I do not know".

Hit and confusion rates are represented in the rows, separately for all inductions (from top to bottom: neutral, empathic concern, personal distress). The neutral expressions were excluded in those matrices because they do not have varying intensities. However, the option to rate *neutral* was always possible, therefore the column for *neutral* was included and presents confusions with the respective emotion label shown in the rows. Neutral facial expressions were correctly labeled 69.04 %. Confusion rates with other emotional expressions were low: anger (ang): 5.26 %, contempt (con): 3.41 %, disgust (dis): 1.24 %, embarrassment (emb): 0.62 %, fear (fea): 1.55 %, joy: 2.79 %, pride (pri): 1.24 %, sadness (sad): 6.50 %, surprise (sur): 1.55 %, IDK: 6.81 %. A visualization of the hit rates for each emotion label, separately for induction condition, can be found in Appendix B (*B1* for high-, and *B2* for low-intensity expressions).

Table 4.

Confusion matrix for low-intensity emotional expression ratings (values in %).

Presented emotion	*Perceived emotion*											
	Ang	**Con**	**Dis**	**Emb**	**Fea**	**Joy**	**(Neu)**	**Pri**	**Sad**	**Sur**	**IDK**	**Ind**
	52.54	10.17	6.78	0	5.08	1.69	10.17	0	5.08	1.69	6.78	*N*
Ang	**61.90**	11.90	7.14	2.38	2.38	0	4.76	0	2.38	2.38	4.76	*EC*
	64.10	0	0	0	5.13	2.56	10.26	0	5.13	5.13	7.69	*PD*
	1.79	**25.00**	1.79	5.36	0	5.36	30.36	5.36	3.57	3.57	17.86	*N*
Con	0	**33.33**	2.22	4.44	0	0	17.78	4.44	8.89	11.11	17.78	*EC*
	0	**30.36**	0	8.93	0	7.14	30.36	1.79	7.14	0	14.29	*PD*
	20.93	9.30	**67.44**	0	0	0	0	0	2.33	0	0	*N*
Dis	18.37	6.12	**69.39**	2.04	2.04	0	0	0	0	0	2.04	*EC*
	15.79	12.28	**64.91**	0	0	0	1.75	0	1.75	1.75	1.75	*PD*
	2.08	8.33	2.08	**33.33**	6.25	2.08	12.50	0	20.83	2.08	10.42	*N*
Emb	3.45	15.52	10.34	**22.41**	1.72	0	12.07	0	8.62	3.45	22.41	*EC*
	6.00	8.00	0	**30.00**	0	6.00	18.00	0	18.00	0	14.00	*PD*
	2.08	8.33	4.17	4.17	**33.33**	0	4.17	0	0	41.67	2.08	*N*
Fea	1.79	1.79	7.14	5.36	**44.64**	0	0	0	1.79	30.36	7.14	*EC*
	3.23	0	4.84	3.23	**41.94**	0	0	0	0	37.10	9.68	*PD*
	0	1.79	1.79	8.93	0	**64.29**	10.71	5.36	1.79	0	5.36	*N*
Joy	0	1.85	1.85	1.85	1.85	**51.85**	7.41	7.41	3.70	0	11.11	*EC*
	0	2.50	0	7.50	0	**62.50**	12.50	10.00	0	0	5.00	*PD*
	0	2.00	0	0	0	60.00	4.00	**16.00**	0	0	18.00	*N*
Pri	0	2.94	0	0	0	41.18	5.88	**32.35**	2.94	2.94	11.76	*EC*
	0	2.33	0	0	0	67.44	4.65	**20.93**	0	0	4.65	*PD*
	0	5.26	10.53	2.63	7.89	0	2.63	0	**63.16**	0	7.89	*N*
Sad	2.78	5.56	8.33	2.78	2.78	0	0	0	**66.67**	2.78	8.33	*EC*
	8.00	6.00	12.00	2.00	2.00	0	4.00	0	**52.00**	2.00	12.00	*PD*
	0	0	0	2.38	9.52	0	0	0	0	**88.10**	0	*N*
Sur	0	0	0	0	7.14	0	0	0	0	**92.86**	0	*EC*
	0	0	0	1.92	7.69	0	0	0	0	**86.54**	3.85	*PD*

Note. Rows present the emotional expressions shown in the videos. Columns show the respective ratings averaged for each induction condition (Ind) from top to bottom: the neutral condition (N), empathic concern (EC) for the mental pain condition, and personal distress (PD) for the physical pain condition. Bold values present the raw hit rates, and red values indicate confusions above 11%. IDK = "I do not know".

Comparing high- and low-intensity expressions, one notices an accuracy drop in the categorical emotion ratings and an increase of IDK selections for low-intensity. *Contempt*, *embarrassment*, and *pride*, commonly classified as complex emotions, show low recognition rates and comparably high IDK proportions. Overall, *pride* was the least recognized emotional expression and *surprise* the most recognized one, partly attributable to the high confusion rates. Facial expressions of *fear* were often labeled as *surprise*, and the label *joy* was often assigned when the presented facial expression was *pride*. In low-intensity expressions, faces conveying *contempt* and *embarrassment* were often labeled as *neutral*.

Table 5 accumulates the information presented in the previously shown confusion matrices to visualize the differences in empathic state conditions. The hit rates of the neutral control condition were subtracted from both empathic state conditions (empathic concern, personal distress) separately for each emotional expression. Positive values indicate better than neutral emotion perception, whereas negative values indicate worse than neutral perception.

Table 5.

Comparisons of hit rates between the empathic concern and personal distress induction condition. Values are compared to the neutral condition and presented in %.

Intensity	*Induction*	Ang	Con	Dis	Emb	Fea	Joy	Pri	Sad	Sur
high	EC	- 7.64	10.52	- 15.47	9.64	- 2.52	- 8.29	5.83	- 6.67	- 5.08
	PD	- 6.40	6.32	13.03	- 3.75	- 3.16	0.62	- 4.15	- 4.77	1.64
low	EC	9.36	8.33	1.95	- 10.92	11.31	- 13.07	16.35	3.51	4.76
	PD	11.56	5.36	- 2.53	- 3.33	8.61	- 1.79	4.93	- 11.16	- 1.56

Note. Empathic concern induction was calculated as (hit EC) – (hit neutral). Personal distress was calculated as (hit PD) – (hit neutral).

Focusing on the low-intensity expressions, one sees that *disgust*, *sadness*, and *surprise* were best recognized in the mental pain condition and worst in the physical pain condition. *Contempt*, *fear*, and *pride* show a similar pattern, with the best recognition in the mental pain condition but the worst in the neutral condition. *Embarrassment* and *joy* were best recognized in the neutral condition and the worst in the mental pain condition. The mental pain induction showed the highest hit rates for all complex emotion expressions (*contempt*, *embarrassment*, *pride*) for high-intensity but scored lowest (except for *fear*) in all the other emotional expression ratings compared to the other induction conditions.

Table 6.

Mean confidence, valence, and arousal ratings for each emotional expression.

	correct	*intensity*	Ang	Con	Dis	Emb	Fea	Joy	Pri	Sad	Sur	Neu
confidence	yes	high	4.40	3.29	4.56	3.91	4.18	4.75	4.08	4.39	4.70	4.10
		low	3.65	2.96	3.91	3.36	3.67	3.79	3.68	4.00	4.37	
	no	high	3.39	3.44	3.93	3.26	4.00	3.55	4.12	3.31	3.70	3.01
		low	3.08	2.98	3.70	3.22	3.70	3.04	3.95	3.50	3.93	
valence	yes	high	1.52	2.13	1.47	2.40	1.67	4.75	4.05	1.90	3.31	3.02
		low	1.90	2.28	1.89	2.32	1.91	3.97	3.93	2.07	3.07	
	no	high	2.39	2.80	1.47	2.37	2.15	4.46	4.28	2.00	2.50	2.45
		low	2.51	2.84	1.72	2.45	2.55	3.00	4.08	2.31	1.93	
arousal	yes	high	3.77	2.89	4.30	3.16	4.11	4.36	3.89	3.58	4.42	2.31
		low	2.73	2.13	2.94	2.34	3.42	2.62	3.21	2.50	3.51	
	no	high	3.65	2.70	4.17	2.89	4.00	3.73	3.91	3.02	4.30	2.39
		low	2.92	2.09	3.49	2.45	3.28	2.08	3.32	2.61	3.79	

Note. *Correct* represents if the same emotional expression was perceived as was presented. *Neutral* only contains one value for correctly and incorrectly rated emotional expressions because those expressions do not vary in intensity.

Additionally to the rating of specific emotion labels, participants also indicated their confidence in this rating and rated the perceived valence and arousal of the facial expression. *Table 6* presents the average ratings for each emotional expression, separated for correctness (overlap between facial expression rating and agreed upon emotion label) and intensity.

State and Trait Empathy and Emotion Perception

The intensity of empathic states was assessed with the ERS and empathic traits with the B-IRI. Pearson correlations were calculated to compare state and trait measures. The trait subscale personal distress was not correlated to the ERS rating in the physical pain condition, which induced personal distress (p = .863), nor was the trait subscale empathic concern correlated to the ERS rating in the emotional pain condition which induced empathic concern (p = .737). However, in the personal distress condition trait personal distress and the oH for high-intensity expressions were negatively correlated r = - .263, p = .017. Assessing the relationship between trait empathy and emotion perception without accounting for the induction condition revealed a negative correlation between trait personal distress and oH for high-intensity expressions (r = - .25, p = .026) but not for low-intensity expressions (r = .19, p = .092). Trait empathic concern was not related to the overall hit rates.

Emotion Regulation Strategies

As previously reported, suppression was negatively correlated with trait empathic concern (see *Table 1*), but neither regulation style was correlated with trait personal distress. However, assessing regulation styles within different induction conditions showed marginally significant negative correlations between state personal distress and reappraisal (in empathic concern: r = - .187, p = .093; in personal distress: r = - .194, p = .081). No significant correlations were found in the neutral condition. Examining individuals high and low in personal distress states (split at the mean, like in the main analysis) in the personal distress condition also indicated marginally significant lower reappraisal scores for those higher in state personal distress $t(80)$ = -1.71, p = .091, d = -.38. Emotion regulation strategies were not correlated to the oH in either high-, or low-intensity expressions.

Discussion

The study aimed to analyze the influence of empathic states on emotion perception in dynamic facial expressions of high- and low-intensity. Overall, the results suggest a strong relationship between expressive intensity and emotion perception accuracy, with an advantage of high- relative to low-intensity emotional expressions. No significant relationships were found between empathic states and general emotion perception accuracy. When assessing the individual emotional expressions separately, the same pattern emerged. Confidence ratings dropped when expressive intensity was low and interestingly varied between induction conditions (post-hoc). Arousal-ratings varied with the expressive intensity only and valence-ratings with neither.

Gender Differences

When comparing questionnaire scales for trait empathy and emotion regulation strategies between genders, the only difference was that male participants rated their trait personal distress significantly higher than female participants. This is contrary to the reported sex differences in the original IRI (Davis, 1983) and the brief version used in this study B-IRI (Ingoglia et al., 2016). In both of those studies, female participants scored higher than males in all four subscales. Males' increased personal distress ratings in this study are exceptional and indicate an unusual sample concerning trait empathy.

Empathic State Induction

The study successfully induced empathic concern and personal distress but could not find a difference in the groups' perception of facial expressions. When talking about successful induction, one has to remember, self-report measures determined that success. Participants could have felt obligated to rate their reaction to vignettes in a socially expected way even though they might have felt differently.

Moreover, the vignettes presented in the current study included losing a loved one and experiencing a heart attack. Some participants might have experienced those scenarios in the immediate or distant past and might have reacted more sensitively. Like others, Israelashvili et al. (2020a) showed that personal distress is linked to lower emotion perception accuracy. Interestingly, people who already experienced a similar event, like the person they empathized with, showed increased personal distress (Israelashvili et al., 2020b). Subsequently, familiarity with the experience led to lower emotion perception accuracy in these individuals than those who had not been in the same situation before. The confounding variable of familiarity with the scenarios was not controlled for in the current study and might have contributed to the non-significant findings.

Lastly, the vignettes were paired with pictures of people of different ages, genders, and ethnicities, validated before (Fabi et al., 2019). This variety was beneficial for a diverse sample because people often show higher empathic reactions to ingroup than outgroup members (e.g., Neumann et al., 2013). Following the induction, participants rated dynamical facial expressions, which also varied in age and gender. However, those actors and actresses were not the same as the ones from the picture stimuli of the induction. It is possible that this mismatch of empathic induction situation and emotion perception situation decreased the influence of state empathy on emotion perception. Stimuli that fulfill a twofold function, namely empathic state induction and emotional expression presentation with the same models, could eliminate those methodological problems.

General Emotion Perception

The results of this study indicate a strong influence of expressive intensity on emotion perception. Coherent with previous studies, high-intensity compared to low-intensity expressions resulted in better emotion perception accuracy (e.g., Wingenbach et al., 2016). However, the study cannot support the hypothesis that higher state empathic concern increases and higher state personal distress decreases emotion perception, in the sense of labeling dynamic facial expressions as predefined. However, post-hoc follow-up analyses showed a trend in the hypothesized direction. Participants experiencing strong empathic concern showed increased emotion perception accuracy, and those experiencing strong personal distress showed decreased emotion perception accuracy in the respective induction condition. This trend was only found for low-intensity expressions but not for high-intensity expressions. Overall, those results suggest that just like variations in trait empathy, variations in state empathy can have different effects on interpersonal skills like emotion perception. Therefore, it might be necessary to assess induction success and the extent of the empathic reactions before conducting subsequent analyses in the future.

Confidence and Arousal Ratings

People were more confident in their high-intensity expression ratings compared to their low-intensity expression ratings. This was expected and conformed with a previous study where higher intensity in texts was correlated to higher confidence ratings (Troiano et al., 2021). Similarly, arousal was rated higher in high-intensity expressions relative to low-intensity expressions. Those findings support the intention of the ADFES-BIV stimuli, distinguishing between easier to recognize, high-intensity videos and harder to recognize, low-intensity videos. Surprisingly, the induction condition also influenced confidence ratings. Participants indicated significantly higher confidence in state personal distress relative to the other induction conditions. Other authors found increased confidence levels under stress before (e.g., in situational awareness, Price et al., 2016). Furthermore, variations in confidence were also found in a previous study that controlled for trait anxiety. Those higher in trait anxiety reacted with underconfidence, and those low in trait anxiety reacted with overconfidence in a competitive situation under stress but no differences were found without stress (Goette et al., 2015). The current results indicate a similar variation of confidence ratings depending on empathic states. State personal distress increased confidence levels in emotional expression ratings relative to state empathic concern or a neutral baseline condition; novel findings that could be investigated in future studies.

Discrete Emotion Perception and Confusions

Pride was one of the most commonly confused emotional expressions in high- (46.16 % rated as *joy*) and low- (56.21 % rated as *joy*) intensity facial expressions. In emotion research, it is often criticized that the only positively valenced emotional expression and label option is *happiness*. For this reason, *happiness* is easy to identify and has low confusion rates compared to negatively valenced emotional expressions. By including *pride* in this study's stimuli set another positively valenced emotional expression was presented. The assumed positivity of *pride* was also supported in the valence ratings, where participants rated emotional expressions of *pride* and *joy* with almost the same valence (see *Table 6*). However, *pride* was excessively confused with *joy* but not the other way round. Ekman asked researchers in the field which emotions they thought were sufficiently grounded in biology and hence expressed universally, based on the current state of research. While *happiness* was rated by 76 %, *pride* only received support from 9 % (Ekman, 2016). Those opinions paired with the results of this study suggest that expressions supposedly showing *pride* might not be as clearly distinguishable as the assumptions of basic emotion theory claim. Contrary to other emotional states like *anger* or *happiness*, *pride* is a social emotion that probably has more alterable connections between feeling states and expressive facial movements. This flexibility could have distorted emotion perception accuracy towards labels of expressions expressed more coherently. Moreover, in the facial action coding system (FACS, Ekman & Friesen, 1978), *pride* overlaps with both action units of *joy* 6 (cheek raiser) and 12 (lip-corner puller), explaining confusions, especially in low-intensity and giving further doubt of the clear distinction between the two emotion-concepts. Additional *pride* units are 53 (head up) and 64 (eyes down), among descriptions of a straight back and an exposed chest. With only facial expressions, essential cues are missing to successfully distinguish *pride* from *joy*, which might have been the problem in the current study.

Similarly, *fear* was often rated as *surprise* in high (21.91 %) and low (36.38 %) intensity facial expressions, while the opposite direction was seldom confused (below 10 % in both expressive intensities). The expression of *surprise* is defined with action units 1, 2, 5, 26, while *fear* has all of those and additional AUs 4, 7, 20. However, in Ekman's poll (2016), *fear* was rated 90%, while *surprise* only reached 40 to 50 %. A possible explanation for the better recognition of *surprise* could be the focus on mouth or eye regions. In a study by Calder et al. (2000), *fear* was better recognized from the top half of the face than the bottom half, while *surprise* was recognized equally well from either half. Consequently, the interindividual differences in facial focus could have led unnoticed to a higher rating frequency of *surprise*. Moreover, contextual cues are essential to distinguish between ambiguous faces and might be particularly important for distinguishing between expressions of *fear* and *surprise*. With those cues missing in stimuli that only present facial expressions, it is possible that due to fewer AUs needed to infer *surprise*, participants rated fearful faces more readily as surprised than the other way round.

In the low-intensity condition, facial expressions conveying *embarrassment* (12 – 18 %) and *contempt* (18 – 30 %) were often rated as *neutral*. For *contempt*, the findings overlap with a previous study where emotions were rated categorically and dimensionally on a circumplex model, with one axis indicating pleasantness (pleasant-unpleasant) and the other axis indicating activeness (active-passive). Among all presented expressions (*embarrassment* not included), *contempt* and *neutral* were not rated significantly differently in the categorical rating nor on the circumplex model (Kilpeläinen & Salmela, 2020). Usually, *embarrassment* is accompanied by a physiological body reaction of blushing. However, the facial expressions in the videos of this study were acted out, hence no blushing was present. A study where neutral faces were manipulated in their blushing showed that stronger facial blushing was rated more likely as *embarrassment* by observers (Thorstenson et al., 2020). Therefore, it is likely, that this cue is vital to distinguish *embarrassment* from *neutral* expressions, especially when expressive intensity is low.

These findings, coupled with the ones in this study, demand a reconsideration of the strict differentiation between some facial expressions. Faces can convey more than just the initial six basic emotions, but simply extending stimuli by adding complex emotional expressions does not seem to enhance classical labeling tasks. Basic emotional expressions are easier to identify, and labels of basic emotions are selected more readily for ambiguous expressions relative to complex emotion labels. It remains a challenge for emotion research to develop stimuli that can sufficiently cover the wide variety of facial expressions and assess individuals' emotion perception more flexibly.

Variations in Trait Empathy and Emotion Regulation

Previous findings showed the positive relationship of trait empathic concern and the negative relationship of trait personal distress with emotion perception (Israelashvili et al., 2020a; Mayukha et al., 2020). The results of this study partly support those finding, mainly for the negative relationship of trait personal distress with emotion perception accuracy of high-intensity expressions. No correlation was found between empathic states and empathic traits, as often indirectly implied in previous studies. The current study indicates that people who score high on self-ratings of trait empathy do not necessarily show empathic states coherent with those scores in a given situation. Nevertheless, state empathic concern and state personal distress pointed into the same directions as their trait measures – with empathic concern being beneficial and personal distress being hampering for emotion perception.

Limitations

The study was conducted during a global pandemic which did not allow for research in person. Therefore, the whole experiment was conducted online, limiting the controllability of participants' attention and engagement with the experiment. Even though empathic induction's success was given, a controlled, quiet environment might have increased the effect of the induction and subsequent emotion perception. Moreover, the experiment was programmed to balance for actor's sex and intensity of expression, but not for specific emotional expressions because the overall research question was not specific to any single emotional expression. It was expected that randomization would balance emotional expressions automatically. This was true for the videos presented overall in the study but not for those per participant; some expressions appeared more often, and others never for most individuals.

Due to ethical reasons, the pictures selected for the inductions that presented people in pain were relatively weak, possibly limiting the strength of the empathic reactions. However, vignettes combined with each picture block were supposed to make a vivid imagination possible and provoke strong reactions regardless. The number of pictures presented was set to eight per block to balance effectiveness and the presentation's time constraints. This was determined based on discussions with students familiar with the field rather than previous research, which has not applied this exact design yet. Likewise, the presentation of 12 emotional expressions after each induction was determined. A pilot study to assess empathic states after presenting different numbers of pictures and videos could have helped determine effectiveness but was not conducted due to time constraints. Unfortunately, it is possible that the induction was successful only for earlier trials and flattened out towards later expressions, thereby distorting the overall results.

Another limitation of the experiment might have constituted the facial expression stimuli. They were selected from an exhaustive database for emotional stimuli and taken as the most suitable instrument for this study. Compared to static stimuli sets which limit ecologic validity, the ADFES-BIV presents facial expressions of basic and complex emotions dynamically and in varying intensities. However, posed expressions differ from naturally arising everyday expressions, and the ones in the videos were acted out. Another weakness already addressed by the authors of the ADFES-BIV is the unnatural development time for different expressions and intensities, kept constant for standardization reasons. Other authors found that basic emotions vary in progression times between 330 ms and 1400 ms, with *anger* and *surprise* developing quickly and *sadness* developing slowly (Yoshikawa & Sato, 2008). Missing variations between facial expressions and artificially stretching out the duration of low-intensity expressions can impact emotion perception. Furthermore, facial expressions not only have varying onsets but also offsets. The ADFES-BIV facial expressions progress from neutral to a specific expression and are cut after the expression is fully present. However, the offset of expressions can significantly contribute to identifying emotions and were missing in the utilized stimuli. Lastly, contextual information, which is crucial for interpreting ambiguous expressions (Clark et al., 2020), was missing. So were sounds and body postures that also contribute to emotion perception in more natural settings. The problems raised here are frequently debated, and suitable measures that account for those variables are yet to be developed. However, the design of this study should not be disregarded because it already improved earlier measures (black and white static pictures in high-intensity) substantially. The inclusion of more complex expressions (like *embarrassment* or *contempt*) broadened the scope of the study. However, it cannot be exclude that those expressions were harder to perceive and more challenging to understand conceptually. Understanding might have been exceedingly complicated for an intercultural sample where English is not the mother tongue of many participants. Short definitions to familiarize the participants with unfamiliar words were provided. However, it is possible that selecting one of those less familiar options was less likely due to precisely those explanations given above. Similarly, the option to honestly respond "I do not know" can increase the answers' validity and produce a clean dataset. However, it could have also distorted results away from more complex emotional expression labels.

Future Research

Future studies could use different empathy-inducing materials, possibly using more potent stimuli, more realistic scenarios, or pictures of related people. Similarly, they could use other emotion perception materials, where more realistic expressions are presented in context. Virtual reality and augmented reality could be interesting technologies utilizable for those purposes.

Eye trackers could record eye movements to analyze the focus of attention, which could vary during induction and emotion perception. While it could work as an additional induction and attention control, it could also show interesting patterns for emotion perception, which might vary with the empathic condition between a more pronounced focus on the eye or the mouth regions. Depending on the facial region, some emotional categories are more easily recognized, with *anger* and *fear* from the eye region and *happiness* and *disgust* from the mouth region (Calder et al., 2000 incl. *sadness* from eyes; Smith et al., 2005 incl. *surprise* from eyes). Older people who tend to focus more on the lower part of the face (Sullivan et al., 2007) showed the best recognition for *disgust* and worst for *anger* (Wong et al., 2005). Furthermore, older people were found to confuse *anger* with *disgust* more often, whereas younger people confused *disgust* with *anger* (Ebner et al., 2011), likely attributable to their difference in attentional focus. Similarly, visual gaze patterns could vary with empathic states and might contribute to differences in subsequent emotion perception accuracy and confusions. Although mainly used in laboratories, advances in the field could allow for applications of eye-tracking in online experiments as well (Semmelmann & Weigelt, 2018). Additional gaze information, which remained undetected in this study, could underly differences in empathic emotion perception.

A very different approach could change the causal direction tested in this study because competing theories about the relationship between empathy and emotion perception exist. Instead of inducing empathic states and subsequently measuring the emotion perception accuracy, one could measure emotion perception accuracy first, and subsequently, the empathic reaction towards the facial expression presented. Some expressions or distinct features may provoke stronger empathic concern while others provoke stronger personal distress, a possibility that was not examined in this study.

Besides controlling for emotion regulation strategies, one could also measure self-awareness, mental flexibility, and attachment styles because those traits can alter the experience of empathy (Decety & Meyer, 2008) and consequently emotion perception. Measuring additional covariates could be helpful in better understanding the non-significant results of the current study.

Conclusion

This study examined the influence of empathic states on emotion perception. Despite well-wrought inductions of empathic concern and personal distress with vignettes and pictures, no influence on emotion categorization accuracy with the ADFES-BIV was detected. Even though multiple improvements to previous designs were implemented, future studies are needed to reexamine the non-significant results with other approaches and elucidate underlying processes that were not covered in this study. Based on the discussed limitations, suggestions for future research were proposed. For now, the influence of empathic states on emotion perception cannot be supported. At least not in the classical *basic emotion theory* sense, where specific facial expressions can be paired with discrete emotion labels.

References

Abbruzzese, L., Magnani, N., Robertson, I. H., & Mancuso, M. (2019). Age and gender differences in emotion recognition. *Frontiers in Psychology*, 1–15. https://doi.org/10.3389/fpsyg.2019.02371

APA dictionary of psychology. (2006). American Psychological Association.

Balconi, M., & Canavesio, Y. (2016). Is empathy necessary to comprehend the emotional faces? The empathic effect on attentional mechanisms (eye movements), cortical correlates (N200 event-related potentials) and facial behaviour (electromyography) in face processing. *Cognition and Emotion*, *30*(2), 210–224. https://doi.org/10.1080/02699931.2014.993306

Calder, A. J., Keane, J., Young, A. W., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 527–551. https://doi.org/10.1037/0096-1523.26.2.527

Caldwell, A. R. ., & Lakens, D. (2019). *Power Analysis with Superpower*. https://aaroncaldwell.us/SuperpowerBook/

Chakrabarti, B., & Baron-Cohen, S. (2010). In the Eyes of the Beholder: How Empathy Influences Emotion Perception. In R. B. Adams, N. Ambady, K. Nakayama, & S. Shimojo (Eds.), *The Science of Social Vision*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333176.003.0013

Chrysikou, E. G., & Thompson, W. J. (2016). Assessing Cognitive and Affective Empathy Through the Interpersonal Reactivity Index: An Argument Against a Two-Factor Model. *Assessment*, *23*(6), 769–777. https://doi.org/10.1177/1073191115599055

Clark, G. M., McNeel, C., Bigelow, F. J., & Enticott, P. G. (2020). The effect of empathy and context on face-processing ERPs. *Neuropsychologia*, *147*. https://doi.org/10.1016/j.neuropsychologia.2020.107612

Davis, M. H. (1980). A Mulitdimensional Approach to Individual Differences in Empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85–104.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498–509. https://doi.org/10.1038/NRN.2017.72

Decety, J., & Meyer, M. (2008). From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and Psychopathology*, *20*(4), 1053–1080. https://doi.org/10.1017/S0954579408000503

Diconne, K., Paltoglou, A., Kountouriotis, G., Parker, A., & Hostler, T. (2021). *KAPODI- The Searchable Database of 364 Available Emotional Stimuli Sets*. PsyArXiv. https://doi.org/10.31234/OSF.IO/6J9MX

Ding, R., Ren, J., Li, S., Zhu, X., Zhang, K., & Luo, W. (2020). Domain-general and domain-preferential neural correlates underlying empathy towards physical pain, emotional situation and emotional faces: An ALE meta-analysis. *Neuropsychologia*, *137*. https://doi.org/10.1016/j.neuropsychologia.2019.107286

Dores, A. R., Barbosa, F., Queirós, C., Carvalho, I. P., & Griffiths, M. D. (2020). Recognizing Emotions through Facial Expressions: A Largescale Experimental Study. *International Journal of Environmental Research and Public Health*, *17*(7420). https://doi.org/10.3390/ijerph17207420

Ebner, N. C., He, Y., & Johnson, M. K. (2011). Age and emotion affect how we look at a face: Visual scan patterns differ for own-age versus other-age emotional faces. *Http://Dx.Doi.Org/10.1080/02699931.2010.540817*, *25*(6), 983–997. https://doi.org/10.1080/02699931.2010.540817

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on Psychological Science*, *11*(1), 31–34. https://doi.org/10.1177/1745691615596992:

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. Pergamon Press.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press.

Fabi, S., Weber, L. A., & Leuthold, H. (2019). Empathic concern and personal distress depend on situational but not dispositional factors. *PLOS ONE*, *14*(11), 1–22. https://doi.org/10.1371/journal.pone.0225102

Gallese, V. (2007). Before and below "theory of mind": Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 659–669. https://doi.org/10.1098/RSTB.2006.2002

Gery, I., Miljkovitch, R., Berthoz, S., & Soussignan, R. (2009). Empathy and recognition of facial expressions of emotion in sex offenders, non-sex offenders and normal controls. *Psychiatry Research*, *165*(3), 252–262. https://doi.org/10.1016/j.psychres.2007.11.006

Goette, L., Bendahan, S., Thoresen, J., Hollis, F., & Sandi, C. (2015). Stress pulls us apart: Anxiety leads to differences in competitive confidence under stress. *Psychoneuroendocrinology*, *54*, 115–123. https://doi.org/10.1016/J.PSYNEUEN.2015.01.019

Gross, J. J., & John, O. P. (2003). Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Håkansson Eklund, J., & Summer Meranius, M. (2021). Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling*, *104*(2), 300–307. https://doi.org/10.1016/j.pec.2020.08.022

Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *Journal of Social Psychology*, *159*(3), 225–243. https://doi.org/10.1080/00224545.2018.1477442

Hall, J. K., Hutton, S. B., & Morgan, M. J. (2010). Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition? *Cognition and Emotion*, *24*(4), 629–637. https://doi.org/10.1080/02699930902906882

Harris, D. A., Hayes-Skelton, S. A., & Ciaramitaro, V. M. (2016). What's in a Face? How Face Gender and Current Affect Influence Perceived Emotion. *Frontiers in Psychology*, *7*(1468). https://doi.org/10.3389/FPSYG.2016.01468

Hengartner, M. P., Ajdacic-Gross, V., Rodgers, S., Müller, M., Haker, H., & Rössler, W. (2014). Fluid intelligence and empathy in association with personality disorder trait-scores: Exploring the link. *European Archives of Psychiatry and Clinical Neuroscience*, *264*(5), 441–448. https://doi.org/10.1007/S00406-013-0441-0

Hess, U., Blairy, S., & Kleck, R. E. (1997). The Intensity of Emotional Facial Expressions and Decoding Accuracy. *Journal of Nonverbal Behavior*, *21*(4), 241–257. https://doi.org/10.1023/A:1024952730333

Hess, U., & Fischer, A. (2013). Emotional Mimicry as Social Regulation. *Personality and Social Psychology Review*, *17*(2), 142–157. https://doi.org/10.1177/1088868312472607

Ingoglia, S., Lo Coco, A., & Albiero, P. (2016). Development of a Brief Form of the Interpersonal Reactivity Index (B–IRI). *Journal of Personality Assessment*, *98*(5), 461–471. https://doi.org/10.1080/00223891.2016.1149858

Israelashvili, J., & Karniol, R. (2018). Testing alternative models of dispositional empathy: The Affect-to-Cognition (ACM) versus the Cognition-to-Affect (CAM) model. *Personality and Individual Differences*, *121*, 161–169. https://doi.org/10.1016/j.paid.2017.09.036

Israelashvili, J., Sauter, D. A., & Fischer, A. H. (2020a). Two facets of affective empathy: concern and distress have opposite relationships to emotion recognition. *Cognition and Emotion*, *34*(6), 1112–1122. https://doi.org/10.1080/02699931.2020.1724893

Israelashvili, J., Sauter, D. A., & Fischer, A. H. (2020b). Different faces of empathy: Feelings of similarity disrupt recognition of negative emotions. *Journal of Experimental Social Psychology*, *87*. https://doi.org/10.1016/j.jesp.2019.103912

Jordan, M. R., Amir, D., & Bloom, P. (2016). Are empathy and concern psychologically distinct? *Emotion*, *16*(8), 1107–1116. https://doi.org/10.1037/emo0000228

Kilpeläinen, M., & Salmela, V. (2020). Perceived emotional expressions of composite faces. *PLoS ONE*, *15*(3). https://doi.org/10.1371/JOURNAL.PONE.0230039

Lakens, D., & Caldwell, A. R. (2019). *Simulation-Based Power-Analysis for Factorial ANOVA Designs*. https://doi.org/10.31234/osf.io/baxsf

Lockwood, P. L., Bird, G., Bridge, M., & Viding, E. (2013). Dissecting empathy: High levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. *Frontiers in Human Neuroscience*, *7*(760). https://doi.org/10.3389/FNHUM.2013.00760

Lockwood, P. L., Seara-Cardoso, A., & Viding, E. (2014). Emotion regulation moderates the association between empathy and prosocial behavior. *PLoS ONE*, *9*(5). https://doi.org/10.1371/JOURNAL.PONE.0096555

Ma-Kellams, C., & Lerner, J. (2016). Trust your gut or think carefully? Examining whether an intuitive, versus a systematic, mode of thought produces greater empathic accuracy. *Journal of Personality and Social Psychology*, *111*(5), 674–685. https://doi.org/10.1037/PSPI0000063

Mayukha, A., Andrade, I., & Cone, J. (2020). Opposing Contributions of Psychologically Distinct Components of Empathy to Empathic Accuracy. *Journal of Experimental Psychology: General*, *149*(11), 2169–2186. https://doi.org/10.1037/xge0000768

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, *40*(4), 525–543. https://doi.org/10.1111/J.1467-6494.1972.TB00078.X

Neumann, D. L., Boyle, G. J., & Chan, R. C. K. (2013). Empathy towards individuals of the same and different ethnicity when depicted in negative and positive contexts. *Personality* *and Individual Differences*, *55*(1), 8–13. https://doi.org/10.1016/J.PAID.2013.01.022

Nezlek, J. B., Feist, G. J., Wilson, F. C., & Plesko, R. M. (2001). Day-to-day variability in empathy as a function of daily events and mood. *Journal of Research in Personality*, *35*(4), 401–423. https://doi.org/10.1006/jrpe.2001.2332

Olderbak, S., Wilhelm, O., Hildebrandt, A., & Quoidbach, J. (2019). Sex differences in facial emotion perception ability across the lifespan. *Cognition and Emotion*, *33*(3), 579–588. https://doi.org/10.1080/02699931.2018.1454403

Phillips, L. H., Channon, S., Tunstall, M., Hedenstrom, A., & Lyons, K. (2008). The Role of Working Memory in Decoding Emotions. *Emotion*, *8*(2), 184–191. https://doi.org/10.1037/1528-3542.8.2.184

Powell, P. A. (2018). Individual differences in emotion regulation moderate the associations between empathy and affective distress. *Motivation and Emotion*, *42*(4), 602–613. https://doi.org/10.1007/S11031-018-9684-4

Price, T., Tenan, M., Head, J., Maslin, W., & Lafiandra, M. (2016). Acute stress causes over confidence in Situation Awareness. *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2016*, 1–6. https://doi.org/10.1109/COGSIMA.2016.7497778

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141. https://doi.org/10.1037/0033-2909.115.1.102

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Schmid, P. C., Schmid Mast, M., Bombari, D., Mast, F. W., & Lobmaier, J. S. (2011). How Mood States Affect Information Processing During Facial Emotion Recognition: An Eye Tracking Study. *Swiss Journal of Psychology*, *70*(4), 223–231. https://doi.org/10.1024/1421-0185/a000060

Schmidt, S. N. L., Hass, J., Kirsch, P., & Mier, D. (2021). The human mirror neuron system—A common neural basis for social cognition? *Psychophysiology*. https://doi.org/10.1111/psyp.13781

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465. https://doi.org/10.3758/S13428-017-0913-7

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*(3), 184–189.

https://doi.org/10.1111/J.0956-7976.2005.00801.X

Stellar, J. E., Anderson, C. L., & Gatchpazian, A. (2019). Profiles in Empathy: Different Empathic Responses to Emotional and Physical Suffering. *Journal of Experimental Psychology: General*, *149*(7), 1398 –1416. https://doi.org/10.1037/xge0000718

Sullivan, S., Campbell, A., Hutton, S. B., & Ruffman, T. (2017). What's good for the goose is not good for the gander: Age and gender differences in scanning emotion faces. *The Journals of Gerontology: Series B*, *72*(3), 441–447. https://doi.org/10.1093/geronb/gbv033

Sullivan, S., Ruffman, T., & Hutton, S. B. (2007). Age Differences in Emotion Recognition Skills and the Visual Scanning of Emotion Faces. *Journal of Gerontology*, *62B*(1), 53–60. https://doi.org/10.1093/GERONB/62.1.P53

Tanaka, J. W., & Sung, A. (2016). The "eye avoidance" hypothesis of autism face processing. *Journal of Autism and Developmental Disorders*, *46*(5), 1538–1552. https://doi.org/10.1007/S10803-013-1976-7

Team, R. C. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Thompson, E. R. (2007a). International Positive and Negative Affect Schedule — Short Form (I-PANAS-SF). *PsycTESTS*. https://doi.org/10.1177/0022022106297301

Thompson, E. R. (2007b). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227–242. https://doi.org/10.1177/0022022106297301

Thorstenson, C. A., Pazda, A. D., & Lichtenfeld, S. (2020). Facial blushing influences perceived embarrassment and related social functional evaluations. *Cognition and Emotion*, *34*(3), 413–426. https://doi.org/10.1080/02699931.2019.1634004

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405. https://doi.org/10.1177/1754073911410747

Trilla, I., Weigand, A., & Dziobek, I. (2021). Affective states influence emotion perception: evidence for emotional egocentricity. *Psychological Research*, *85*(3), 1005–1015. https://doi.org/10.1007/s00426-020-01314-3

Troiano, E., Padó, S., & Klinger, R. (2021). *Emotion Ratings: How Intensity, Annotation Confidence and Agreements are Entangled*.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality* *and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV): A Set of Videos Expressing Low, Intermediate, and High Intensity Emotions. *PLOS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0147112

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2018). Sex differences in facial emotion recognition across varying expression intensity levels from videos. *PLoS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0190634

WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. (2013). *World Medical Journal*, *59*(5), 199–202. https://eds-b-ebscohost-com.ludwig.lub.lu.se/eds/detail/detail?vid=5&sid=4d144a1c-87a4-42f3-adeb-3a81299613c7%40sdc-v-sessmgr02&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=93262173&db=a9h

Wong, B., Cronin-Golomb, A., & Neargarder, S. (2005). Patterns of visual scanning as predictors of emotion identification in normal aging. *Neuropsychology*, *19*(6), 739–749. https://doi.org/10.1037/0894-4105.19.6.739

Yoshikawa, S., & Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cognitive, Affective and Behavioral Neuroscience*, *8*(1), 25–31. https://doi.org/10.3758/CABN.8.1.25

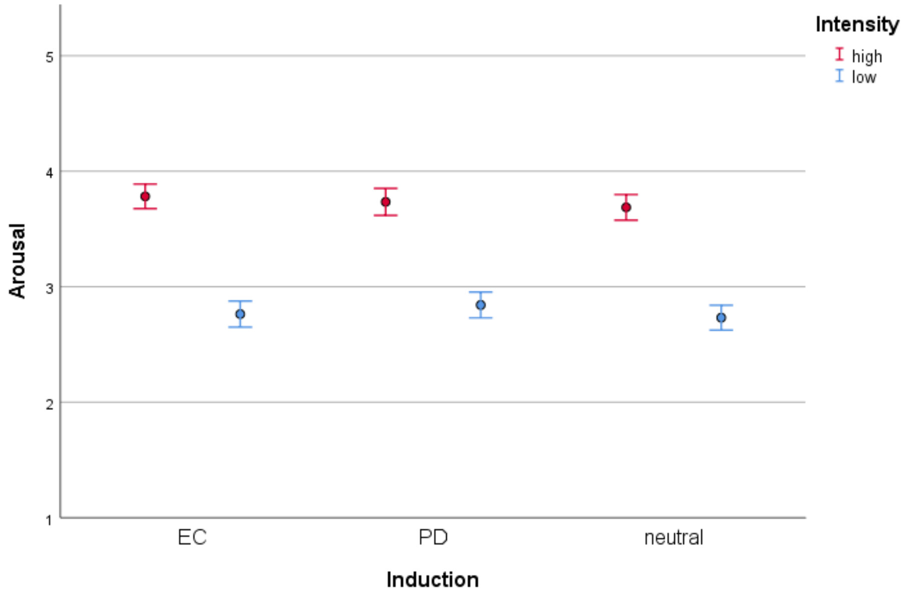
Zaki, J. (2014). Empathy: A motivated account. In *Psychological Bulletin* (Vol. 140, Issue 6, pp. 1608–1647). American Psychological Association Inc. https://doi.org/10.1037/a0037679

Appendix A

Additional Figures for the Dimensional Ratings of Facial Expressions

A1.

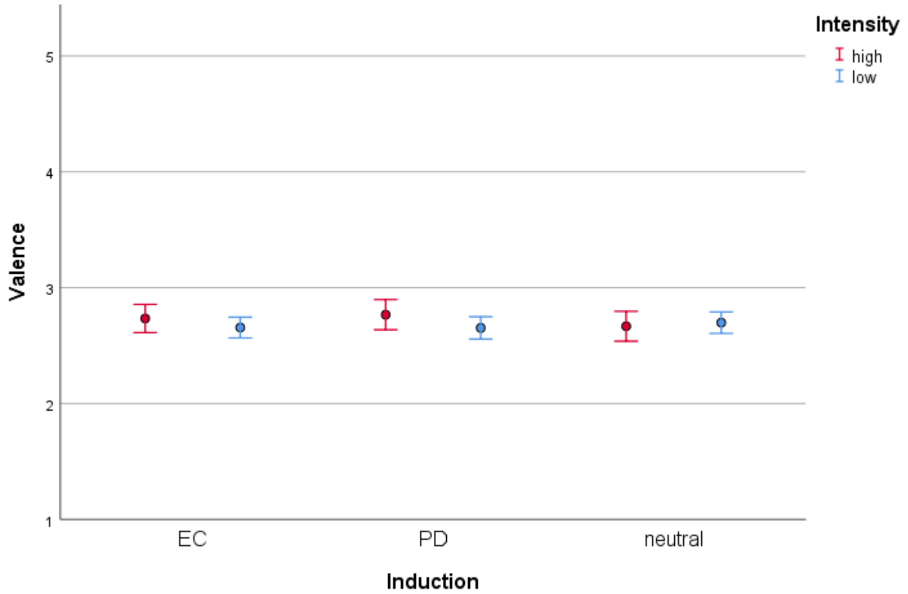
Graphical depiction of arousal ratings per empathic induction condition and expressive intensity.



Note. Red markers show the average arousal rates of high-intensity expressions, and blue markers the average arousal rates of low-intensity expressions (ratings were made on a scale from 1-5). Induction conditions on the x-axis are empathic concern (EC), personal distress (PD), and neutral. Error bars represent 95 % confidence intervals.

A2.

Graphical depiction of valence ratings per empathic induction condition and expressive intensity.



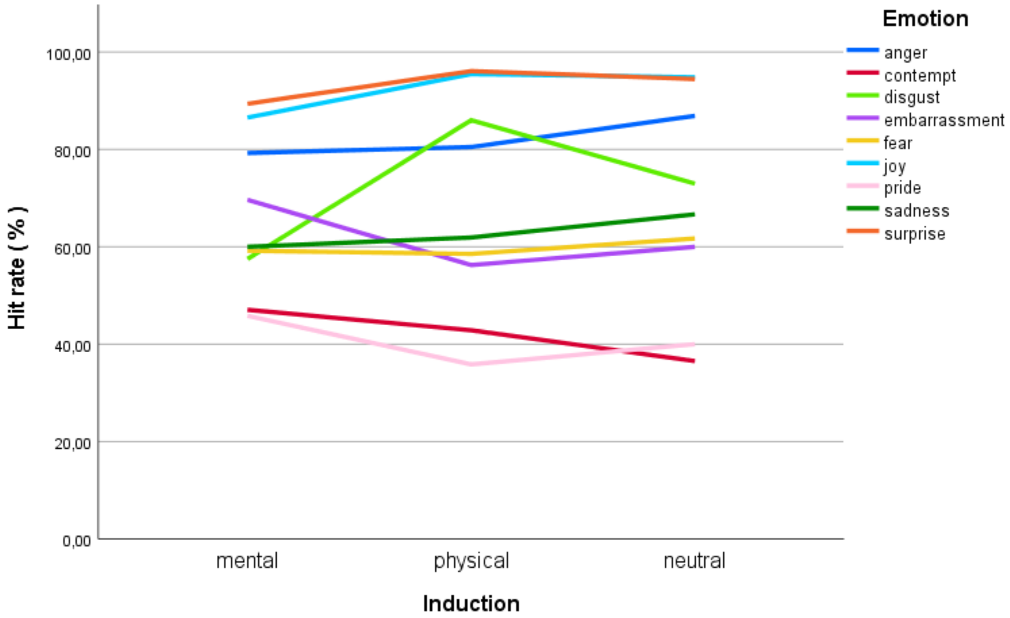
Note. Red markers show the average valence rates of high-intensity expressions, and blue markers the average valence rates of low-intensity expressions (ratings were made on a scale from 1-5). Induction conditions on the x-axis are empathic concern (EC), personal distress (PD), and neutral. Error bars represent 95 % confidence intervals.

Appendix B

Additional Figures for the Main 3 (Induction) x 2 (Intensity) Analysis per Emotional Expression

B1.

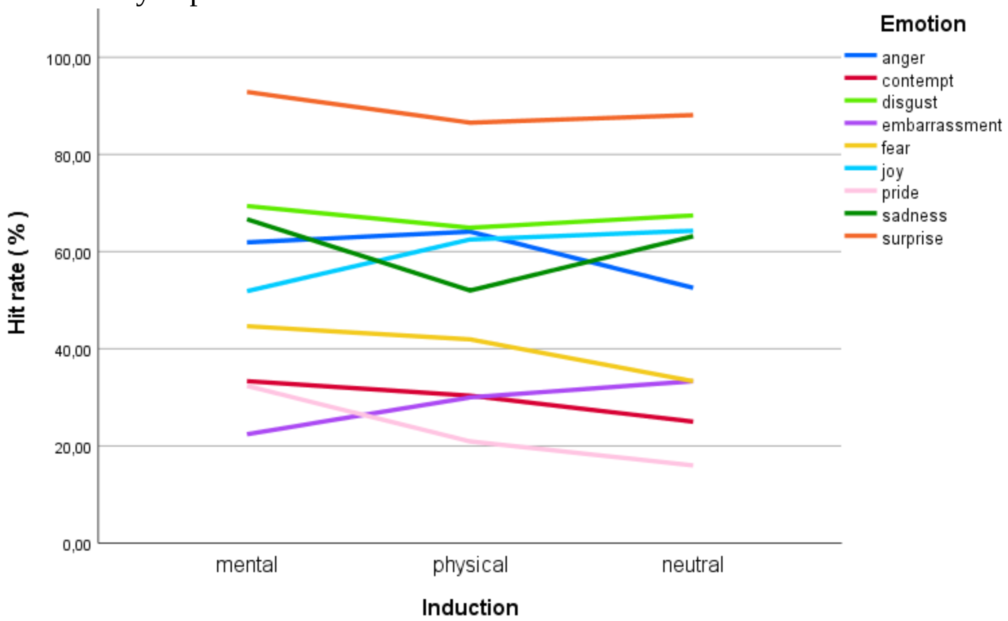
Graphical depiction of correct ratings per emotion and empathic induction condition in % for high-intensity expressions.



Note. Mental pain was presented to induce empathic concern. Physical pain was presented to induce personal distress. The neutral induction included a vignette where neighbors greet one another.

B2.

Graphical depiction of correct ratings per emotion and empathic induction condition in % for low-intensity expressions.



Note. Mental pain was presented to induce empathic concern. Physical pain was presented to induce personal distress. The neutral induction included a vignette where neighbors greet one another.