Emotion in Action:
Correlates of Emotion Recognition from Body Movements

Inauguraldissertation zur Erlangung des Doktorgrades

der Naturwissenschaften

doctor rerum naturalium (Dr. rer. nat.)

Julia Bachmann

Gießen, 2022

Emotion in Action:

Correlates of Emotion Recognition from Body Movements

Inauguraldissertation zur Erlangung des Doktorgrades

der Naturwissenschaften

doctor rerum naturalium (Dr. rer. nat.)

Justus - Liebig - Universität Gießen

Fachbereich 06 - Psychologie und Sportwissenschaften

Vorgelegt von

Julia Bachmann

Als Dissertation vom Fachbereich Psychologie und Sportwissenschaft

der Justus-Liebig-Universität Gießen angenommen.

Erstgutachter: Prof. Dr. Jörn Munzert

Zweitgutachterin: PD Dr. Britta Krüger

Danksagung

Ich möchte mich bei allen Menschen bedanken, die mich während meiner Promotion begleitet und unterstützt haben.

Ich danke meinem Doktorvater Prof. Dr. Jörn Munzert, der mir die Möglichkeit gegeben hat, mich als Wissenschaftlerin zu entwickeln und meine Forschungsfähigkeiten zu erweitern. Danke auch für die Denkanstöße, die meine Arbeiten verbessert haben.

Ich bedanke mich auch herzlich bei PD Dr. Britta Krüger, die mich auf beruflicher und privater Ebene unterstützt hat und immer ermutigende Worte gefunden hat. Außerdem bedanke ich mich für die zahlreichen fruchtbaren Diskussionen, die mich auf meinem Weg voran gebracht haben.

Ich bedanke mich außerdem bei Prof. Dr. Gudrun Schwarzer und Prof. Dr. Gebhard Sammer, dafür, dass sie Teil meiner Prüfungskommission sind.

Desweiteren bin ich dankbar, Teil des internationalen Graduiertenkollegs (IRTG 1901) gewesen zu sein, welches mir meine Promotion ermöglicht hat.

Besonders bedanken möchte ich mich bei meinem guten Freund und Kollegen Adam Zabicki, mit dem ich einen großen Teil meines Weges zur Promotion gemeinsam bestreiten durfte, der in allen Situationen für mich da war und mich auch in weniger guten Zeiten aufgemuntert und unterstützt hat. Danke!

Natürlich möchte ich mich auch bei allen weiteren Kollegen für die schöne Zeit und die gemeinsamen Erlebnisse bedanken. Viele von euch sind über die Zeit enge Freunde geworden, die ich nicht missen möchte.

Ich möchte meiner Familie, die mich bei all meinen Vorhaben unterstützt und in meinem Leben eine wichtige Stütze ist, ganz besonders danken.

Ich möchte besonders Anton von ganzem Herzen danken, der meine Erfolge mit mir gefeiert und mich in schwierigen Zeiten unterstützt hat. Danke für deine Unterstützung, deine Liebe und das Vertrauen, das du in mich setzt. Diese Arbeit ist ganz besonders dir gewidmet.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus eröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.

Gießen, der 22.12.2022, ____________________________________

Julia Bachmann

Abstract

Human emotions can be viewed as a principal currency for relationships as they are often the underlying motivational force for human behavior. Thus, the ability to perceive and correctly interpret the emotions of our conspecifics constitutes a core competence and is of great adaptive value, as it allows us to navigate our behavior towards an adequate response. The human body has been identified as a potent source to quickly retrieve a wealth of information about affective states. In this regard, previous research has discussed factors that modulate emotion recognition, including kinematic and postural properties of the movement, the presence (or absence) of an interacting partner as well as individual characteristics of the observer.

The major aim of this dissertation is to establish further factors that modulate the perception of emotions from body movements from interactions. To tease these factors apart, three experiments were carried out. The first experiment investigates which parts of the body convey the most salient information that ultimately allows for an identification of the respective affective state. Moreover, it is addressed how the emotional expressivity of the observer influences emotion recognition. The second experiment assesses how the spatiotemporal coupling of the agent‘s actions influence emotion recognition. The third experiment extends previous research designs by implementing a novel virtual reality apporach to investigate whether the recognition of emotion is influence by the perceived (co)presence with the interacting agents. Moreover, gender differences are assessed using a subjective valence rating scale.

The results indicate that the information that is retrieved by the observer to recognize an emotion depends on the type of emotion that is displayed. Whereas the emotions anger and happiness can be recognised more reliably from arm movements, sadness seems to be perceived mainly via the cues from the head and torso. Moreover, our results suggest that the recognition of anger and affection depends on a spatiotemporal coupling of the agent’s actions while the recognition of happiness and sadness does not. With respect to interindividual differences, we found that more emotionally expressive individuals display higher emotion recognition accuracies than individuals with lower subjective emotional expressivity. Lastly, our results do not support the notion that there are gender differences regarding the valence perception of emotional interactions.

Contents

1 Relevance and background

„The human body is the best

picture of the human soul."

– Ludwig Wittgenstein

1.1 Emotions as action dispositions

Imagine a person walking towards you, expanding their chest, clenching and lifting their fists - breathing heavily. Instantly, we may feel that the person we are looking at is filled with anger and that we should probably draw away from this potential threat.

Humans are able to quickly retrieve a wealth of information through mere observation of a person. Being able to correctly identify and interpret the internal state of our conspecifics allows us to form adequate responses to these states (Bradley et al., 2001). For instance, when we see a loved one cry, we soothe them; when a friend is happy, we are happy with them; if a person is angry, we may give them space. Thus, in a way, emotions can be viewed as one principal currency for human relationships as they are often the underlying motivational force for our behavior (Bradley et al., 2001; Motta et al., 2009; de Gelder & de Borst, 2015).

Not being able to identify the emotional states of our conspecifics can lead to interpersonal difficulties or even social avoidance and isolation (Persad & Polivy, 1993). Different psychiatric disorders such as schizophrenia, major depressive disorder, or autism spectrum disorder are characterized by marked emotion recognition difficulties which are thought to contribute to the difficulties in everyday social functioning (e.g. Kee et al., 2003; Kaletsch et al., 2014; Kohler et al., 2010; Tse et al., 2004; Trevisan & Birmingham, 2016; Krüger et al., 2018). Thus, the ability to identify and infer emotional states can be critical to successful social interactions.

In fact, the ability to recognize emotions develops early in life. Infants progress from *discrimination* (i.e. the ability to detect differences between two emotional body postures or vocalizations) to the *recognition* (i.e. the ability to, interpret the emotional expression and exhibit some level of understanding of the underlying affect) of emotion in faces during the first half year of life (Walker-Andrews, 1998).

Heck and colleagues (2018) showed that at five-months old, infants display successful *intermodal matching*. More specifically, the five month-old infants looked at videos of upright emotional body movements for a longer period of time if they were paired with matching emotional vocalizations. For instance, when angry movements were presented with angry voices, infants watched the videos for a longer period of time than when angry movements were paired with happy voices. This suggests that the five-month old infants were able to match the emotional body movements to the corresponding vocal emotions and thus understand their meaning to a certain degree.

In fact, there are multiple channels through which affective information can be conveyed. For instance, faces, bodies and prosody. Although research on emotion recognition has mainly been based on emotions expressed via the face or via prosody for a long time, another channel, that has been largely neglected, has come into the spotlight in the past decade: the human body. It has become evident that the human body is a potent source for conveying internal states, such as intentions or emotions (e.g., Atkinson et al. 2004; de Gelder, 2009; Bachmann et al., 2018).

However, research continues to tackle many questions that are associated with the ability to retrieve information about emotional states from body movements. Influencing factors that shape our perception may include the type or amount of information that is availabe or individual factors such as personality traits. In the following, I will give an overview of factors that influence our ability to infer mental states from the (body) movements. Moreover, I will elaborate on why our own experiences and behavior influences what we ultimately perceive.

1.2 Emotional Body Language

"Considering the emotional value of bodily expressions, it is somewhat surprising that the study of perception of whole-body expressions lags so far behind that of facial expressions." (Van den Stock et al, 2007)

While the role of the body in emotion recognition has not been a primary focus of research for an extended period, theorists such as Darwin (1872) and Frijda et al. (1989) have long acknowledged the strong connection between emotions and actions, leading them to highlight the importance of the body (de Gelder, 2009). For more than a decade now, however, behavioral as well as neuroimaging research has expanded its view, including body movements to investigate emotion recognition (for a review, see de Gelder, 2009).

1.2.1 Emotion models

Due to the emphasis on facial expressions of emotion, most emotion models only indirectly allow for predictions about emotional body movements and the cues that observers use to interpret them. Take the basic emotion theory (BET), for instance (Ekman, 1992, Keltner et al., 2003; Izard, 1992; Karon & Tomkins, 2008). Although perhaps over-simplified, the BET views emotions as a limited set of discrete entities (i.e. fear, anger, happiness, disgust, sadness, surprise), with each of these *basic* emotions manifesting within a *fixed* behavioral pattern that can be observed and identified even across different cultures (Ekman et al., 1969). Although Eckman (1992) suggested a fixed behavioral pattern of emotions, most observations were centered around *facial expressions*. Thus, one can only speculate that this assumption also holds true for the bodily expression of emotion.

Other models, in contrast, do not reduce emotions and their expression to a limited number of discrete entities or dimensions (as seen in dimensional models of emotion, see for instance Osgood, 1966; Russell 1983) but rather highlight the complexity of emotional experience and expression and its dependence on cultural influences (Russell, 1994). Componential models of emotion, for instance, suggest that (emotional) motor expression is a direct effect of appraisal (e.g. intrinsic pleasantness, goal conduciveness, coping potential) and can thus have multiple different outcomes (e.g. Scherer et al., 2009). In this context, Scherer (2009) has made predictions with respect to bodily representations of emotion. He suggested, for instance, that unpleasant stimulus events might be accompanied by centrifugal hand and arm movements, shrinking posture and avoidance locomotion.

Constructionist theories (e.g. Averill, 1980; Barrett, 2006), in turn, view emotions as categories in a much broader sense, inspired by Charles Darwin who, for instance, viewed the concept of a *species* in terms of a population of unique individuals with different phenotypes (Darwin, 1859). Applied to the (bodily) expression emotion, this means that variation is the norm (Barrett, 2006). Happiness, for instance, can have many different expressions. You might jump for joy, you might smile or laugh loudly, you might scream or even stand still, enjoying the tingling feeling inside, all depending on whatever fits the situation. Based on this view, we construct an emotion and its corresponding behavior, depending on the situation, resulting in a number of different expressions.

1.2.2 Measuring emotion recognition from body movements

Overall, emotional body movements, often termed *emotional body language* (EBL), can be viewed as a composition of body parts resulting in coordinated movements and often meaningful action. The question arises, however, what information in particular we use to make sense of it, i.e. to recognize a certain emotion. In order to answer this question, researchers in cognitive sciences have established different paradigms to measure the recognition of EBL. In most cases, participants are asked to observe images or video sequences of emotional displays and subsequently assign it to an emotion category (e.g. anger, happiness, sadness, etc.) or judge its emotional valence (i.e. negative or positive) on a dichotomous scale or on a continuum (e.g. Atkinson et al., 2004; Alaerts et al., 2011; Lorey et al., 2012; Ross et al., 2012).

The types of emotional displays vary in their level of abstraction, with less abstract stimuli such as static images and short video sequences, or more abstract stimuli such as point light displays (PLDs) presented using desktop designs. PLDs provide only kinematic and configural information stemming from a few points representing the joints of the body (Johansson, 1973). So far, many studies have confirmed that emotional expressions can indeed be identified from even minimal kinematic information such as PLDs (e.g. Atkinson et al., 2004; Alaerts et al., 2011; Lorey et al., 2012).

However Atkinson et al. (2004) showed that the use of full-light displays leads to slightly better recognition accuracies as compared to PLDs, perhaps due to the fact that not only configural but also form information is available. Yet, images (i.e. full light displays) are harder to control and the observer is more susceptible to possible confounding factors such as attractiveness, sympathy or cultral aspects (Hoffmann et al., 2010).

While most of these experimental designs can be thought of as classical desktop designs (i.e. non-immersive and two-dimensional), in which the participants observe a certain stimulus on the screen, more recent paradigms extend these methods by using virtual reality. A virtual environment allows for an administration of highly-controllable dynamic stimuli, yet within a more realistic setting (Geraets et al., 2021). Thus, the experience within a virtual reality allows for a subjective feeling of *presence*. A high degree of presence is thought to create a sense of physical presence (i.e. “being there”) which leads participants to behave as if they were in the real world (Slater et al., 1997; Heater, 1992). Similarly to PLDs, they are highly controllable, yet preserve form and movement information. As this is a relatively new paradigm within the field of emotion recognition research, the body of research on emotion perception from avatars is still limited (for an overview, see Marin-Morales et al., 2020; Geraets et al., 2021; Souto et al., 2020).

1.2.3 The role of kinematic and postural information

Using these various methods, researchers have investigated which parameters influence the perception of emotions. In an early study, Michalak and colleagues (2009), for instance, aimed to assess the relationship between gait patterns and sadness as well as depression. In two different experiments they examined patients suffering from major depression as well as undergraduate students who were subjected to a mood induction (positive and negative). They found that gait patterns associated with sad mood and depression are characterized by reduced walking speed and arm swing, vertical head movements as well as a slumped posture and larger swaying movements of the upper body.

Studies focusing on the *perception* of emotions highlighted that there may be distinctive patterns of posture and movement depending on the type of emotion that is displayed. While more active emotions (e.g. anger and happiness) are associated with greater gestural movement, speed, force and expansiveness (de Meijer,1989; Wallbott, 1998; Wallbott & Scherer, 1986), emotions such as sadness are characterized by a lack of movement (Dael, et al., 2012). When postural (e.g. limb angles and symmetry) and kinematic features (e.g. velocity, acceleration) are directly compared, postural features seem to discriminate better between emotion categories overall (Poyo Solanas et al., 2020).

With respect to emotion recognition processes, it has been suggested that emotional states can be recognized either by a recognition-as-a-whole strategy, in which the entire object is presented or even by a recognition-by-parts strategy in which the stimulus is decomposed into simple elements such as individual body parts (Neri, 2009). This idea is supported by Pollick et al. (2001), who demonstrated that emotions can readily be from point-light displays of isolated arm movements. The researchers presented knocking and drinking movements with ten different affects (i.e. afraid, angry, excited, happy, neutral, relaxed, sad, strong, tired and weak) and then asked participants to categorize the respective affect. They found that the respective emotion can readily be identified from arm movements and that the perceived emotional activation (i.e. intensity) relates positively to temporal movement parameters, such as velocity, acceleration, or jerk.). In this regard, Neri (2009) argued that individual body-part processing occurs at an early, bottom-up processing stage in which actions can already be identified. They emphasize that this part-based representation allows for more flexible encoding of actions.

1.2.4 Recognizing emotion from interactions

Emotional expression, however, does not only occur at the intrapersonal level but also within dynamic interactions between the individual and the environment to serve its communicatory function (Blair, 2003). For instance, already children as young as two years old display non-approach (i.e. *not* moving towards) or even avoidance behavior (moving away) when confronted with an angry person (e.g. de Rosnay et al., 2006; Walle et al., 2017).

Yet, most research on emotion recognition has been carried out using single agent paradigms in which one agent expresses a certain emotion. Interestingly, research suggests that the recognition of affective states seems to be facilitated by the observation of dyadic interactions rather than an individual's emotional display (e.g. Clarke et al., 2005; Lorey et al., 2012). Clarke and colleagues (2005), for instance, asked participants to categorize emotions (i.e. sadness, anger, joy, disgust, fear, romantic love) from point-light displays of either dyadic interactions or single agents that were isolated from the interaction sequence. The authors found that performances for the recognition of joy and romantic love were markedly decreased when single agents were displayed. Moreover, Lorey and colleagues (2012) found that observing social interactions (i.e. dyadic displays of emotion) facilitates the perception of the corresponding valence (i.e. positive vs. negative) of the scene and increases how confident participants feel about their evaluation.

Within interactions, not only purely kinematic and postural cues of each agent give information about the (emotional) content of the interaction. Research paradigms investigating interactive activity found that the visual system relies on the *spatiotemporal coupling* between the interactive agents, i.e. the dependence of the agent's actions in space and time. More specifically, within meaningfully synchronized interactions (e.g. fighting or dancing), the actions of one agent serve as the predictor for the expected actions of the other agent (e.g. Neri et al., 2006; Manera et al., 2011, 2013; von der Lühe (2016). Manera and colleagues (2011, 2013) showed that observing communicative gestures of one agent enhances the visual discrimination of a second agent who responds to the gesture - a phenomenon that has been termed as interpersonal predictive coding. It suggests that we possess implicit knowledge about the natural dynamics of human interaction that guides the processing of motion patterns generated within an interaction. Thus, when we try to make sense of interactions, we seem to anticipate how it will unfold (von der Lühe, et al., 2016).

1.2.5 Interindividual differences

So far, we have established that our perceptual impression of emotion may be formed not only through kinematic and postural information but also by the way actions between two agents are coordinated. However, it is assumed that our perception of the world and the people around us is shaped profoundly by individual experiences that are stored as representations in our brain (for a review, see Decety & Sommerville, 2003). This idea is strengthened by research which demonstrates that individual characteristics can modulate the way we perceive emotions. These characteristics include the gender of a person, their mood, or their own ability to get in touch with their own emotions (e.g. Alaerts et al., 2011; Lorey et al., 2012; van der Veen et al., 2007). There is evidence which suggests that females are generally faster and more accurate at identifying emotions conveyed through bodily cues (Alaerts et al., 2011; Sokolov et al., 2011). However, with regard through emotions expressed through facial expressions, this gender difference appears to only be present for more subtle, as opposed to stereotypical, displays of emotion (Hoffmann et al., 2010; Montagne et al., 2005).

Clinical research studies have frequently shown that persistent negative mood, as it can be observed in patients with affective disorders such as depression, for instance, can lead to a negative bias, i.e. depicted emotional content is rated more negatively compared to healthy controls (e.g. Kaletsch et al., 2014). Moreover, Qiao-Tasserit and colleagues (2017) investigated how induced affective states as well as affective traits such as depression and anxiety influence emotion perception (i.e. fear and happiness). They found that negative (vs.

neutral) mood increased participants' tendency to classify ambiguous faces as fearful. Moreover, affective traits such as anxiety and depression had a stronger effect than temporary affective states and increased the tendency to classify ambiguous faces as fearful (Qiao-Tasserit et al., 2017).

In a non-clinical sample, Lorey et al. (2012) used point-light displays of emotional body movements to investigate whether individuals who have shortcomings in recognizing and describing their own emotions are compromised in their ability to assign the sequences to a valence category. They showed that these individuals were significantly less confident when deciding whether a presented scene is positive or negative although their valence ratings (i.e. whether the observed scene is positive or negative) did not differ from people without these constraints. Goldman and Sripada (2005) also postulated a close relationship between the ability and willingness to produce emotional behaviour and the ability to recognize affective states in others.

2 Research gap and outline

Although the perception and recognition of emotional body language has been studied more extensively using different tasks (see section *1.2.2*) in the past decade, the debate of which information in particular is used to create meaning from what we are seeing is still ongoing. Multiple factors have been discussed to drive and modulate the recognition process, including kinematic and postural properties of the movement (e.g. de Meijer,1989; Wallbott & Scherer, 1986, Dael, et al., 2012; Poyo-Solanas et al., 2020), the presence of an interacting partner (e.g. Clarke et al. 2005; Lorey et al., 2012) as well as individual characteristics of the observer, such as the gender of a person or one's own ability to recognize emotions in oneself and others (Alaerts et al., 2011; Lorey et al., 2012).

The present dissertation aims to expand our understanding of factors underlying emotion recognition from body movements, a skill of great societal relevance. Specifically, we extent the existing literature by using bodily displays of emotional interactions, rather than single agents, to investigate which information in particular is retrieved to form the perceptual impression of an emotion. Investigating these factors has great potential to advance our knowledge of how it is possible to infer mental states by observing others and paves the way towards understanding why in some disorders, this ability is compromised (Kaletsch et al., 2014; Dalili et al. 2015; Krüger et al., 2018). In the following, three studies will be outlined with respect to research aims and the methods that have been used.

The first study presented in this dissertation investigates which parts of the body convey *key information* that ultimately allows for an identification of affective states from interactions. Although specific kinematic and postural parameters have been linked to different emotional states (for an overview, see Kleinsmith and Bianchi-Berthouze, 2013), it remains unclear what parts the body in particular convey the most salient information. Moreover, as a close relationship between the ability and willingness to produce emotional behavior and the ability to recognize affective states in others has been postulated (Goldman & Scripada, 2005). Here, we picked up on this issue by addressing how the emotional expressivity of the observer influences emotion recognition. To do so, an emotion recognition paradigm was used in which participants observed two types of impoverished PLDs depicting four different emotional interactions (i.e. happiness, anger, sadness, affection). In the first step, participants were asked to observe and subsequently rate the emotion category. In the next step, they were asked to rate the confidence in their judgment. Emotional expressivity of the observer was assessed via one of the subscales of the German version of the Emotional Competence Questionnaire (Emotionale-Kompetenz-Fragebogen, EKF; Rindermann, 2009). The interactions were either depicted as arm or trunk movements, which allowed us to identify key body parts. Moreover, it was evaluated how the the use of the information that is presented interacts with the characteristics of the observer (i.e. the observer's own emotional expressivity).

In **the second study**, we further expand current research by identifying how spatiotemporal properties that are specific to emotional interactions (i.e. between two agents) influence the recognition. In particular, we asked whether the recognition of affective states depends on a spatiotemporal coupling of the agent's movements. To answer this question, an emotion recognition task including four different affective interactions (anger, sadness, happiness, affection) was used. The actions of two interacting PLDs were manipulated by implementing various temporal offsets which delayed the onset of one of the agent's actions (+ 0ms, +500ms, +1000ms, +2000ms). Participants observed the scenes and were asked to determine the emotion category (anger, sadness, happiness, affection) as well as the emotional valence intensity (i.e. how positive or negative) on a 11-point likert scale. The latter allows for a deeper understanding of more subtle perceptual differences as compared to the category rating.

Finally**, the third study** uses a novel multi-system virtual reality research design to investigate whether the perceptual impression of emotional interactions is shaped by the perceived presence of the interacting agents. Since most of the existing evidence is based on desktop designs in which the observer is physically separated by a screen, a virtual reality approach allows for the observer to share a space with the observed agents. Previous research has demonstrated that immersive virtual environments (i.e., an environment in which the participant feels physically “present”) can increase the emotional response of the viewer and lead to a more intense perception of the content being viewed (Estupiñán et al. 2014; Visch et al. 2010). Thus, in the present study, two avatars displayed emotional interactions (anger, sadness, affection, happiness) within two different conditions. In the “pictorial” condition, the emotionally interacting partners were displayed on a screen within the virtual environment, thus creating a sense of physical separation. In the “visual” condition, participants were able to share a common space with the agents, with the aim of creating an increased sense of co-presence and agency. After each presentation, explicit valence perception (i.e. positive or negative) was assessed on a 11-point likert scale. In this regard, it was also assessed whether there are gender differences with respect to subjective emotion recognition.

3 Emotional expressivity of the observer mediates recognition of affective states from human body movements

This chapter has been published as:

Bachmann, J., Zabicki, A., Munzert, J., & Krüger, B. (2020). Emotional expressivity of the observer mediates recognition of affective states from human body movements. *Cognition and Emotion*, *34*(7), 1370–1381. doi: 10.1080/02699931.2020.1747990

COGNITION AND EMOTION
2020, VOL. 34, NO. 7, 1370–1381
https://doi.org/10.1080/02699931.2020.1747990

Emotional expressivity of the observer mediates recognition of affective states from human body movements

Julia Bachmann, Adam Zabicki, Jörn Munzert and Britta Krüger

Department of Psychology and Sport Science, Justus-Liebig-University, Giessen, Germany

ABSTRACT
Research on human motion perception shows that people are highly adept at inferring emotional states from body movements. Yet, this process is mediated by a number of individual factors and experiences. Within this study, we tackle two questions. Firstly, we ask which part of the body transmits the key information that is used to infer affective states. Secondly, we address how the observer's own emotional expressivity influences the recognition process. We used two types of impoverished point-light displays depicting the same emotional interactions as either arm or trunk movements. Results showed that participants used different sources of information in an emotion-specific manner. Participants with richer self-reported emotional expressivity showed higher recognition accuracies overall but also benefited more from information delivered by arm gestures. We interpret our findings in terms of embodied simulation, suggesting that emotion perception constitutes a function of the expressing body and the individual observer.

ARTICLE HISTORY
Received 4 July 2019
Revised 7 January 2020
Accepted 10 March 2020

KEYWORDS
Emotional expressivity; emotion recognition; point-light displays; embodied simulation

In higher primates and particularly in humans, the ability to perceive and experience emotions constitutes a core competence. Emotions can be viewed as the "principal currency" for human relationships, and they are often the underlying motivational force for human behaviour. This capability to perceive and correctly interpret another individual's affective state automatically – that is, without conscious effort – is of great adaptive value, because it allows us to navigate our behaviour toward an adequate response.

Emotional states can be inferred from different information sources. Alongside facial expressions and speech, human body movements present one major source of information for recognising an individual's affective state (Bachmann et al., 2018; de Gelder, 2006; de Gelder et al., 2014). Individuals are able to infer emotional states by observing body movements, even when the provided stimulus information is reduced to a minimum. Point-light displays (PLDs), for instance, present such a form of highly reduced stimulus material, as they contain only kinematic and configural information, eliminating possible confounding factors, such as attractiveness, sympathy or cultural aspects (Hoffmann et al., 2010).

Generally, human body language can be considered a composition of body posture, trunk and arm movements, gestures, and the use of space to convey information such as the felt emotion (Dael et al., 2012). Several studies have demonstrated that human body posture (de Meijer, 1989; Hadjikhani & de Gelder, 2003; Schindler et al., 2008; Wallbott, 1998) and arm gestures (Dael et al., 2012; Montepare et al., 1999; Pollick et al., 2001) convey emotional information quite reliably. In this regard, Neri (2008) argued that individual body-part processing occurs at an early, bottom-up processing stage in which actions can already be identified. They emphasise that this part-based representation allows for more flexible encoding of actions. However, the question remains which part of the body carries more salient information for emotion recognition. The answer to this question may rely on the characteristics of the specific emotional category. Wallbott and Scherer (1986) pointed out that there seem to be distinctive

CONTACT Julia Bachmann julia.bachmann@sport.uni-giessen.de
Supplemental data for this article can be accessed at https://doi.org/10.1080/02699931.2020.1747990.

patterns of movement and postural behaviour associated with certain emotions, which may allow for a reliable distinction between emotional categories. More "active" emotions such as anger and happiness are expressed through greater gestural movement activity, speed, force, or expansiveness (de Meijer, 1989; Wallbott, 1998; Wallbott & Scherer, 1986). In contrast, sadness is an emotion related to apathy, hypo activity, and the absence of every form of action (Dael et al., 2012). Thus, it seems possible that emotions like anger and happiness are perceived more easily from arm movement than movement of other body parts, while this might not be the case for sadness sequences due to the relative lack of gestural movements.

It has been proposed that one underlying mechanism mediating the perception and recognition of emotions in human body movements is embodied simulation. This means that an observer arrives at a mental attribution by simulating, replicating, or reproducing her or his own representations of being in the same state as the observed person (Barsalou et al., 2003; Gallese, 2003; Gallese & Goldman, 1998; Goldman & Sripada, 2005; Rizzolatti & Sinigaglia, 2010). In the case of emotion perception, seeing someone else's emotional expression might be linked to a simulation of the respective emotion in oneself and, therefore, to experiencing one's own emotions (Goldman & Sripada, 2005; Niedenthal et al., 2009; Wicker et al., 2003). Thus, embodied simulation suggests that our perception of the world and the people around us is shaped profoundly by individual experiences held in our brain's representations (for a review, see Decety & Sommerville, 2003). This notion has been underpinned by research on the modulating influence of observer characteristics on perceptual processes. For example, it has been demonstrated that individual characteristics such as gender, mood, or the ability to get in touch with one's own emotions modulate emotion perception (Alaerts et al., 2011; Lorey et al., 2012; Van der Veen et al., 2007). Furthermore, Goldman and Sripada (2005) postulated a close relationship between the ability and willingness to produce emotional behaviour and the ability to recognise affective states in others.

On this background, the present study formulates two major issues: First, it is still not clear which aspect of human body language carries the most salient information from which to infer affective states. Thus, we tried to disentangle which part of a dynamic body movement provides the key information for emotional action recognition and whether this is emotion-specific. Second, it remains unclear how the ability to express one's emotions affects the recognition of emotions expressed by others. More specifically, we asked whether differences in this ability relate, for example, to a specific usage of information given by trunk or arm movements to infer an emotional state. Therefore, we used PLDs of human interactions depicting the kinematics and configural information of various emotions (anger, sadness, affection, happiness). The PLDs were presented within an occlusion paradigm as arms-only (i.e. shoulders, elbows, wrists) or trunk-only (i.e. head, trunk, and legs) displays with full-body displays serving as a baseline condition. Participants were asked to rate the depicted emotional category of the PLDs and report their subjective confidence in their assessments. We created post hoc groups with varying levels of emotional expressivity to examine the influence of this factor on emotion recognition.

We hypothesised that (1) participants would be able to infer emotions from the depicted PLDs (serving as the precondition); (2) participants with an enriched emotional expressivity would show higher recognition accuracies in their evaluations of full-body displays; (3) the depiction of reduced stimuli (i.e. dynamic trunk and arm movements) would negatively impact recognition accuracies, presumably in an emotion-specific manner; and (4) richness of emotional expressivity in the observer would positively relate to recognition accuracies and confidence ratings for reduced stimuli.

1. Methods and materials

1.1. Participants

Thirty adults, including 17 females (mean age = 23.3 years, $SD = 5.3$ years) and 13 males (mean age = 25.2 years, $SD = 4.3$ years) with normal or corrected-to-normal vision participated in the study. None reported any history of psychiatric or neurological disorders and they had no history or current use of psychoactive medication. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the local ethics committee. The German version of the Emotional Competence Questionnaire (Emotionale-Kompetenz-Fragebogen, EKF; Rindermann, 2009) was used to assess the participant's score within four ability domains. The first subscale assesses the ability to recognise one's own emotions; the second, the ability to recognise emotions in others; the third, the ability to regulate and control one's emotions; and the fourth, the ability to express one's emotions, both nonverbally and verbally. The EKF contains 62 items that have to be rated on 5-point scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Scores ranged from 91.68 (within average) to 125.68 (above average) ($M = 108.57$, $SD = 10.04$).

Beck's Depression Inventory (BDI-II; Beck et al., 1996) and the State–Trait Anxiety Inventory (STAI; Spielberger et al., 1983) were applied to control for each participant's current emotional state. BDI-II scores ranged from 0 to 20 ($M = 9.0$, $SD = 5.5$). Five participants displayed scores greater 13, indicating a depressed mood. STAI-X1 scores ranged from 24 to 60 ($M = 38.5$, $SD = 8.3$); STAI-X2, from 27 to 57 ($M = 39.7$, $SD = 9.5$). Although some of these participants showed clinically relevant scores (i.e. a BDI-II total score above 13), it did not affect their emotion recognition behaviour (see supplementary material).

1.2. Creating point-light stimuli

Eight pairs of non-professional actors were asked to portray four emotional states within an interaction: happiness, sadness, anger, and affection. Each emotional state was acted out in three intensities (low, medium, and high) in order to obtain a variety of emotional expressions. To ensure a symmetric behavioural pattern, all participants received scripts of emotional situations that they were instructed to perform. Subjects were asked to act intuitively within the context of the given situation, thereby allowing freedom and subsequently variability in their expression. Moreover, both actors in a scene were instructed specifically to perform the same emotion in order to produce a congruent behavioural pattern.

The interactions were recorded with an optical motion capture system (Vicon Motion Systems, Oxford, England) equipped with 10 cameras. Forty-one reflective markers were attached to predefined anatomical landmarks on the whole body of the actor. Marker placement was based on the Vicon "plug-in-gait" model originally comprising 39 markers per person (see Figure 1(A)). In our recording sessions, we placed two additional markers on the left and right little toe of the participants. This was done so that we would be able to create different stimuli from the motion capture data (i.e. virtual reality avatars, point-light displays). The three-dimensional trajectories of the markers were tracked with a spatial accuracy of 1 mm at a sampling rate of 100 Hz. After capturing, the data was post processed using Nexus 1.8.5 (Vicon Motion Systems, Oxford, England). In this step, only the parts of a sequence successfully representing the emotion were retained. We then created video files of 4-s sequences from the original coordinate 3D (C3D) files using MATLAB software (MathWorks, Natick, MA). Within this process, only 15 markers per person were plotted as white spheres on a black background, which represent a standard PLD model (Troje et al., 2005).

1.3. Stimulus validation and selection

Prior to the experiment, we tested the recognisability of the scenes within a pilot study. We asked 18 participants who did not participate in the present experiment to evaluate the valence and the emotional category of the interaction. Valence was judged on an 11-point scale ranging from −5 (*extremely negative*) to + 5 (*extremely positive*). Emotion categories included anger, sadness, happiness, and affection. We applied two validation criteria for the stimuli: The first was that at least 50% of the participants should recognise the emotion displayed in the trial (e.g. anger). The second criterion ensured that the second most rated emotion (e.g. sadness) should not exceed a percentage of 25%. This allowed us to identify and exclude ambiguous scenes in which a specific emotion could not be recognised reliably. Finally, we randomly selected 12 scenes for each emotion that met both criteria, forming a set of 48 stimuli for this study. To ensure that our emotion categories did not differ significantly in terms of perceived intensity, we averaged valence ratings per stimuli across participants. Next, absolute values (i.e. intensities) were calculated and entered into a one-way ANOVA with emotional category as the independent variable. Results confirmed that the mean intensities did not differ significantly between categories, $F(3, 47) = .17$, $p = .92$, $\eta^2 = .01$. Lastly, we created two reduced forms of the PLD sequences for the experimental manipulation. The first reduced form comprised only arm markers – that is, shoulders, elbows, and wrists ("arms-only"). The second form of reduced PLDs included all markers except arm markers – that is, head, clavicular, hips, knees, and ankles

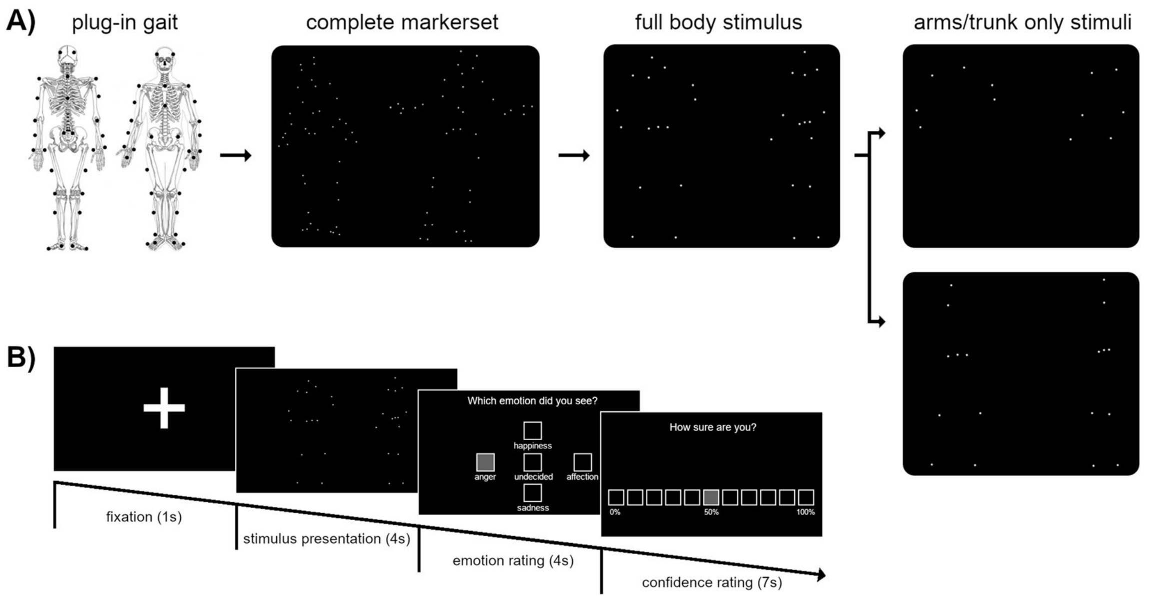

Figure 1. Stimulus creation and experimental timeline. (A) First, point-light displays using Vicon Motion Capture System were created with 41 markers attached to predefined anatomical landmarks (modified Vicon "plug-in-gait" model, adapted from Galna et al., 2014). Full-body stimuli consisted of 15 markers from which arms-only and trunk-only stimuli were derived. (B) Temporal structure of one trial.

("trunk-only"). Figure 1(A) depicts the different stimulus categories.

1.4. Experimental procedure

Prior to the experiment, participants were familiarised with the task and subsequently performed a test run. In the test run, they categorised four different emotional interactions. To do so, a sequence was presented and participants subsequently had to choose one out of four emotional categories (i.e. sadness, anger, happiness, affection) to classify the sequence. All scenes were depicted as full bodies, arms-only and trunk-only, resulting in a total of 4 × 3 test sequences. The actual experiment comprised a series of 144 sequences (4 emotions × 12 scenes × 3 types of display). Each sequence was presented twice, resulting in a total of 288 trials. Sequences were displayed in a pseudorandomised order to prevent the same trial from being presented consecutively.

Each trial started with a fixation phase (1s), followed by a stimulus sequence (4s) and then two ratings (7s each). The first rating assessed the emotion category and participants could choose from five options: anger, happiness, sadness, affection, and "undecided". After each category rating, a second rating assessed rating confidence on an 11-point scale ranging from 0 to 100% confidence (see Figure 1(B)). In total, the experimental session took about 45 min per participant.

1.5. Data analysis and statistics

1.5.1. Influence of emotional expressivity on emotion recognition

In the first step of our analyses, we calculated recognition rates of full body displays based on our validation data. In order to ensure a sufficient degree of recognisability of the stimuli, we tested each emotional category against chance (25%), using Bonferroni-corrected one-sample *t*-tests.

Next, a between-group factor was generated. The emotional expressivity subscale of the EKF was sorted by scores and split into three equally sized groups ($n = 10$): a low expressivity ($M = 97.5$, $SD = 3.7$), a medium expressivity ($M = 108.4$, $SD = 3.6$), and a high expressivity group ($M = 116.4$, $SD = 1.5$). We created three groups to ensure good differentiability, while keeping the sample size per group as large as possible. The groups differed significantly from each other with respect to their mean emotional expressivity scores, as revealed by a one-way ANOVA ($F(2, 29) = 45.21$, $p < .001$). Post hoc analyses indicated that the mean scores of each group were significantly different from each other (all $ps < .001$).

We then calculated a 3 (emotional expressivity group: low, medium, high expressivity) × 4 (emotion category: anger, sadness, happiness, affection) repeated measures ANOVA to explore whether emotional categories differ in terms of their recognisability and whether recognition performances are mediated by the group factor.

1.5.2. Influence of stimulus reduction on emotion recognition

In the last step, we explored the effect of stimulus reduction on recognition performances. To do so, the full body ratings of a participant were used as an individual baseline measure (in contrast to the previous analysis, in which validation data were used as a baseline). Having presented the full body stimuli twice, we were able to identify ambiguous interactions on an individual level, which allowed us to exclude them from further analysis (5.6% of all trials), and simultaneously ensured good intra-rater reliability of the full body baseline ratings. On the basis of this individual baseline, we calculated relative recognition accuracies by comparing the category ratings of the reduced stimuli (i.e. arms-only, trunk-only) to the respective full body rating.

Finally, we calculated two 4 (emotion category: anger, sadness, happiness, affection) × 3 (emotional expressivity groups: low, medium, high expressivity) × 2 (stimulus type: arms-only, trunk-only) repeated measures ANOVAs for the dependent variables recognition performance and rating confidence. All statistics were calculated using SPSS software (Version 25) and alpha was set at .05 for all statistical tests. All post hoc pairwise comparisons were Bonferroni-corrected.

Due to violations of the normal distribution, we applied a two-step relative rank transformation according to Templeton (2011) to normalise our data. All recognition rates are presented in percentages. The respective statistics rely on normalised data.

2. Results

2.1. Emotion recognition of full-body stimuli

Overall, emotion recognition accuracies were high (see Table 1). All four emotions were classified above chance level (anger: $t[29] = 31.44$, $p < .001$; happiness $t[29] = 27.94$, $p < .001$; affection: $t[29] = 19.35$, $p < .001$; sadness: $t([29] = 19.44$, $p < .001)$). On average, categorisation judgments showed an accuracy of 80.7%. Anger sequences were categorised with the highest accuracy, displaying a mean recognition performance of 87.8%. The second most accurately categorised emotion was happiness at 85.6%. This was followed by sadness with a mean recognition performance of 82.1%. Affection, however, showed a markedly lower recognition performance of 67.6%, yet still within a fully satisfying range (see Figure 2(A)).

2.2. Emotion recognition of full-body stimuli and the influence of emotional expressivity

To investigate the relationship between emotion recognition of full body stimuli and emotional expressivity, we conducted a 3 (emotional expressivity group) × 4 (emotion category) repeated measures ANOVA (see Figure 2(B)). Results revealed a main effect of emotional expressivity group, $F(2, 27) = 10.44$, $p < .001$, $\eta_p^2 = 0.44$. Post hoc analyses revealed a significantly lower recognition accuracy for the low expressivity group as compared to the medium ($p < .001$) or high expressivity group ($p < .05$). No significant interaction was found between expressivity group and emotion category ($F(6, 81) = 1.62$, $p = .15$, $\eta_p^2 = 0.11$).

2.3. Influence of stimulus reduction on recognition performance and confidence

In the next step, we examined how the reduction of kinematic information influenced recognition performance. We computed two separate 3 (emotional expressivity group) × 2 (stimulus type) × 4 (emotion category) repeated measures ANOVAs with emotional expressivity as a between-subject factor to determine the significance of both the emotion recognition performance as well as the confidence ratings.

For recognition performance, analyses revealed a significant main effect of stimulus type, $F(1, 27) = 50.06$, $p < .001$, $\eta_p^2 = 0.65$. Participant's performances were significantly higher when presented with arms-only stimuli ($M = 68.9\%$, $SEM = 1.7\%$) compared to trunk-only sequences ($M = 57.8\%$, $SEM = 1.4\%$). Interestingly, this effect was further differentiated by an interaction between stimulus type and emotion category, $F(3, 81) = 66.21$, $p < .001$, $\eta_p^2 = 0.71$ (see Figure 3(A)), indicating that the effect did not hold for all emotions equally. Whereas anger and happiness were recognised more easily from arms-only compared to trunk-only displays (both $ps < .001$), affection interactions did not lead to stimulus-

Table 1. Confusion matrix of emotion categories. Displays how often (in %) a specific target emotion was recognised as another emotion.

		recognised emotion [%]				
		anger	happiness	affection	sadness	undec.
target emotion	anger	87.8	3.3	0.4	3.6	4.7
	happiness	10.1	85.6	1.0	0	3.3
	affection	3.9	14.7	67.6	6.0	7.8
	sadness	4.2	2.5	0.8	82.1	10.4

Note: Percentage values are calculated across all sequences and all participants ($n = 2880$).

specific recognition performances. For sadness sequences, however, the effect was even reversed. Here, recognition rates were significantly higher for trunk-only compared to arms-only sequences ($p < .001$). Therefore, the main effect "stimulus type" originated from an advantage of arms-only displays in two emotion categories (i.e. anger and happiness), whereas trunk-only displays led to an advantage for only one emotion category (i.e. sadness).

With regard to confidence ratings, we also found a main effect of stimulus type on confidence ratings, $F(1, 27) = 21.68$, $p < .001$, $\eta_p^2 = 0.45$, indicating that participants were more confident when rating arms-only displays ($M = 75.2\%$, $SEM = 1.8\%$) compared to trunk-only sequences ($M = 70.2\%$, $SEM = 2.3\%$). Again, this effect was augmented by a significant interaction between stimulus type and emotion category, $F(3, 81) = 28.56$, $p < .001$, $\eta_p^2 = 0.51$. Whereas anger and happiness were judged with a higher confidence when observed as arms-only stimuli, the effect was reversed for sadness sequences in which trunk-only stimuli were judged more confidently (all ps $< .001$). Confidence for affection sequences did not reveal significant differences for the two stimulus types ($p = .14$; see Figure 3(B)). Therefore, the results for recognition scores and the associated confidence ratings seem to be congruent.

2.4. Influence of emotional expressivity on reduced stimulus recognition performance and confidence

Finally, we investigated whether the emotional expressivity groups were affected differently by the reduction of the stimulus material. Results revealed a main effect of expressivity group, $F(2, 27) = 9.28$, $p = .001$, $\eta_p^2 = 0.41$. Participants with low emotional expressivity displayed significantly lower recognition accuracies than participants with medium ($p < .01$) or high ($p < .05$) emotional expressivity (see Figure 3(C)).

Interestingly, there was a significant interaction between group and stimulus type, $F(2, 27) = 4.46$, $p < .05$, $\eta_p^2 = 0.25$. This interaction showed no significant difference between arms-only and trunk-only recognition for the low emotional expressivity group; however, medium and highly expressive participants differed significantly in their recognition performance depending on the type of stimulus presented. For these individuals, recognition performances were better for arms-only stimuli compared to trunk-only stimuli ($p < .001$ in both groups; see Figure 3(D)). These results indicate that medium and high expressivity group individuals showed higher recognition performances driven by higher recognition accuracies for arms-only stimuli.

Whereas these effects were found for recognition performance, confidence did not differ depending on one's own emotional expressivity, $F(2, 27) = 1.38$, $p = .27$, $\eta_p^2 = 0.09$.

3. Discussion

The present study demonstrates that participants are able to infer the emotional states from interactions depicted by PLDs. However, it seems that, depending on the emotion which is displayed, different body parts can carry the key information that allows for a decoding of the affective state. More precisely, the depicted emotions anger and happiness are recognised more reliably and more confidently from arm movements compared to movements of the rest of the body. Regarding interactions that depict sadness, we found the opposite effect: These PLD interactions are easier to recognise and are recognised more confidently from trunk rather than from arm movements. Interactions displaying affection, however, can be recognised equally well from both body and arm movements.

With respect to one's ability and willingness to express emotions, the present data reveal a relationship with recognition performance: Participants with an enriched emotional expressivity show higher recognition accuracy. More specifically, our data show that individuals with lower emotional expressivity

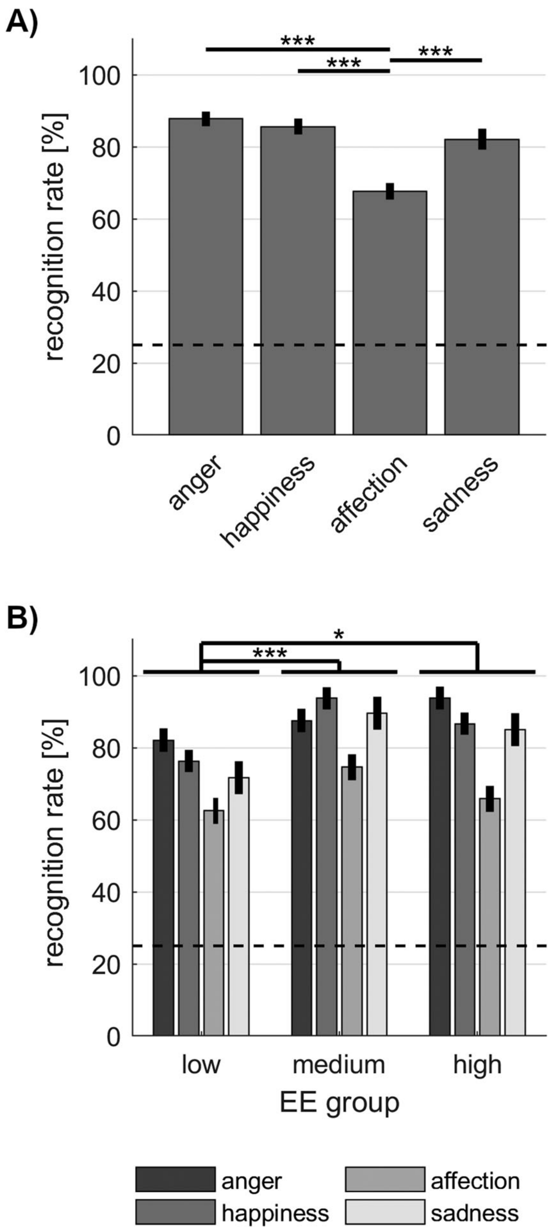

Figure 2. Emotion recognition rate of full-body stimuli for (A) all four emotion categories and (B) separated by emotional expressivity. Bars (and their standard errors) showing average emotion recognition rate in % as (A) a function of emotion and (B) additionally separated by group (low, medium, high emotional expressivity [EE]). Horizontal lines indicate significant differences between (A) emotions, respectively between (B) participant groups. Significance level is indicated by asterisks (*: $p < .05$; ***: $p < .001$). Dashed line represents threshold of chance recognition (25%).

display an equally decreased performance for arms-only as well as trunk-only displays. Individuals with higher emotional expressivity, however, benefit differentially from information on arm movements, as their recognition is enhanced especially when arm stimuli are presented.

Our results suggest the following main conclusions: (1) Humans are highly adept at recognising emotions from extremely impoverished stimulus information and (2) use emotion-dependent movement features to infer the emotional content of an interaction. (3) One's own ability and willingness to express emotions modulates emotion recognition. (4) Individuals with higher emotional expressivity seem to profit from information conveyed by gestures to infer observed emotions. In the following, we shall discuss the main findings in more detail.

3.1. Understanding emotions from body movement and body language

Emotion perception and recognition are an inherently multimodal phenomenon based on different cues in different modalities – namely gestural, postural, facial, or vocal (Atkinson et al., 2004; Pollick et al., 2001). There is a broad body of literature demonstrating that people have little difficulty in decoding emotionally relevant cues from others' bodily expressions, and that they regularly use these to infer emotional states (Atkinson et al., 2004; de Gelder, 2006). Because a number of critical features might contribute to recognising emotions from body language, one major claim of research in this field has been to define the relative impact of different motor parameters and movement features.

In principle, it can be stated that both, static properties of postures (Coulson, 2004; Tracy & Robins, 2004) and dynamic qualities of body movement enable emotion perception (Atkinson et al., 2004; Dittrich et al., 1996; Glowinski et al., 2011; Wallbott, 1998). For example, when Pollick et al. (2001) investigated emotion perception from arm movements, they found that perceived emotional activation (arousal) relates positively to more temporal movement parameters (such as velocity, acceleration, or jerk). Studies investigating affective gait patterns such as sad gait have come to a similar conclusion (Michalak et al., 2009): Sad walking is rather slow-paced with a clearly diminished arm swing, thereby underpinning also the importance of movement-related temporal parameters. Wallbott (1998) reported that produced movement features such as spatial extension or movement energy also differentiate significantly between expressed emotions. Regarding relevant body parts,

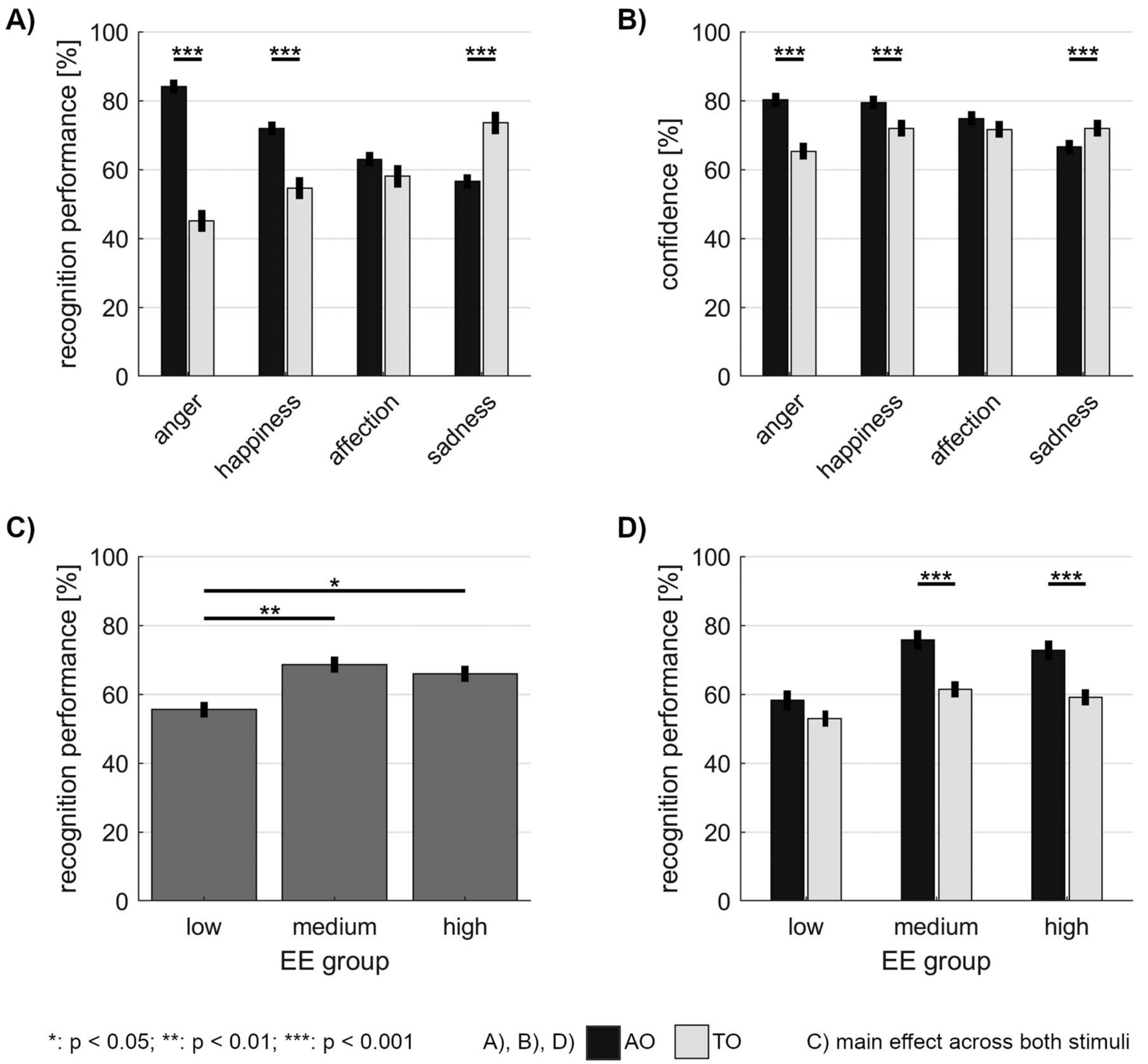

Figure 3. Stimuli type dependent categorisation performances and confidences. Emotion-specific recognition performance (A) and confidence (B) for reduced stimuli (arms-only [AO] and trunk-only [TO]) across all participants. Influence of emotional expressivity (low, medium, high EE group) shown as (C) main effect of EE group and (D) interaction with stimulus type on recognition performances. All bars (and their standard errors) displaying average categorisation performance and confidence rates in %. Horizontal lines indicate significant differences. Significance level is indicated by asterisks (*: $p < .05$; ***: $p < .001$).

Dael et al. (2012, 2013) showed that it is especially characteristics of arm and hand movements that allow a reliable emotional attribution. Hence, a complex interaction between timing and spatial parameters of different body parts and body form contributes to the percept of a specific emotional state.

In the present study, we used an occlusion paradigm to elucidate the salience of body and arm movements for recognising four different emotional states displayed as interactions: namely, anger, sadness, happiness, and affection. In line with the literature, the present data demonstrate that emotions can be conveyed via both, arm as well as trunk movements. However, we further showed that the movement features used to infer the emotional content of an interaction are emotion dependent. That is, anger and happiness are recognised more reliably from arm movements compared to trunk movements, but the opposite is the case for sad interactions.

In this regard, it has been argued that more active emotions (e.g. hot anger, elated joy, pride, irritation) are associated with more expansive and more forceful arm gesturing (Dael et al., 2013; de Meijer, 1989; Wallbott, 1998; Wallbott & Scherer, 1986). Thus, it is plausible that arm movements might be mandatory for the recognition process and can, therefore, be considered as one key information carrier for emotions such as happiness and anger. In particular, they seem to encode emotion-specific action tendencies of the observed agent, and their decoding helps the observer to adapt to actual environmental demands (e.g. information processing, preparation, and direction of action, communication of reaction, and behavioural intention).

In contrast, sadness is characterised by a reduction or even absence of every form of action, including arm movement, vertical movements of the head and lateral swaying movements as well as by a slumped posture (Dael et al., 2012; Michalak et al., 2009). Thus, in particular information stemming from a specific trunk posture seems to provide the most salient emotion-specific information, perhaps due to the relative lack of any emotion-specific effector movements.

In support of these findings, the investigation of the bodily topography of feelings revealed that different feelings were associated with distinct bodily fingerprints (Nummenmaa et al., 2018). While all emotions are *felt* within the head as well as within the trunk, anger and happiness are predominantly felt within the upper limbs (i.e. arms and hands). Sadness, in contrast, is felt within the trunk but not within the upper or lower limbs. The feeling of love seems to spread across the body. Thus, it seems plausible that the same body parts in which an emotion is *felt* are also used to *express* the emotion and will ultimately serve as information carrier during emotion perception.

3.2. How does individual emotional expressivity shape emotion recognition?

The subjective tendency to express emotions varies inter-individually (Kring & Gordon, 1998). In the present study, we demonstrated that this individual characteristic impacts on the recognition of emotional interactions: Individuals with higher subjective emotional expressivity are better at recognising emotions than individuals with lower subjective emotional expressivity. Our results further show that individuals with low emotional expressivity reveal no differences in their recognition performance for the two types of reduced stimuli presented here (i.e. arms-only vs. trunk-only PLDs). In contrast, individuals with medium or high emotional expressivity profited especially from displays presenting arm movements.

In this regard, several studies revealed that individual characteristics mediate the observer's perception of affective states in others (Bernaerts et al., 2016; Edey et al., 2017; Kaletsch et al., 2014a, 2014b; Van der Veen et al., 2007). Lorey et al. (2012), for instance, showed that participants with a reduced ability to identify and describe own feelings are significantly less confident about their judgment of emotional interactions depicted as PLDs. Moreover, a study by Edey et al. (2017) showed that an individual's own walking kinematics were directly related to the judgment they made about another person's affective state depicted by emotional PLDs.

A shared mapping of feelings and subjective states across individuals is also supported by a large body of empirical literature in the domain of action perception addressing the principle of action-perception loops (Decety & Sommerville, 2003; Prinz, 1997). One core assumption is that the perception of a given behaviour in another individual automatically activates one's own representations of that behaviour (Barsalou et al., 2003; Knoblich & Flach, 2001; Preston & De Waal, 2002). Hence, seeing someone else's emotional behaviour might be linked to a simulation of the way we express that emotion ourselves (Gallese, 2003; Singer & Lamm, 2009; Wicker et al., 2003).

However, based on our data we cannot rule out that the underlying process of emotion recognition can also be explained by a perceptual learning process and perceptual expertise. Grossman et al. (2004), for instance, showed that participants who were visually exposed to biological motion displayed as PLDs did not only show improvements in their recognition performance but also displayed increased BOLD-signal within areas associated with the recognition of motion and form. With respect to our findings, it may be possible that highly expressive individuals acquire visual instead of motor representations of their very own emotional body expressions, which then can be retrieved to understand another person's expressive behaviour.

4. Conclusions

We can draw two main conclusions from our findings: First, we demonstrated that emotions can be conveyed via arm as well as trunk movements. More precisely, we showed that the movement features used to infer the emotional content of an interaction are emotion-dependent. Whereas anger and happiness can be recognised more reliably from arm movements, sadness seems to be perceived mainly via cues from the head and torso. Thus, body parts that express the emotional state are also the most salient parts from which the observer can perceive it and, therefore, underpin the notion of an action perception loop across the agent and the observer (Hari & Kujala, 2009).

Second and more importantly, this recognition process seems to be linked to one's own ability to express emotions. We demonstrated that individuals with higher emotional expressivity show better recognition performances for emotional PLDs. Higher emotional expressivity is also associated with better recognition of emotional arm as compared to trunk movements, thereby showing a certain sensitivity of more expressive individuals for arm movements suggesting, again, a shared representation of action and perception in the framework of emotional body movements. Thus, emotion perception from body language seems to form a function of the expressing body and the individual observer.

5. Limitations and future implications

In order to assess emotional expressivity, we used a self-report measure. However, to our knowledge, there have not been any studies investigating the link between reported emotional expressivity and actual emotional expressivity, as measured by kinematic parameters. As self-report measures might be influenced by social desirability or low introspective ability, we cannot ascertain how closely the self-reported subjective impression of one's own emotional expressivity relates to the actual usage of body language to express emotions. Yet, despite of this methodological limitation we found a significant influence of this self-report measure on objective measures on emotion recognition.

However, a follow-up study should tackle this issue of subjectivity by specifying objective kinematic parameters (e.g. joint velocity, acceleration, jerk) that are affected by emotional states and investigate how these specific parameters relate to one's self-reported emotional expressivity.

Acknowledgements

The authors thank Jonathan Harrow for his helpful comments. We also thank Marie-Claire Englert for her support during data collection. Conceived and designed the experiments: JB, AZ, JM, BK. Performed the experiments and analysed the data: JB, AZ. Discussed the results and wrote the paper: JB, AZ, JM, BK.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the Deutsche Forschungsgemeinschaft – DFG IRTG 1901 "The Brain in Action" under a grant for JB.

Data availability statement

The data that support the findings of this study are available from the corresponding author JB upon reasonable request.

ORCID

Adam Zabicki http://orcid.org/0000-0002-0527-2705

References

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: An investigation of gender differences. *PloS One*, *6*(6), e20989. https://doi.org/10.1371/journal.pone.0020989

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746. https://doi.org/10.1068/p5096

Bachmann, J., Munzert, J., & Krüger, B. (2018). Neural underpinnings of the perception of emotional states derived from biological human motion: A review of neuroimaging research. *Frontiers in Psychology*, *9*, 176. https://doi.org/10.3389/fpsyg.2018.01763

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 43–92). Academic.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-II*. Psychological Corporation.

Bernaerts, S., Berra, E., Wenderoth, N., & Alaerts, K. (2016). Influence of oxytocin on emotion recognition from body language: A randomized placebo-controlled trial. *Psychoneuroendocrinology*, *72*, 182–189. https://doi.org/10.1016/j.psyneuen.2016.07.002

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, *28*(2), 117–139. https://doi.org/10.1023/B:JONB.0000023655.25550.be

Dael, N., Goudbeek, M., & Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception*, *42*(6), 642–657. https://doi.org/10.1068/p7364

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085–1101. https://doi.org/10.1037/a0025737

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends Cogn Sci.*, *7*(12), 527–533. https://doi.org/10.1016/j.tics.2003.10.004

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, *7*(3), 242. https://doi.org/10.1038/nrn1872

de Gelder, B., Terburg, D., Morgan, B., Hortensius, R., Stein, D. J., & van Honk, J. (2014). The role of human basolateral amygdala in ambiguous social threat perception. *Cortex*, *52*, 28–34. https://doi.org/10.1016/j.cortex.2013.12.010

de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, *13*(4), 247–268. https://doi.org/10.1007/BF00990296

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738. https://doi.org/10.1068/p250727

Edey, R., Yon, D., Cook, J., Dumontheil, I., & Press, C. (2017). Our own action kinematics predict the perceived affective states of others. *Journal of Experimental Psychology Human Perception and Performance*, *43*(7), 1263–1268. https://doi.org/10.1037/xhp0000423

Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society B*, *358*(1431), 517–528. https://doi.org/10.1098/rstb.2002.1234

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., & Rochester, L. (2014). Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait & Posture*, *39*(4), 1062–1068. https://doi.org/10.1016/j.gaitpost.2014.01.008

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., & Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, *2*(2), 106–118. https://doi.org/10.1109/T-AFFC.2011.7

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*(3), 193–213. https://doi.org/10.1016/j.cognition.2004.01.005

Grossman, E. D., Blake, R., & Kim, C. Y. (2004). Learning to see biological motion: Brain activity parallels behavior. *Journal of Cognitive Neuroscience*, *16*(9), 1669–1679. https://doi.org/10.1162/0898929042568569

Hadjikhani, N., & de Gelder, B. (2003). Seeing fearful body expressions activates the fusiform cortex and amygdala. *Current Biology*, *13*(24), 2201–2205. https://doi.org/10.1016/j.cub.2003.11.049

Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: From concepts to brain imaging. *Physiological Reviews*, *89*(2), 453–479. https://doi.org/10.1152/physrev.00041.2007

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, *135*(3), 278–283. https://doi.org/10.1016/j.actpsy.2010.07.012

Kaletsch, M., Krüger, B., Pilgramm, S., Stark, R., Lis, S., Gallhofer, B., Zentgraf, K., Munzert, J., & Sammer, G. (2014a). Borderline personality disorder is associated with lower confidence in perception of emotional body movements. *Frontiers in Psychology*, *5*(1262), https://doi.org/10.3389/fpsyg.2014.01262

Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., Stark, R., Lis, S., Gallhofer, B., Sammer, G., Zentgraf, K., Munzert, J., & Lorey, B. (2014b). Major depressive disorder alters perception of emotional body movements. *Frontiers Psychiatry*, *5*, 4. https://doi.org/10.3389/fpsyt.2014.00004

Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, *12*(6), 467–472. https://doi.org/10.1111/1467-9280.00387

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology*, *74*(3), 686–703. https://doi.org/10.1037/0022-3514.74.3.686

Lorey, B., Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., Stark, R., Zentgraf, K., & Munzert, J. (2012). Confidence in emotion perception in point-light displays varies with the ability to perceive own emotions. *PLoS One*, *7*(8), e42169. https://doi.org/10.1371/journal.pone.0042169

Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., & Schulte, D. (2009). Embodiment of sadness and depression – gait patterns associated with dysphoric mood. *Psychosomatic Medicine*, *71*(5), 580–587. https://doi.org/10.1097/PSY.0b013e3181a2515c

Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, *23*(2), 133–152. https://doi.org/10.1023/A:1021435526134

Neri, P. (2008). Wholes and subparts in visual processing of human agency. *Proceedings of the Royal Society*, *276*(1658), 861–869. https://doi.org/10.1098/rspb.2008.1363

Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, *96*(6), 1120–1136. https://doi.org/10.1037/a0015574

Nummenmaa, L., Hari, R., Hietanen, J. K., & Glerean, E. (2018). Maps of subjective feelings. *Proceedings of the National Academy of Sciences*, *115*(37), 9198–9203. https://doi.org/10.1073/pnas.1807390115

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*(2), B51–B61. https://doi.org/10.1016/S0010-0277(01)00147-0

Preston, S. D., & De Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–20. https://doi.org/10.1017/S0140525X02000018

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Rindermann, H. (2009). *Emotionale-Kompetenz-Fragebogen: EKF; Einschätzung emotionaler Kompetenzen und emotionaler Intelligenz aus Selbst-und Fremdsicht*. Hogrefe.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264–274. https://doi.org/10.1038/nrn2805

Schindler, K., Van Gool, L., & de Gelder, B. (2008). Recognizing emotions expressed by body pose: A biologically inspired neural model. *Neural Networks*, *21*(9), 1238–1246. https://doi.org/10.1016/j.neunet.2008.05.003

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologist Press.

Templeton, G. F. (2011). A two-step approach for transforming continuous variables to normal: Implications and recommendations for IS research. *Communications of the Association for Information Systems*, *28*(1), 41–58. https://doi.org/10.17705/1CAIS.02804

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science*, *15*(3), 194–197. https://doi.org/10.1111/j.0956-7976.2004.01503008.x

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics*, *67*(4), 667–675. https://doi.org/10.3758/BF03193523

Van der Veen, F. M., Evers, E. A., Deutz, N. E., & Schmitt, J. A. (2007). Effects of acute tryptophan depletion on mood and facial emotion perception related brain activation and performance in healthy women with and without a family history of depression. *Neuropsychopharmacology*, *32*(1), 216–224. https://doi.org/10.1038/sj.npp.1301212

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, *28*(6), 879–896. https://doi.org/10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, *51*(4), 690–699. https://doi.org/10.1037/0022-3514.51.4.690

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664. https://doi.org/10.1016/S0896-6273(03)00679-2

4 When the Timing is Right: The Link Between Temporal Coupling in Dyadic Interactions and Emotion Recognition

This chapter has been published as:

Bachmann, J., Krüger, B., Keck, J., Munzert, J., & Zabicki, A. (2022). When the timing is right: The link between temporal coupling in dyadic interactions and emotion recognition. *Cognition*, *229*, 105267. doi: 10.1016/j.cognition.2022.105267

Cognition 229 (2022) 105267

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

When the timing is right: The link between temporal coupling in dyadic interactions and emotion recognition

Julia Bachmann [a,*], Britta Krüger [a], Johannes Keck [a], Jörn Munzert [a,b], Adam Zabicki [a]

[a] Neuromotor Behavior Laboratory, Department of Psychology and Sport Science, Justus Liebig University Giessen, Giessen, Germany
[b] Center for Mind, Brain and Behavior (CMBB), Philipps University of Marburg and Justus Liebig University Giessen, Germany

ARTICLE INFO

Keywords:
Emotion recognition
Dyadic interaction
Temporal coupling
Point-light displays
Interpersonal predictive coding
Simulation theory

ABSTRACT

Affective states can be understood as dynamic interpersonal processes developing over time and space. When we observe emotional interactions performed by other individuals, our visual system anticipates how the action will unfold. Thus, it has been proposed that the process of emotion perception is not only a simulative but also a predictive process - a phenomenon described as interpersonal predictive coding. The present study investigated whether the recognition of emotions from dyadic interactions depends on a fixed spatiotemporal coupling of the agents. We used an emotion recognition task to manipulate the actions of two interacting point-light figures by implementing different temporal offsets that delayed the onset of one of the agent's actions (+0 ms, +500 ms, +1000 ms or + 2000 ms). Participants had to determine both the subjective valence and the emotion category (happiness, anger, sadness, affection) of the interaction. Results showed that temporal decoupling had a critical effect on both emotion recognition and the subjective impression of valence intensity: Both measures decreased with increasing temporal offset. However, these effects were dependent on which emotion was displayed. Whereas affection and anger sequences were impacted by the temporal manipulation, happiness and sadness were not. To further investigate these effects, we conducted post-hoc exploratory analyses of interpersonal movement parameters. Our findings complement and extend previous evidence by showing that the complex, noncoincidental coordination of actions within dyadic interactions results in a meaningful movement pattern and might serve as a fundamental factor in both detecting and understanding complex actions during human interaction.

1. Introduction

Social interactions are highly complex phenomena in which people rely on efficient mechanisms to derive information about each other. We are able to quickly retrieve a wealth of explicit and implicit information through the mere observation of a person. In particular, the human body presents a rich and reliable source of nonverbal information. Body motion tells us not only *what* a person is doing but also *why* they are doing it. It communicates both the mental state of a person and their action demands to the perceiver (for a review, see de Gelder, 2006).

Individuals can make a number of inferences about a person even if the available information is reduced to a minimum, as seen in point-light displays (PLD) in which shape information is removed while kinematic and configural information is preserved. It is not only possible to judge the gender or the identity of a person by simply observing these light points representing the kinematics of joints, we can also infer intentions and even affective states (Bachmann, Zabicki, Munzert, & Krüger, 2020; Hill & Pollick, 2000; Lorey et al., 2012; Mather & Murdoch, 1994; Runeson & Frykholm, 1983; Troje, 2002). It has been argued that affective states, often termed as *emotions*, are thought to inherently carry information about an organism's needs and goals and thus are closely linked to action tendencies. In his componential theory, Frijda (1986, 2007) described that action tendency or change in action readiness may represent a central component in emotional experience and thus closely ties to motor expression. In turn, the ability to retrieve relevant information allows us to generate adequate responses and integrate complex behavioral repertoires in order to achieve a certain goal (de Gelder, 2006; de Gelder & Hortensius, 2014; Frijda, 1986).

Interestingly, research shows that the recognition of affective states seems to be facilitated by observation of dyadic interactions rather than an individual's (monadic) emotional display (T. J. Clarke, Bradshaw, Field, Hampson, & Rose, 2005; Lorey et al., 2012). Clarke et al. (2005),

* Corresponding author at: Kugelberg 62, 35394 Giessen, Germany.
E-mail address: julia.bachmann@sport.uni-giessen.de (J. Bachmann).

https://doi.org/10.1016/j.cognition.2022.105267
Received 7 October 2021; Received in revised form 5 August 2022; Accepted 25 August 2022
Available online 1 September 2022

for instance, asked participants to identify the emotion category from interactions displayed as point-light displays. These were either displayed as single actors (monads) or as dyads. They found that the recognition of joy and romantic love was impaired by the absence of the interacting partner. Regarding the expression of emotions, it was demonstrated that the expression of joy is *heightened* in the presence of others (Fridlund, 1992). Both aspects highlight the fact that affective states are not merely intrapersonal but dynamic interpersonal processes that develop over time and space (T. J. Clarke et al., 2005; Lorey et al., 2012).

It has been proposed this ability to assign intentions or affective states from mere observation can be explained by *simulation* theory (Gallese & Goldman, 1998). Although simulation theory was thought to explain action understanding, it was soon extended and applied to intention and emotion understanding as well (Gallese, 2009; Rizzolatti & Craighero, 2004). Within this framework, perceiving bodily expressions of emotion in others is thought to rely on one's own (motor) representations of these emotional movements and states (Gallese, 2009; Gallese, Keysers, & Rizzolatti, 2004). More specifically, we recognize another individual's emotional state by internally simulating the same emotional state in ourselves (Wicker et al., 2003).

However, it has also been proposed that action perception is not only a simulation of what is happening but also a *prediction* of what is going to happen (Kilner, Friston, & Frith, 2007). This view is supported by behavioral evidence. In a seminal study, Neri, Luu, and Levi (2006) demonstrated that when interactive activity requiring close body contact is observed (e.g., fighting or dancing), the human visual system relies on the spatiotemporal coupling between the agents to retrieve information relating to each agent individually. More specifically, within meaningfully synchronized interactions, the actions of one agent serve as the predictor for the expected actions of the other agent. This is particularly true if actions of both agents are tightly coupled as in the context of fighting and dancing.

With respect to social interactions that do not require close body contact, Manera, Becchio, Schouten, Bara, and Verfaillie (2011); Manera, Schouten, Verfaillie, and Becchio (2013) and von der Lühe et al. (2016) demonstrated that observing dynamic communicative gestures of one agent enhances the visual discrimination of a second agent who is responding to this communicative gesture—a phenomenon that has been referred to as *interpersonal predictive coding*. Their results suggest that we possess implicit knowledge about the natural dynamics of human interaction that guides the processing of motion patterns generated by each of the two agents. Thus, when we observe actions performed by other individuals, we seem to anticipate how the action will unfold (von der Lühe et al., 2016). When we are engaged in a direct social interaction with a partner, prediction of the other person's actions helps us adjust our movements "online" (i.e., in real time) in order to plan an appropriate response (Becchio, Sartori, & Castiello, 2010). However, up to now, little is known about how such a spatiotemporal relatedness of the actions of two agents is important for recognizing different emotions in social interactions. Here, we explicitly address two questions: (a) Does the recognition of emotions from dyadic interactions depend on a fixed spatiotemporal coupling of the actions of the individuals? (b) Is this coupling emotion-specific?

To address these questions, we employed an emotion recognition task in which we manipulated the actions of two interacting agents by implementing different temporal offsets. More specifically, we delayed the onset of one of the agent's actions by either +0 ms, +500 ms, +1000 ms, or + 2000 ms. Participants were asked to determine the subjective valence—that is, how positive or negative a stimulus is perceived to be—as well as the emotion category (happiness, anger, sadness, or affection) displayed within a given scene. In light of experimental evidence demonstrating that spatiotemporally congruent actions are critical for the perception of social interactions, we anticipate that recognition will be impaired as a result of the temporal offset (Manera et al., 2011; Neri et al., 2006; Petrini, Piwek, Crabbe, Pollick, & Garrod, 2014; Thurman & Lu, 2014). Furthermore, we expect that because some emotions have a stronger interpersonal character (e.g., affection is characterized by stronger bodily interaction than sadness), the recognition of certain emotions will be affected more strongly by the temporal decoupling (T. J. Clarke et al., 2005).

Finally, we aimed to explore whether objective movement parameters are influenced by the temporal offset, and how these potential changes correspond to the subjective measure of emotion perception. Therefore, we quantified objective interpersonal movement features of the observed scenes and related these to recognition performance.

2. Methods

2.1. Participants

Forty-four adults (19 females and 25 males, mean age = 24.68 years, *SD* = 6 years) with normal or corrected-to-normal vision participated in the study. None reported any history of psychiatric or neurological disorders, and they had no history or current use of psychoactive medication. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the local ethics committee. The sample size was determined according to previous studies using a similar research paradigms (see for instance, Manera et al., 2011, 2013; von der Lühe et al., 2016; Bachmann et al., 2020). We used the German version of the Emotional Competence Questionnaire (Emotionale-Kompetenz- Fragebogen, EKF; Rindermann, 2009) to assess the participant's score in four abilities. To control for the participant's emotional state, we applied the State–Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) and the Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 1996). STAI-X1 scores ranged from 21 to 56 (M = 33.91, SD = 7.43); STAI-X2, from 20 to 70 (M = 38.39, SD = 11.10), with greater scores indicating greater anxiety. BDI-II scores ranged from 0 to 48 (M = 8.40, SD = 8.30). Although some of these participants showed clinically relevant scores (i.e., a BDI-II total score above 13), this did not affect their emotion recognition performance (see Appendix A, Fig. A.1).

2.2. Stimuli

Point-light displays of emotional interactions were taken from a motion-capture data set created by Bachmann et al. (2020). In brief (*for full details see* Bachmann et al., 2020), eight pairs of nonprofessional actors were asked to portray the following four emotional states within an interaction: happiness, sadness, anger, and affection. Each emotional state was acted out in varying intensities in order to obtain a variety of emotional expressions. Actors were asked to perform intuitively within the context of the given situation, thereby allowing freedom and subsequently variability in their expression. Moreover, both actors were instructed specifically to act out the same emotion in order to produce a congruent behavioral pattern. Interactions were recorded with an optical motion capture system (Vicon Motion Systems, Oxford, England). Next, video files of 4-s sequences were created from the original coordinate 3D (C3D) files using MATLAB software (MathWorks, Natick, MA). The final stimuli consisted of two point-light figures, each consisting of 15 markers indicating the major joints of each actor (head, clavicle, shoulders, elbows, wrists, hips, knee, and ankles). In the next step, recognizability was tested within a pilot study. Eighteen independent participants who were naive to the point-light sequences were asked to evaluate the valence and the emotional category of the interactions. We then excluded ambiguous scenes from the dataset.

From this dataset, we randomly selected 68 point-light interactions (i.e., 17 scenes per emotion category: affection, anger, happiness, sadness) to create the final stimulus set. Using MATLAB software (MathWorks, Natick, MA), four different temporal conditions were assembled per sequence. In the *no offset* condition (+0 ms), the actions of the actors were displayed in their original form. All other conditions contained temporal offsets between the actions of the agents. Within these conditions, the onset of the actions of one randomly selected agent was delayed by either 500 ms (+500 ms), 1000 ms (+1000 ms), or 2000 ms (+2000 ms). Finally, all sequences were cut down from 4 to 2 s in order to obtain overlapping actions (see Fig. 1A).

2.3. Apparatus and procedure

Stimuli were displayed on a 17.3-in. full HD screen (display resolution: 1920 × 1080; refresh rate 60 Hz) using MATLAB (MathWorks, Natick, MA) software. Participants were seated at a viewing distance of approximately 60 cm from the screen. The dots of the point-light sequences were light grey against a black background.

Prior to the experiment, participants were familiarized with the task and subsequently performed a test run. In the test run, they categorized four different emotional interactions. They were presented with eight point-light interactions and subsequently had to judge (a) the valence of the scene and (b) the emotional category that it displayed (i.e., sadness, anger, happiness, or affection). All scenes were depicted in their original temporal order—that is, with no temporal manipulation.

The actual experiment comprised a series of 272 sequences (4 emotion categories × 17 point-light sequences × 4 temporal offsets). Each sequence was presented once, except for the *no-offset* condition (+0 ms) that was presented twice in order to obtain an individual baseline for each participant. This resulted in a total of 340 trials. Sequences were displayed in a pseudorandomized order to prevent the same trial from being presented twice consecutively.

A forced-choice paradigm was employed: Each trial started with a fixation phase (1 s), followed by a stimulus sequence (2 s) and two ratings (see Fig. 1B). In the first rating, participants were asked to judge the subjectively perceived valence of the interaction on an 11-point scale ranging from −5 (*extremely negative*) to 5 (*extremely positive*) with 0 (*neutral*) marking the center of the scale. Next, participants were asked to assign an emotional category to the observed sequences (i.e., affection, anger, happiness, or sadness). They were instructed to make intuitive judgments; however, they were not restricted in their rating time. In total, the experimental session took about 45 min per participant.

2.4. Data analysis and statistics

2.4.1. Recognizability of emotional interactions

In the first step of our analyses, we calculated a confusion matrix by comparing the emotion category of baseline sequences (i.e., first presentation of each +0 ms interaction) with the category ratings from the validation study. To ensure a sufficient degree of recognizability for the stimuli, we tested the recognition rates of each emotional category against chance (25%) using a Wilcoxon signed-ranks test. Following Rosenthal (1994), we calculated effect sizes r as the Z statistic divided by the square root of the sample size N, with N being the total number of participants.

2.4.2. Influence of temporal decoupling on emotion recognition accuracy

In the next step, we explored the effect of different temporal offsets on recognition performance. We used the emotion category rating of the first presentation of each +0 ms sequence as an individual baseline measure (in contrast to the previous analysis, in which validation data were used as a baseline). This was done in order to prevent instances in which an individual's + 0 ms rating is incorrect (not matching the

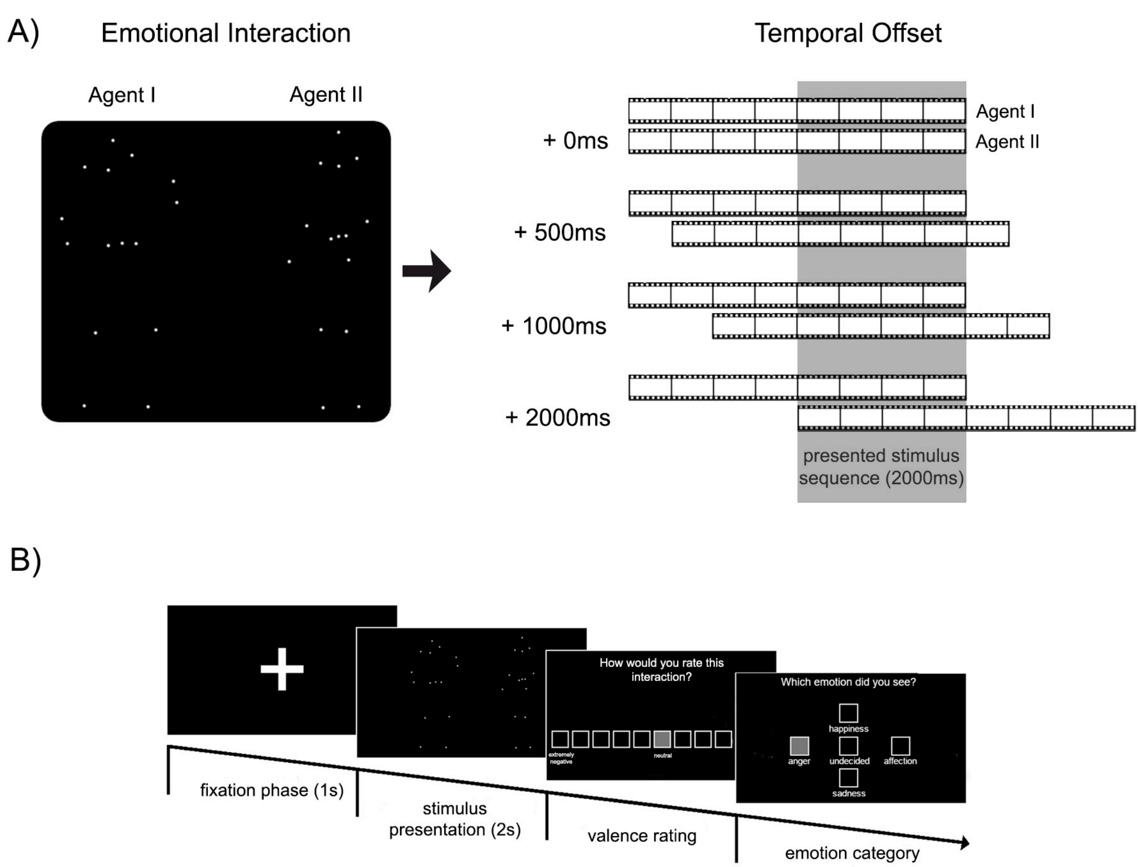

Fig. 1. *Stimulus creation and experimental timeline.* A) Point-light displays of emotional interactions were taken from a motion-capture data set created by Bachmann et al. (2020). Next, temporal offsets (+0 ms, +500 ms, +1000 ms, +2000 ms) were implemented, and the duration of stimulus sequences was reduced to 2000 ms in total. B) Temporal structure of one trial. Each trial started with a fixation phase (1 s) followed by the stimulus sequence (2 s). Subsequent to the stimulus presentation, a valence as well as emotion category rating was employed with no maximum response time.

validation) and the temporal offset leads to the individual identifying the "correct" (matching the validation) emotion, as this would falsely indicate that the individual's performance *improved* with increasing temporal offset although he or she did not recognize the respective (+0 ms) emotion and thus decreased in recognition performance. On the basis of this individual baseline, we calculated relative recognition accuracies by comparing the category ratings of the remaining sequences containing different temporal offsets (i.e., +0 ms, +500 ms, +1000 ms, +2000 ms), to the respective baseline rating.

Due to a violation of the normal distribution assumption, we applied a two-step relative rank transformation according to Templeton (2011) to normalize our data. Finally, we calculated a 4 (emotion category: affection, anger, happiness, sadness) × 4 (temporal offsets: +0 ms, +500 ms, +1000 ms, +2000 ms) repeated measures ANOVA for the dependent variable recognition performance.

In order to investigate whether changes in recognition performances are also accompanied by changes in sensitivity and response biases, we calculated sensitivity (d') and criterion (c) values for each emotion category and temporal offset. Sensitivity is a measure of the individual's ability to discriminate whether the 'signal' (here, the respective emotional category) is presented or not. Higher values of d' indicate better discrimination ability, i.e. the emotional categories can be distinguished from each other. Criterion values give information about the proportion of correctly identified as well as falsely identified sequences, allowing for an identification of response biases, i.e. if one or more category is chosen disproportionately more frequently than others. Thus, for each of the sixteen conditions, we calculated the proportion of hits (when the displayed emotion category was consistent with that identified) and false alarms (when an emotion category was falsely identified). This was done for each participant. Proportions of 0 were replaced with $0.5/N_{p/a}$, and proportions of 1 were replaced with $(N_{p/a} - 0.5)/N_{p/a}$ (in which N_p is the number of baseline sequences that were identified as a specific emotion category by a participant, and N_a the number of the remaining baseline sequences not identified as the respective category: i.e., $N_a = 68 - N_p$). Finally, we calculated two separate 4 (emotion category: affection, anger, happiness, sadness) × 4 (temporal offsets: +0 ms, +500 ms, +1000 ms, +2000 ms) repeated measures ANOVA using d' and c values.

2.4.3. Influence of temporal decoupling on valence judgment

Next, we explored the effect of different temporal offsets on subjective valence perception by calculating a 4 (emotion category: affection, anger, happiness, sadness) × 4 (temporal offsets: +0 ms, +500 ms, +1000 ms, +2000 ms) repeated measures ANOVA for the dependent variable valence intensity. Here, we used absolute values to be able to make inference about how intense a stimulus was perceived to be. Thus, higher valence intensity can therefore be interpreted as more negative or more positive, depending on the category. It is noteworthy that, on average, all anger and sadness sequences were identified as negative while all happiness and affection sequences were identified as positive with respect to the validated data (see Appendix A, Fig. A.2).

2.4.4. Movement features and emotion recognition

In a last step, we performed exploratory analyses to elucidate how specific quantitative movement features are potentially linked to the recognition of the emotional sequences. First, we used the SAMI toolbox (Similarity Analysis of Human Movements and Interactions, see Zabicki & Keck, 2021) to calculate relevant movement features of the interaction sequences. This enabled us to identify changes in interpersonal movement parameters within the stimuli caused by the introduced temporal offset. Specifically, we determined 12 parameters describing the interaction of the observed two agents (see Table 1). These features included, for instance, *interpersonal distance* (i.e., the average distance between the two agents), *personal space* (i.e., the time each agent spent in the personal space of the interactive partner), and so-called *motion energy balance* (provides information about how much of the total amount of body movement over time can be ascribed to each interactive partner). A perfectly balanced motion energy (equals a maximum value of 1) would be characterized by two agents who each take up the same amount of the total motion energy presented in the sequence.

Table 1
Summary of interaction-specific kinematic features calculated by the SAMI Toolbox*.

Kinematic interaction feature	Abbr.	Short description
Average of Interpersonal Distance	IPDavg	Spatial distance between both agents. Calculated as the mean over time, respectively the standard deviation.
Variance in Interpersonal Distance	IPDvar	
Average of Interpersonal Orientation	IPOavg	Time spent facing each other + time spent by one agent facing the other and vice versa.
Balance of Interpersonal Orientation	IPObal	absolute value of the difference between orientation times of each agent divided by the sum of orientation times; higher value indicating greater balance level
Correlation between Spatial Distance and Velocity	CorrDistVel	Absolute correlation between spatial distance between the agents and mean velocity of whole-body movements
Correlation between Spatial Distance and Acceleration	CorrDistAcc	Absolute correlation between spatial distance between the agents and mean acceleration of whole-body movements
Correlation between Spatial Distance and Limb Contraction	CorrDistLC	Absolute correlation between spatial distance between the agents and limb contraction
Correlation between Spatial Distance and Limb Volume	CorrDistLV	Absolute correlation between spatial distance and the space taken up by the three-dimensional total body extension
Synchronization of Velocity	SyncVel	Absolute correlation between the agent's velocity profiles
Synchronization of Acceleration	SyncAcc	Absolute correlation between the agent's acceleration profiles
Motion Energy Balance	MEB	Total amount of body movement over time that can be ascribed to each interactive partner Sum of averaged inter-frame Euclidean displacements of each marker of the agents, divided by the total amount of body movement in the scene; higher value indicating greater balance level
Personal Space	PS	Time spent in the personal space of the interactive partner

* For a detailed description of each kinematic feature and how they are calculated, see Zabicki and Keck (2021).

Each feature was calculated for each interaction sequence (68 in total) and for each temporal offset (+0 ms, +500 ms, +1000 ms, +2000 ms). Next, we conducted a one-way repeated measures ANOVA with temporal offset as the within factor to find out whether the kinematic parameters were altered significantly by the temporal offset. We did this only for those emotion categories that showed diminished recognition accuracies caused by the temporal offset, because our aim was to elucidate whether decreased recognition accuracies could potentially be explained by changes of movement features due to the implemented temporal offset.

In case a kinematic parameter would prove to be modulated in an analogous way as the corresponding subjective recognition accuracy, we conducted further analyses to see whether the respective kinematic parameters of a scene differ with respect to the recognition accuracy. We operationalized this by sorting movement parameters on the basis of whether the sequence was correctly assigned to its emotional category (based on the individual baseline rating). This was done for each trial per participant, again only for the emotion categories of interest. In a next step, we averaged movement features of both categories (correctly vs. incorrectly identified) across trials for each participant. Finally, we conducted a non-parametric Wilcoxon signed-rank test to test for a significant difference between the respective categories. Effect size was calculated by means of a matched-pairs rank biserial correlation.

2.4.5. General procedure

In case of violations of the normal distribution, we applied a two-step relative rank transformation according to Templeton (2011) to normalize our data when applying parametric tests. All recognition rates are presented in percentages. The respective statistics rely on normalized data. We report Huyn-Feldt- or Greenhouse-Geisser-corrected p values in order to counteract observed violations of sphericity. Moreover, all post hoc pairwise comparisons were Bonferroni-corrected.

3. Results

3.1. Recognizability of emotional interactions

Overall, the emotion category ratings of the baseline interactions were consistent with the categories identified within the validation study, as displayed by the confusion matrix in Fig. 2A. In the next step, we tested the recognition rates of each emotional category against chance (25%) in order to ensure a sufficient degree of recognizibility. The Wilcoxon signed-ranks test revealed that sequences of each emotional category were classified significantly above chance level (affection: $Z = 5.78$, $p < .001$, $r = 0.87$; anger: $Z = 5.81$, $p < .001$, $r = 0.88$; happiness $Z = 5.79$, $p < .001$, $r = 0.87$; sadness: $Z = 5.79$, $p < .001$, $r = 0.87$; see Fig. 2B). Anger sequences were categorized with the highest accuracy, displaying a mean recognition performance of 88%. The second most accurately categorized emotion was happiness at 76%. This was followed by sadness with a mean recognition performance of 68%. Affection, however, showed a markedly lower recognition performance of 58%, yet still within a fully satisfactory range, that is, *above* the 25% *chance level (see* Fig. 2B*)*.

3.2. Temporal decoupling and emotion recognition

In the following, the respective statistics rely on normalized data, and all descriptive data are presented in actual values. Using a two-way repeated measures ANOVA, we investigated whether the recognition of emotional interactions was affected by temporal offsets. Results revealed a main effect of temporal offset, $F(3, 129) = 30.531$, $p < .001$, $\eta_p^2 = 0.42$. Bonferroni-corrected post hoc analyses showed that recognition rates for the +0 ms condition were significantly higher than for all other conditions (all $p < .01$). Recognition rates decreased linearly from the +0 ms to the +2000 ms interaction sequences. However, there was no significant difference between +1000 ms and + 2000 ms (see Fig. 3A). Moreover, we found a main effect of emotion category, $F(2.75, 118.12) = 7.75$, $p < .001$, $\eta_p^2 = 0.15$. Post hoc analyses indicated that affection scenes had significantly lower recognition rates than any of the other emotions (all $p < .001$). Recognition rates for anger, happiness, and sadness scenes did not differ from each other (all $p > .05$).

There was a significant interaction between temporal offset and emotion category, $F(9, 387) = 3.79$, $p < .001$, $\eta_p^2 = 0.04$. Considering each emotion category separately, we found that only the recognition of affection and anger sequences was affected by the temporal offset. More specifically, results showed that for affection scenes, the recognition rates between +0 ms and + 500 ms did not differ significantly ($p = .08$). However, recognition rates for +0 ms sequences were significantly higher compared to +1000 ms and + 2000 ms interactions (all $p < .001$). With respect to anger scenes, recognition rates decreased linearly from the +0 ms to the +2000 ms interaction sequences. Post hoc analyses showed no differences between the temporal offset conditions for happiness and sadness interactions (all $p > .05$) (see Fig. 3B).

To further explore the interaction effects of temporal offset and emotion category on recognition performances, we conducted a sensitivity (d') and criterion (c) analysis. Thus, we investigated how introducing a temporal offset relates to the participants' discrimination abilities and response biases. For d', we found a main effect of temporal offset, $F(3, 129) = 37.21$; $p < .001$, $\eta_p^2 = 0.46$, decreasing linearly from the +0 ms to the +2000 ms temporal offset condition (all $p < .05$) (see Fig. 4A). Moreover, we found a main effect of emotion category, $F(2.75, 118.12) = 7.75$, $p < .001$, $\eta_p^2 = 0.15$, indicating significantly lower values for happiness sequences compared to anger ($M_{diff} = 0.21$, $p < .01$, 95% CI [0.04, 0.39]) and sadness ($M_{diff} = 0.28$, $p < .001$, 95% CI [0.11, 0.46]) sequences. Finally, we found a significant interaction between temporal offset and emotion category, $F(8.19, 352.01) = 2.12$, $p < .05$, $\eta_p^2 = 0.05$. Post hoc tests showed that for affection sequences, d' values did not differ significantly between +0 ms and + 500 ms. However, the +0 ms differed significantly from the +1000 ms and + 2000 ms sequences (all $p < .05$). For anger scenes, we found that d' values were significantly higher for the +0 ms sequences than for all other sequences (all $p < .001$). Moreover, d' values for the +500 ms as well as the +1000 ms sequences differed significantly from d' values for the +2000 ms sequences (all $p < .001$). For happiness sequences, post hoc analysis indicated that d' values were significantly higher for the +0 ms sequences than for all other sequences (all $p < .05$). Moreover, d' values for the +500 ms sequences were significantly higher as compared to +1000

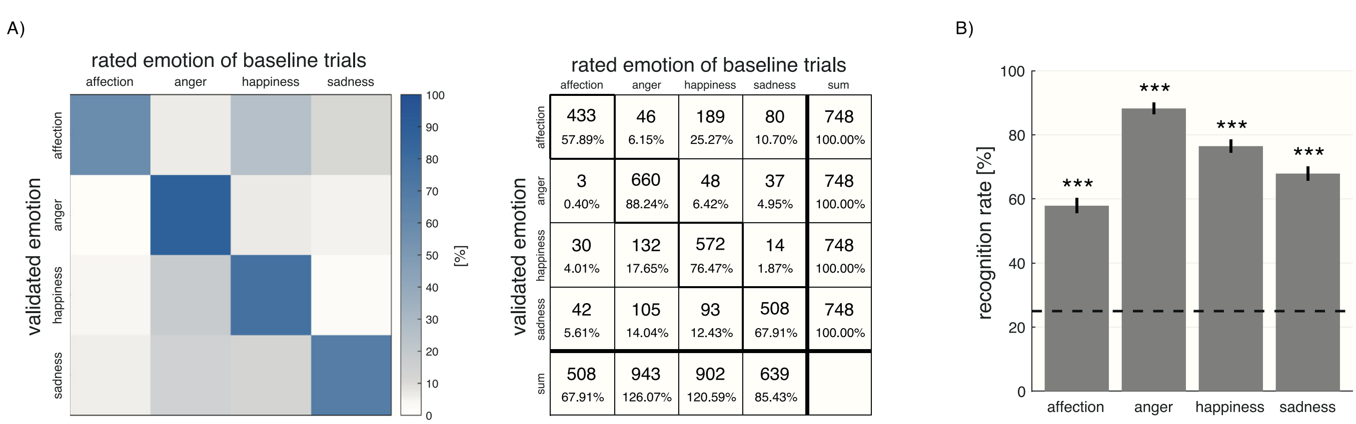

Fig. 2. *Confusion matrix and recognition rates.* A) Confusion matrix displaying number and percentage of participants' baseline emotion category ratings compared with the validated emotion category rating. B) Emotion recognition rates of baseline interactions for all emotion categories (i.e., affection, anger, happiness, sadness). Bars (and their standard errors) show average emotion recognition rate in %. Significance level is indicated by asterisks (***$p < .001$). Dashed line represents threshold of chance recognition (25%).

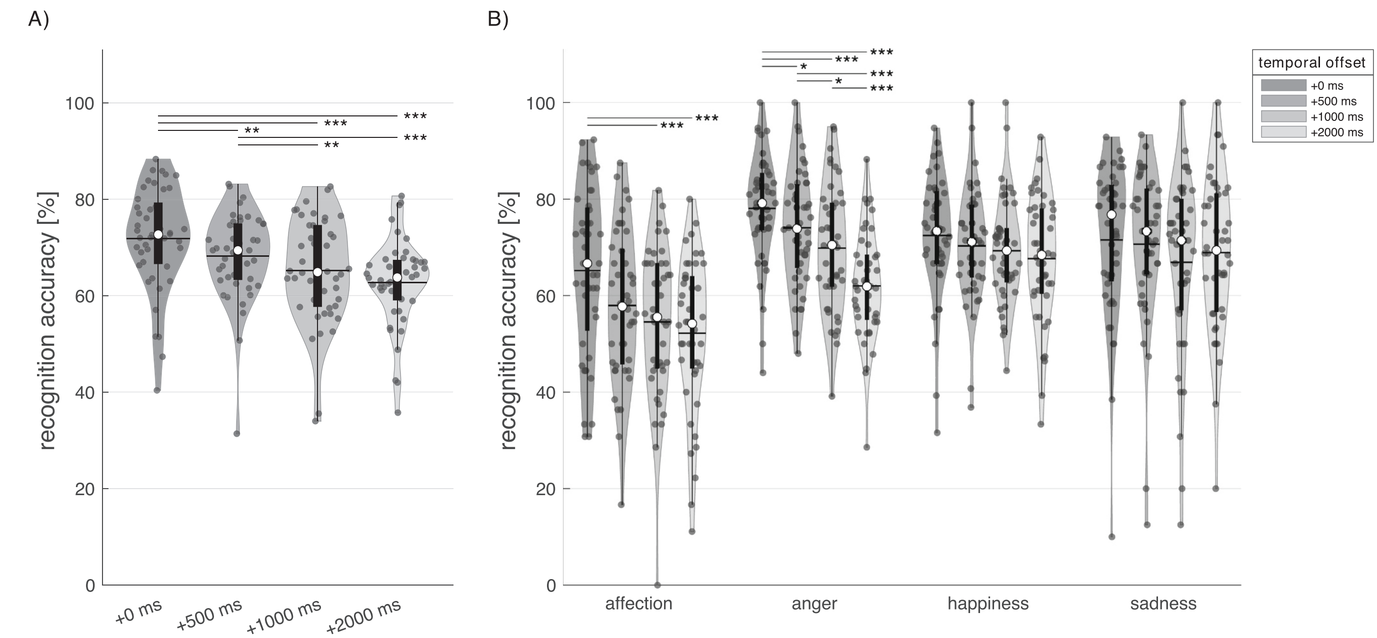

Fig. 3. *Influence of temporal offset on recognition performance.* Emotion recognition accuracies (in %) for A) all four temporal offsets (+0 ms, +500 ms, +1000 ms, +2000 ms) and then B) additionally separated by emotion category (affection, anger, happiness, sadness). Violin plots display the distribution of individual performances (single, colored dots). Each violin plot contains information about the median (i.e., white dot), the 25th and 75th percentiles (i.e., bold vertical line), and the mean (i.e., horizontal line). Horizontal lines between variables indicate significance of difference between temporal offsets. Significance level is indicated by asterisks (*p < .05; **p < .01; ***p < .001).

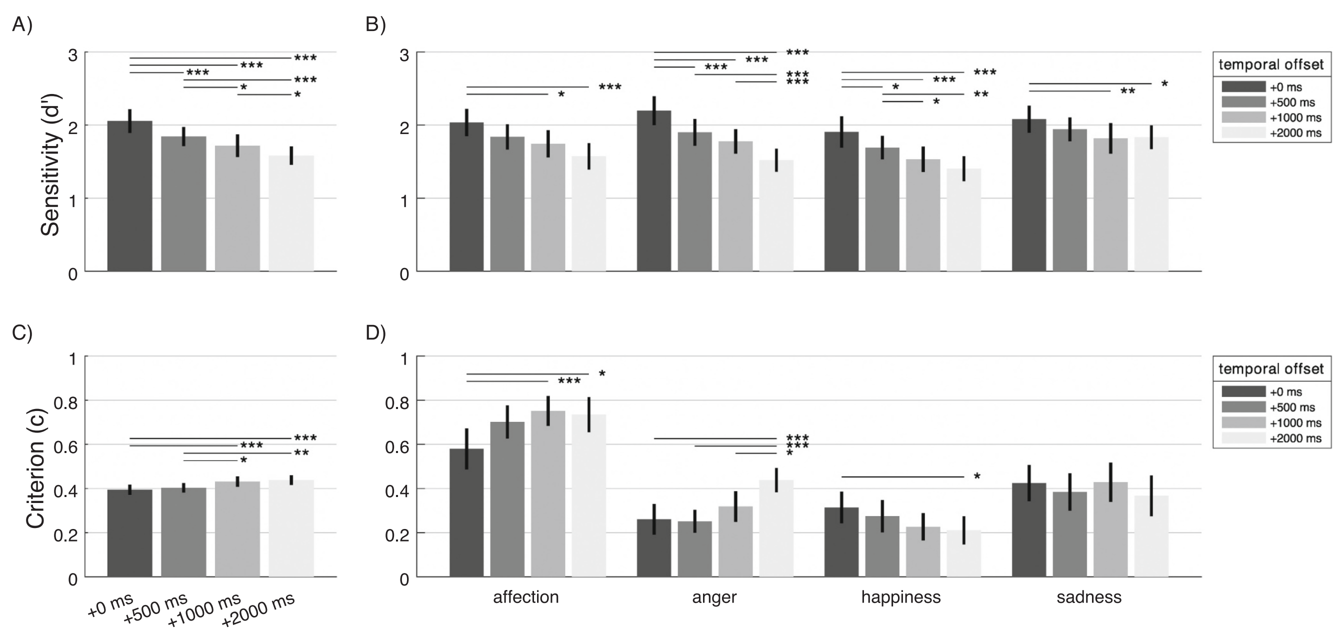

Fig. 4. *Influence of temporal offset on sensitivity (d') and response bias (c)* A) + C) for all four temporal offsets (+0 ms, +500 ms, +1000 ms, +2000 ms) and B) + D) additionally separated by emotion category (affection, anger, happiness, sadness). Bars (and their standard errors) show average d' as well as c values. Horizontal lines indicate significant differences between temporal offsets. Significance level is indicated by asterisks (*p < .05; **p < .01; ***p < .001).

ms and + 2000 ms sequences (all $p < .05$). Finally, post hoc tests indicated that for sadness trials, d' values were significantly higher for +0 ms sequences as compared to +1000 ms and + 2000 ms sequences (all $p < .05$) but not for +500 ms sequences ($p = .62$) (see Fig. 4B).

With regard to criterion values, we found a main effect of temporal offset ($F(3, 129) = 10.99$, $p < .001$, $\eta_p^2 = 0.20$), increasing linearly from the +500 ms to the +2000 ms temporal offset condition (all $p < .05$) (see Fig. 4C). Moreover, we found a main effect of emotion category, $F(3, 129) = 41.79$, $p < .001$, $\eta_p^2 = 0.49$, indicating significantly higher values for affection sequences compared to all other categories (anger: $M_{diff} = 0.38$, $p < .001$, 95% CI [0.28, 0.48]; happiness: $M_{diff} = 0.44$, $p < .001$, 95% CI [0.31, 0.57]; sadness: $M_{diff} = 0.29$, $p < .001$, 95% CI [0.18, 0.39]) as well as significantly higher values for sadness sequences as compared to happiness ($M_{diff} = 0.15$, $p < .05$, 95% CI [0.01, 0.29]).

A significant interaction was found between temporal offset and emotion category ($F(6.72, 288.75) = 5.16$, $p < .001$, $\eta_p^2 = 0.11$). Post hoc tests showed that for affection, criterion values indicate that +0 ms differed significantly from the +1000 ms and + 2000 ms sequences (all $p < .05$), indicating that both, the false alarm rate and the hit rate, *decrease* with increasing temporal offset. For anger, we found that values were significantly higher for +2000 ms sequences as compared to all others (all $p < .05$). Finally¸we found that for happiness, criterion values for +0 ms sequences were significantly higher as compared to +2000 ms. In contrast to affection scenes, the decrease of the criterion values for happiness indicates that both the hit as well as the false alarm rate *increase*.

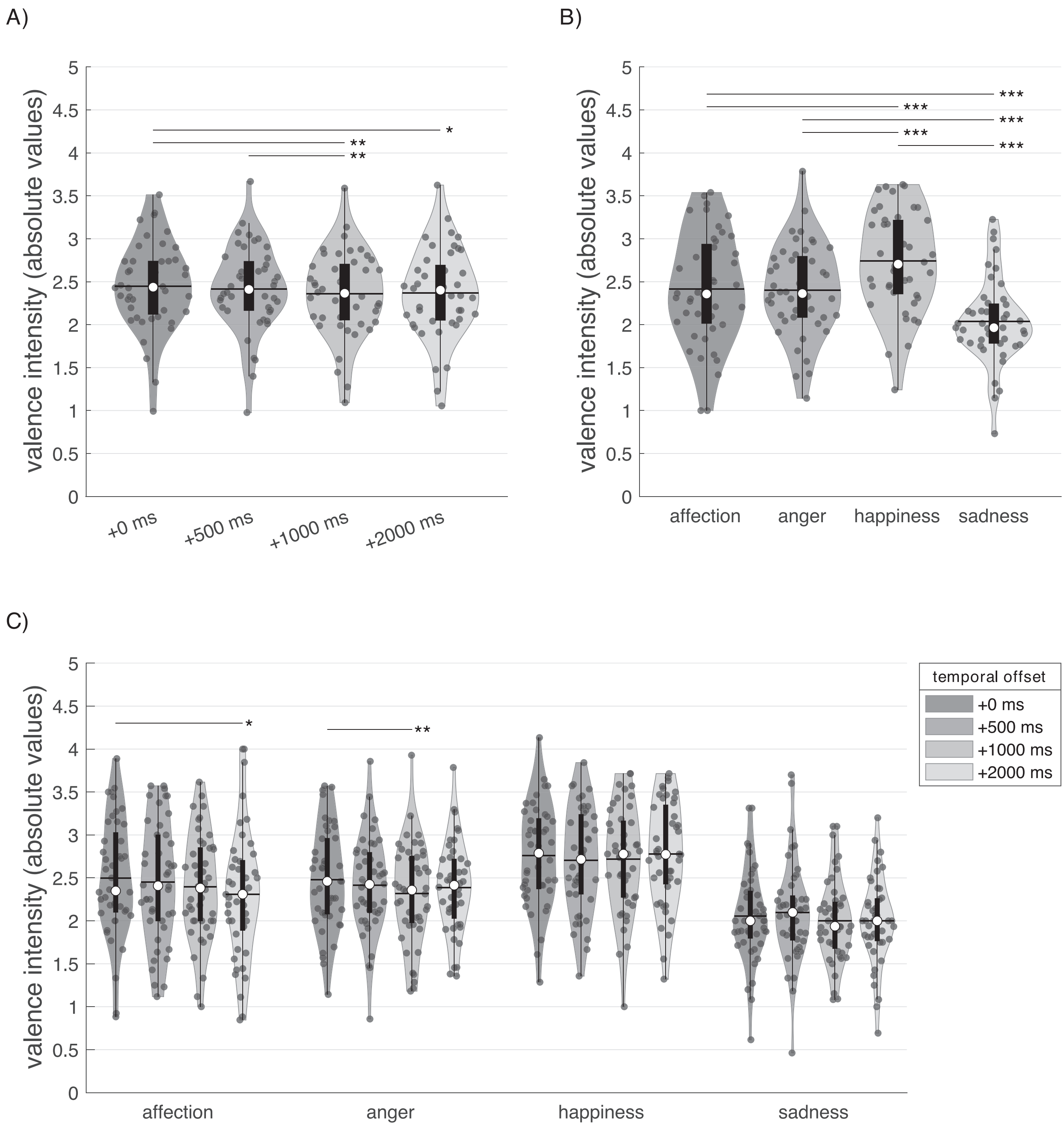

Fig. 5. *Influence of temporal offset on subjective valence perception.* Absolute valence rating for A) all four temporal offsets (+0 ms, +500 ms, +1000 ms, +2000 ms); and B) all four emotion categories (affection, anger, happiness, sadness); and C) per emotion category separated by temporal offset. *Note*: Values for affection and happiness indicate *positive* valence intensity while values for anger and sadness indicate *negative* valence intensity (see Appendix A, Fig. A.2). Violin plots display the distribution of individual performances (single, colored dots). Each violin plot contains information about the median (i.e., white dot), the 25th and 75th percentiles (i.e., bold vertical line), and the mean (i.e., horizontal line). Horizontal lines between variables indicate significance of difference between temporal offsets. Significance level is indicated by asterisks (*$p < .05$; **$p < .01$; ***$p < .001$).

3.3. Temporal decoupling and the perception of valence intensity

We further investigated whether the perception of valence intensity (i.e. absolute valence) was affected by temporal offsets using a two-way repeated measures ANOVA. Again, our results revealed a main effect of temporal offset, $F(2.47, 106.41) = 6.78, p < .001, \eta_p^2 = 0.14$. Bonferroni-corrected post hoc analyses showed that valence intensity ratings did not differ between the +0 ms and + 500 ms condition ($M_{diff} = 0.03, p = 1$). However, valence ratings in the +0 ms condition were judged to be more *intense* (i.e., absolute valence) compared to the +1000 ms ($M_{diff} = 0.09$, $p < .01$, 95% CI [0.02, 0.16]) and + 2000 ms ($M_{diff} = 0.08, p < .05$, 95% CI [0.01, 0.15]) conditions. Moreover, valence ratings within +500 ms sequences were judged to be significantly more intense than +1000 ms sequences ($M_{diff} = 0.06$, $p < .01$, 95% CI [0.02, 0.10]) (see Fig. 5A).

Additionally, we found a main effect of emotion category, $F(3, 129) = 36.81, p < .001, \eta_p^2 = 0.46$. Post hoc analyses indicated that happiness received the significantly highest valence intensity ratings of all emotion categories (all $p < .001$). In contrast, sadness received the lowest valence intensity ratings of all categories (all $p < .001$; see Fig. 5B). Finally, we found an interaction between temporal offset and emotion category, $F(9, 387) = 2.02$, $p < .05$, $\eta_p^2 = 0.05$. Post hoc analyses showed that for affection, valence intensity ratings were significantly higher for +0 ms sequences compared to +2000 ms sequences ($M_{diff} = 0.19, p < .05$, 95% CI [0.01, 0.36]). With respect to anger sequences, we found significantly greater intensity ratings for +0 ms sequences compared to +1000 ms sequences ($M_{diff} = 0.16$, $p < .01$, 95% CI [0.05, 0.27]) (see Fig. 5C).

3.4. Associations between movement features and emotion recognition

Results showed that the temporal decoupling led to a decrease in recognition performance exclusively for affection and anger sequences. Therefore, we explored whether movement patterns of affection and anger sequences were also affected systematically by the temporal offset. With respect to affection sequences, repeated measures ANOVA revealed a significant effect for the parameter *motion energy balance*. For anger sequences, in contrast, we found effects for the parameters *average interpersonal distance* and *personal* space (for an overview of all results, see Appendix A, Fig. A.3 and Table A.4).

3.4.1. Motion energy balance

In affection sequences, we found a significant main effect for motion energy balance, $F(1.64, 26.17) = 4.832$, $p < .05$, $\boldsymbol{\eta_p^2} = 0.23$, indicating that this was significantly impacted by the temporal offset. Bonferroni-corrected post hoc tests showed a trend towards a decline in motion energy balance for an increased temporal offset, yet failed to attain the significance level of $p < .05$ (post hoc comparison of +0 ms sequences with +2000 ms sequences resulted in: $M_{diff} = 0.11$, $p = .074$, 95% CI [−0.008, 0.232], see Fig. 6A).

A Wilcoxon signed-rank test revealed that with respect to motion energy balance, there is a significant difference between incorrectly and correctly identified sequences ($Z = 3.26, p < .01, r = 0.56$, see Fig. 6D). More specifically, MEB values were greater for correctly identified trials ($M = 0.84$, $SEM = 0.004$) as compared to incorrectly identified sequences ($M = 0.81$, $SEM = 0.006$).

3.4.2. Average interpersonal distance

For anger sequences, we found effects for average interpersonal distance and personal space. With respect to average interpersonal distance, there was a main effect of temporal offset, $F(3, 48) = 4.168$, $p < .05$, $\eta_p^2 = 0.21$. Bonferroni-corrected post hoc tests indicated significantly greater distances for a temporal offset of +1000 ms compared to a temporal offset of +2000 ms ($M_{diff} = 105.64$, $p < .05$, 95% CI [7.36, 203.93] (see Fig. 6B).

3.4.3. Personal space

We also found a main effect of temporal offset for personal space in anger sequences, $F(1.95, 31.12) = 3.81, p < .05, \eta_p^2 = 0.19$. Bonferroni-corrected post hoc tests indicated significantly less time spent in the personal space of the other agent for a temporal offset of +1000 ms compared to a temporal offset of +2000 ms ($M_{diff} = -13.03$, $p < .05$, 95% CI [−25.63, −0.43], see Fig. 6C).

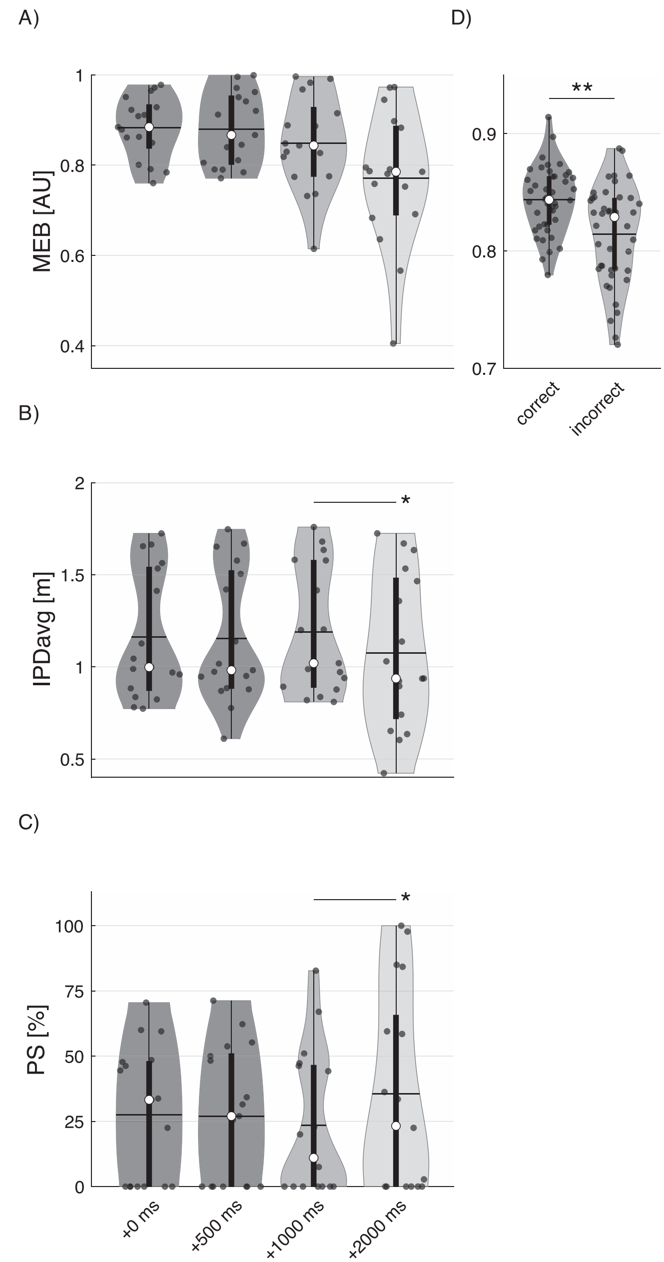

Fig. 6. *Differences in kinematic parameters.* A, B, C: Kinematic differences between temporal offsets (+0 ms, +500 ms, +1000 ms, +2000 ms) for A) motion energy balance (MEB) in affection sequences, B) average interpersonal distance (IPDavg) in anger sequences, and B) personal space (PS) in anger sequences. Violin plots display the distribution of all sequences (single, colored dots represent individual interaction sequences). D: Kinematic differences between correctly and incorrectly identified sequences for motion energy balance in affection sequences. Violin plots display the distribution of mean motion energy balance values (single, colored dots represent average values for each subject). Each violin plot contains information about the median (i.e., white dot), the 25th and 75th percentiles (i.e., bold vertical line), and the mean (i.e., horizontal line). Horizontal lines between variables indicate significance of difference between categories. Significance level is indicated by asterisks (*$p < .05$, *$p < .01$).

4. Discussion

Human sensitivity to emotional body language has been documented in a broad body of literature (for a review, see de Gelder, 2006) that emphasizes our ability to quickly retrieve the state of a person even if observation is based on highly reduced stimuli. We are able to make a number of inferences that lead to attributions of emotional states and intentions, trigger predictions of others' behaviors, and motivate corresponding adjustments of our own behavior. Especially physical properties of the observed movements such as spatiotemporal coupling of actions seem to be a critical source of information for such inferences (see, e.g., Dael, Goudbeek, & Scherer, 2013).

In the present study, we showed that recognition of emotions from dyadic interactions does, indeed, depend on a fixed spatiotemporal coupling of the actions: Temporal decoupling of one of the agent's actions within an emotional interaction leads to a *decrease* in recognition performance. However, this is true exclusively for affection and anger sequences. Recognition of happiness and sadness sequences is not impacted significantly by a temporal decoupling. Furthermore, we found that the perception of valence intensity is also affected in an emotion-specific manner: Affection and anger scenes are perceived as being *less intense* when the sequences are displayed with a temporal offset. However, there is no change in subjective perception of valence intensity when happiness or sadness sequences are displayed with a temporal offset. These results demonstrate that the temporal decoupling of an agent's actions within an emotional interaction alters the perception of emotional intensity in an observed scene, and—depending on the displayed emotion—this leads to a *decrease* in emotion recognition performance.

Using sensitivity measures, we further found that with increasing temporal offset, discrimination abilities worsen. This can either be due to a reduction of 'hits', i.e. the respective emotion is not correctly identified or an increase of 'false alarms', i.e. an emotion category is falsely identified. Criterion values showed that with increasing temporal offset, affection and anger scenes were less likely to be chosen (decrease of hits), while happiness was more likely to be chosen with increasing temporal offset (increase in false alarm). Thus, the tendency to identify a sequence as happiness increases when movements are less tightly coupled.

In a last step, we explored how the administered temporal offset leads to changes to specific movement features that might explain changes in the subjective impression and recognition of the observed emotional interactions. For affection sequences, we found that motion energy balance between the observed agents decreases linearly with increasing temporal offset. This corresponds directly to the recognition performances. Furthermore, we found that motion energy balance is significantly greater for correctly compared to incorrectly identified sequences. With respect to anger sequences, we found changes in the *average interpersonal distance* and *personal space*—albeit not corresponding to the recognition performances. In the following, we shall discuss the main findings in detail.

4.1. Temporal coupling and emotion recognition

Within dyadic interactions, it is necessary to be able to adjust one's actions to those of another person. This involves choosing an appropriate complementary action at an appropriate time, and it requires the ability to predict what others will do next—that is, when and how their actions will unfold (Sebanz & Knoblich, 2009). Even if we are not part of the social interaction ourselves, we seem to possess implicit knowledge about the natural dynamics of the observed human interaction (von der Lühe et al., 2016). Studies on visual discrimination have shown that, with respect to interactive activities, the actions of one agent can be used to predict what the second agent will do (Manera et al., 2011, 2013; Neri et al., 2006; von der Lühe et al., 2016). Manera et al. (2013), for instance, demonstrated that manipulating the timing of the actions of two agents had a detrimental effect on visual discrimination performance only within a *communicative* condition—that is, a condition in which the gestures performed by one agent were related to those performed by the other agent. In an individual condition in which the actions of the two agents were not related, no effect of timing manipulation was found.

The present findings show that a temporal decoupling of emotional interactions between two agents affects the recognition of the emotional content. More specifically, recognition performances decline linearly with increasing temporal offset. Interestingly, this effect is emotion-specific as the decline of recognition performances is observed only for temporally decoupled affection and anger sequences. Recognition of happiness and sadness is not affected by temporal decoupling. This suggests that some emotions might have a stronger interpersonal and communicative character than others and depend more on reciprocal actions between the agents, whereas others may be acted out more on an individual level.

Indeed, this idea seems to be underpinned by different studies. Clarke et al. (2005), for instance, found that romantic love and joy are recognized more reliably when they are depicted within an interaction (i.e., dyad) instead of individually (i.e., monad). The authors argued that this may be because these emotions are particularly socially expressive and have a stronger interpersonal character. We, however, did not observe a decrease in recognition performance for happiness sequences. Here, one has to take the methodological differences of both studies into account. Clarke et al. (2005) presented monads (i.e. single agents) taken from the dyadic interactions. In our case, both agents were depicted, but with different temporal offsets. Thus, in the present study, movement information of a second agent was continuously available to the observer, perhaps impacting recognition performances beneficially.

Another approach by Abramson, Petranker, Marom, and Aviezer (2021) suggests that some emotions can be recognized more easily if the social context is *congruent*, i.e. both agents display the same emotions, while other emotions are recognized through *functional-relational* contexts. One example for this is the functional-relational context between fear and anger. Abramson et al. (2021) showed that fear recognition is facilitated by an interacting angry agent more strongly than by a fearful agent. Although both agents were specifically instructed to display the same emotion in our study, especially anger may inherently contain more complementary, or *functional-relational* actions. Hence, temporally decoupling the actions may not allow the perceiver to 'match' the respective actions leading to the reduced recognition.

In affection scenes, movements are rather similar and less complementary. For instance, if one agent offers a hug, the second agent is likely expected to reciprocate the hug within a certain time frame. However, if the hug is not returned, it may create ambiguity and a weakened perception of connection between the agents. Studies showed that interactional synchrony (i.e. movements matched in time) is associated with affiliation and romantic attraction between people (Hove & Risen, 2009; Prochazkova, Sjak-Shie, Behrens, Lindh, & Kret, 2022). In this regard, Hove and Risen (2009) suggested that a third person observer will infer closeness between two individuals when noticing interactional synchrony. In our case, temporal decoupling may lead to a reduction of interactional synchrony within affection sequences, resulting in a decrease of recognition performances.

Our criterion analyses revealed that while anger and affection were less frequently identified with increasing temporal offset, the tendency to identify the respective trial as happiness increased. Although speculative, this underpins the idea that happiness recognition, as compared to affection and anger, is less dependent on a strong coupling of the agent's actions.

However, it should be noted that by introducing a temporal offset, we do not only decouple but also remove certain movements. Thus, it is likely that our results cannot solely be attributed to a lack of spatiotemporal contingency. If a certain emotion, anger for instance, is indeed dependent on a functional-relational context, and a complementary movement is not only presented at the wrong time but not presented at all (i.e. action contingency is fully removed), this might worsen the recognition ability even more strongly.

4.2. The role of interindividual movement parameters

The assumption that certain emotions may have a stronger interpersonal character is seemingly supported by our exploratory kinematic analyses. One has to bear in mind, however, that these analyses were conducted post-hoc and cannot establish a causal link between movement characteristics and participant judgments. Yet, we found that *motion energy balance* seems to play a role in the recognition of affection. Motion energy balance is a parameter that provides information about how much of the total amount of body movement within one sequence can be ascribed to each interactive partner. With respect to the temporal offset that was introduced, we found a linear decrease in motion energy balance with increasing temporal offset that corresponds directly to the linear decrease in recognition accuracy. Moreover, we found that motion energy balance is significantly greater in correctly identified trials as compared to incorrectly identified trials. This means that for correctly identified sequences, the total amount of body movement was more balanced between the agents as compared to incorrectly identified sequences.

Although these analyses do not allow a causal conclusion, we speculate that recognition accuracy of affection may be facilitated by highly balanced motion energy between the agents. In this vein, Keck, Zabicki, Bachmann, Munzert, and Krüger (2022) revealed that balance within emotional interactions seems to be an important property for the observer to generate an emotional percept. For example, they found that especially motion energy balance between two agents allows for a perceptual distinction between positive and negative valence in observed interactions. Hence, social context information such as motion energy balance is particularly important for inferring the emotional content, especially when the depicted emotions depend on reciprocal actions (T. J. Clarke et al., 2005; Kret & de Gelder, 2010).

With respect to interpersonal movement parameters and anger recognition, we were unable to identify changes in movement parameters that correspond directly to the behavioral ratings. Although we observed an effect for the average interpersonal distance and the personal space shared by the agents, these effects do not give reason to assume that they drive the behavioral effects caused by the temporal offset as they do not correspond to the participant ratings.

4.3. Temporal coupling and the subjective perception of valence intensity

So far, recognition performance for emotional stimuli has been investigated only with respect to the influence of a temporal decoupling between two interacting agents. However, we were also interested in a more subjective dimension of emotion perception: the perceived emotional valence intensity. Therefore, we explored how temporal coupling is linked to how negatively or positively a scene is perceived.

We found that perception of valence intensity is dependent on a temporal coupling between the actions of two agents. Although there was no overall difference between the original timing and a delay of +500 ms, the present data reveal that the sequences are perceived as *more intense* in terms of valence when they are presented in their original form as compared to a delay of +1000 and + 2000 ms. This effect, again, is emotion-specific. Whereas valence intensity judgments for sadness and happiness do not differ regardless of the temporal offset, anger and affection trials receive higher valence intensity ratings (i.e., more negative for anger and more positive for affection) and are perceived as more intense in the original timing set.

To the best of our knowledge, the link between temporal coupling and valence perception of emotional body movements has not been investigated so far. Here, we show for the first time that the perception of valence intensity depends on (inter-)actions that occur in a time-locked manner—albeit this is not equally true for all emotion categories. Specifically, affection and anger scenes are perceived as more *intense* when the actions of both agents occur in a temporally and spatially meaningful pattern. In contrast, no effect of timing manipulation is observed when happiness and sadness sequences are displayed. This further supports the notion of the relevance of contextual information.

Taking these results together, time-locked contextual social information seems to enhance the perceived intensity of emotional body language and, thereby, might also facilitate emotion recognition.

4.4. Temporal coupling, movement contingency, and action prediction

In light of the current results, the question arises why spatiotemporal coupling plays such a fundamental role in the perception of certain emotional interactions. Evidence suggests that the contingency of movement patterns – that is, the temporal and spatial relatedness of movements – facilitates the perception of “meaning” in a visual stimulus (Castelli, Happé, Frith, & Frith, 2000; Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007). Consequently, the term “social contingency” has been used to describe the relatedness of the complex, coordinated actions within dyadic interactions that result in a meaningful movement pattern (Moran, Dumas, & Symons, 1992). Hence, motion congruency might serve as a fundamental factor in both detecting and understanding complex actions in human interaction. Dissolving these movement contingencies by delaying or removing the actions of one of the agents, in contrast, results in changes of interaction-specific movement parameters, decreased emotion recognition performance, and also a diminished perception of emotional intensity. This suggests that these (spatiotemporal) action contingencies are critical to the perception of emotional dyadic interactions.

4.5. What mechanisms do humans utilize to understand affective states when they are observing complex human interactions?

Findings regarding action simulation processes suggest that the observation of actions is a predictive activity (Flanagan & Johansson, 2003; Graf et al., 2007; Verfaillie & Daems, 2002). When we observe someone's current actions, we automatically anticipate and predict their future ones. In a series of behavioral experiments, Graf et al. (2007) demonstrated that we seem to make real-time predictions about how actions will unfold. Besides predicting *how* an action will unfold, it has been suggested that we also predict *when* it will unfold. More precisely, we are able to accurately apply temporal predictions generated in our own motor system to observed actions (Sato, 2008; Wilson & Knoblich, 2005).

For the observation of social interactions, it has been demonstrated that the movements of one agent are used to predict *how* and *when* the second agent will respond – provided that the actions of the two agents are related in time (Manera et al., 2013). Thus, it seems plausible that a temporal decoupling will cause a decrease in emotion recognition as well as a diminished perceived intensity of the emotion due to a violation of predicted interaction consequences – however, only under the condition that the interactive movements are functionally related or have a communicative function.

On a neurophysiological level, it is suggested that action prediction within dyadic interactions is rooted in activation of the so-called action observation network (AON; Georgescu et al., 2014). The AON is believed to integrate observed actions of others with an individual's personal motor repertoire, thereby making sense of an action (*for a meta-analytical overview, see* Caspers, Zilles, Laird, & Eickhoff, 2010). Georgescu et al. (2014) found that the AON is preferentially engaged by contingency within movement patterns when interacting dyads are displayed instead of individually acting monads. The authors argued that the complexity of action representation is determined not merely by the communicative nature of the observed behavior but, indeed, by the relational context in which such behavior is performed. Thus, the AON could be considered to be an early key processing component that supports and contributes to the understanding of nonverbal social interaction; and an automatic movement analysis might be performed to adequately understand an observed agent's social intentions (Gallese, 2006; Gallese & Goldman, 1998; Jacob & Jeannerod, 2005).

5. Conclusion

The present findings extend previous evidence on the perception of emotional body language in social interactions by emphasizing the fundamental role of the spatiotemporal coupling of movements within emotional interactions. We show that a valid spatiotemporal coupling between actions is especially necessary for the recognition of anger and affection, but not for happiness or sadness. Moreover, the same effect becomes visible for the perception of the valence intensity of a sequence – however, especially for anger and affection sequences. Hence, our findings suggest that the recognition of emotions seems to operate through *functional relations* between two agents that are specific to the emotion that is displayed.

6. Limitations and future implications

With regard to limitations and future implications, two aspects should be highlighted. As already mentioned, one possible limitation in the interpretation of our data might result from the fact that the introduced temporal offsets do not only lead to a stepwise temporal decoupling of (inter-)actions, but also lead to a change in (kinematic) information that is provided. Thus, the temporal decoupling is also accompanied by a change in action contingencies between the two agents. In case of affection, for instance, the reduced recognition performance as well as the reduced perceived valence intensity might, therefore, be driven by the fact that if one agent initiates a hug, the other agent reciprocates the hug at an inappropriate time or does not reciprocate it at all. This means that a hug or a threatening fist-raise is missing its corresponding response, rather than just being temporally misaligned. This makes it difficult to draw the final conclusion that the present results are solely the result of disrupted temporal contingencies. For future studies, it would therefore be important to introduce either longer sequences or, shorter delays that still preserve most of the original scene information. That way, one would not only be able to identify critical point in time at which a certain delay worsens the recognition of an emotion but also ensure that no action contingency is lost completely.

The second aspect relates to the post-hoc analysis of quantitative movement features. This analysis was done in order to gain a deeper understanding of the observed behavioral effects by evaluating objective movement properties of our stimuli. However, due to the present experimental design, we were unable to conduct analyses that allow for an identification of a direct link between participant judgments and kinematic parameters. Thus, we cannot draw causal conclusions about the changes in recognition accuracy and the changes in certain kinematic properties of the stimulus. Future studies should make use of prediction models that allow for an identification of all relevant kinematic parameters that drive the perception of the participant. In this regard, one should not only investigate interpersonal parameters but also intrapersonal parameters to explain why certain emotions may not be affected by a loss of (temporal) action contingency.

Data and code availability statement

The stimulus set, experimental data, and code to reproduce the results presented here are openly available in OSF at https://doi.org/10.17605/OSF.IO/GT3QU or on Gitlab (https://gitlab.com/azabicki/tcei). The SAMI toolbox is archived in Zenodo (Zabicki & Keck, 2021) as well as openly available on Github (https://github.com/azabicki/SAMI).

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under a grant (IRTG 1901) for JB and JK.

CRediT authorship contribution statement

Julia Bachmann: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Britta Krüger:** Validation, Writing – review & editing, Supervision. **Johannes Keck:** Formal analysis, Writing – review & editing. **Jörn Munzert:** Writing – review & editing, Supervision, Funding acquisition. **Adam Zabicki:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.

Declaration of Competing Interest

The authors have no competing interests to declare.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank Jonathan Harrow for native speaker support. We also thank Marie-Claire Englert, Franziska Schulz and Fabian Kowallik for their support during data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105267.

References

Abramson, L., Petranker, R., Marom, I., & Aviezer, H. (2021). Social interaction context shapes emotion recognition through body language, not facial expressions. *Emotion, 21*(3), 557–568. https://doi.org/10.1037/emo0000718

Bachmann, J., Zabicki, A., Munzert, J., & Krüger, B. (2020). Emotional expressivity of the observer mediates recognition of affective states from human body movements. *Cognition and Emotion, 34*(7), 1370–1381. https://doi.org/10.1080/02699931.2020.1747990

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science, 19*(3), 183–188. https://doi.org/10.1177/0963721410370131

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *Psychological Corporation*. https://doi.org/10.1037/t00742-000

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage, 50*(3), 1148–1167. https://doi.org/10.1016/j.neuroimage.2009.12.112

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage, 12*(3), 314–325. https://doi.org/10.1006/nimg.2000.0612

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception, 34*(10), 1171–1180. https://doi.org/10.1068/p5203

Dael, N., Goudbeek, M., & Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception, 42*(6), 642–657. https://doi.org/10.1068/p7364

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424*(6950), 769–771. https://doi.org/10.1038/nature01861

Fridlund, A. J. (1992). The behavioral ecology and sociality of human faces. In M. S. Clarke (Ed.), *Emotion* (pp. 90–121). Sage Publications, Inc.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion*. Lawrence Erlbaum Associates Publishers.

Gallese, V. (2006). Embodied simulation: From mirror neuron systems to interpersonal relations. In G. Bock, & J. Goode (Eds.), *Vol. 278. Empathy and fairness: Novartis foundation symposium*. John Wiley.

Gallese, V. (2009). Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic Dialogues, 19*(5), 519–536. https://doi.org/10.1080/10481880903231910

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*(9), 396–403. https://doi.org/10.1016/j.tics.2004.07.002

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience, 7*(3), 242–249. https://doi.org/10.1038/nrn1872

de Gelder, B., & Hortensius, R. (2014). The many faces of the emotional body. In J. Decety, & Y. Christen (Eds.)*, Vol. 21. New frontiers in social neuroscience* (pp. 153–164). Springer. https://doi.org/10.1007/978-3-319-02904-7_9.

Georgescu, A. L., Kuzmanovic, B., Santos, N. S., Tepest, R., Bente, G., Tittgemeyer, M., & Vogeley, K. (2014). Perceiving nonverbal behavior: Neural correlates of processing movement fluency and contingency in dyadic interactions: Perceiving nonverbal interactive behavior. *Human Brain Mapping, 35*(4), 1362–1378. https://doi.org/10.1002/hbm.22259

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience, 19*(11), 1803–1814. https://doi.org/10.1162/jocn.2007.19.11.1803

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage, 36*, T22–T32. https://doi.org/10.1016/j.neuroimage.2007.03.017

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science, 11*(3), 223–228. https://doi.org/10.1111/1467-9280.00245

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition, 27*(6), 949–961. https://doi.org/10.1521/soco.2009.27.6.949

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: A critique. *Trends in Cognitive Sciences, 9*(1), 21–25. https://doi.org/10.1016/j.tics.2004.11.003

Keck, J., Zabicki, A., Bachmann, J., Munzert, J., & Krüger, B. (2022). Decoding spatiotemporal features of emotional body language in social interactions. *PsyArXiv.*. https://doi.org/10.31234/osf.io/62uv9

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Kret, M. E., & de Gelder, B. (2010). Social context influences recognition of bodily expressions. *Experimental Brain Research, 203*(1), 169–180. https://doi.org/10.1007/s00221-010-2220-8

Lorey, B., Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., … Munzert, J. (2012). Confidence in emotion perception in point-light displays varies with the ability to perceive own emotions. *PLoS One, 7*(8), Article e42169. https://doi.org/10.1371/journal.pone.0042169

von der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 371*(1693), 20150373. https://doi.org/10.1098/rstb.2015.0373

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS One, 6*(1), Article e14594. https://doi.org/10.1371/journal.pone.0014594

Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time will show: Real time predictions during interpersonal action perception. *PLoS One, 8*(1), Article e54949. https://doi.org/10.1371/journal.pone.0054949

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Biological Sciences, 258*(1353), 273–279. https://doi.org/10.1098/rspb.1994.0173

Moran, G., Dumas, J. E., & Symons, D. K. (1992). Approaches to sequential analysis and the description of contingency in behavioral interaction. *Behavioral Assessment, 14*(1), 65–92.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience, 9*(9), 1186–1192. https://doi.org/10.1038/nn1759

Petrini, K., Piwek, L., Crabbe, F., Pollick, F. E., & Garrod, S. (2014). Look at those two!: The precuneus role in unattended third-person perspective of social interactions: Precuneus and third-person interaction. *Human Brain Mapping, 35*(10), 5190–5203. https://doi.org/10.1002/hbm.22543

Prochazkova, E., Sjak-Shie, E., Behrens, F., Lindh, D., & Kret, M. E. (2022). Physiological synchrony is associated with attraction in a blind date setting. *Nature Human Behaviour, 6*(2), 269–278. https://doi.org/10.1038/s41562-021-01197-3

Rindermann, H. (2009). *Emotionale-Kompetenz-Fragebogen: EKF*. Einschätzung emotionaler Kompetenzen und emotionaler Intelligenz aus Selbst-und Fremdsicht: Hogrefe.

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

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of reseach synthesis* (pp. 231–244). Russell Sage Foundation.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation. *Gender Recognition, and Deceptive Intention., 112*(4), 585–615. https://doi.org/10.1037/0096-3445.112.4.585

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and Cognition, 17*(4), 1219–1227. https://doi.org/10.1016/j.concog.2008.01.003

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science, 1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Consulting Psychologist Press.

Templeton, G. F. (2011). A two-step approach for transforming continuous variables to normal: Implications and recommendations for IS research. *Communications of the Association for Information Systems, 28*(4), 41–58. https://doi.org/10.17705/1CAIS.02804

Thurman, S. M., & Lu, H. (2014). Perception of social interactions for spatially scrambled biological motion. *PLoS One, 9*(11), Article e112539. https://doi.org/10.1371/journal.pone.0112539

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*(5), 2. https://doi.org/10.1167/2.5.2

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition, 9*(1–2), 217–232. https://doi.org/10.1080/13506280143000403

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655–664. https://doi.org/10.1016/S0896-6273(03)00679-2

Wilson, M., & Knoblich, G. (2005). *The case for motor involvement in perceiving conspecifics., 131*(3), 460–473. https://doi.org/10.1037/0033-2909.131.3.460

Zabicki, A., & Keck, J. (2021). SAMI: Similarity analysis of human movements and interactions. *Zenodo.*. https://doi.org/10.5281/zenodo.4764552

5 Does co-presence affect the way we perceive and respond to emotional interactions?

This chapter has been published as:

Bachmann, J., Zabicki, A., Gradl, S., Kurz, J., Munzert, J., Troje, N. F., & Krueger, B. (2021). Does co-presence affect the way we perceive and respond to emotional interactions? *Experimental Brain Research*, *239*(3), 923–936. doi: 10.1007/s00221-020-06020-5

Experimental Brain Research
https://doi.org/10.1007/s00221-020-06020-5

RESEARCH ARTICLE

Does co-presence affect the way we perceive and respond to emotional interactions?

Julia Bachmann[1] · Adam Zabicki[1] · Stefan Gradl[2] · Johannes Kurz[1] · Jörn Munzert[1,3] · Nikolaus F. Troje[4] · Britta Krueger[1]

Received: 21 August 2020 / Accepted: 15 December 2020

Abstract

This study compared how two virtual display conditions of human body expressions influenced explicit and implicit dimensions of emotion perception and response behavior in women and men. Two avatars displayed emotional interactions (angry, sad, affectionate, happy) in a "pictorial" condition depicting the emotional interactive partners on a screen within a virtual environment and a "visual" condition allowing participants to share space with the avatars, thereby enhancing co-presence and agency. Subsequently to stimulus presentation, explicit valence perception and response tendency (i.e. the explicit tendency to avoid or approach the situation) were assessed on rating scales. Implicit responses, i.e. postural and autonomic responses towards the observed interactions were measured by means of postural displacement and changes in skin conductance. Results showed that self-reported presence differed between pictorial and visual conditions, however, it was not correlated with skin conductance responses. Valence perception was only marginally influenced by the virtual condition and not at all by explicit response behavior. There were gender-mediated effects on postural response tendencies as well as gender differences in explicit response behavior but not in valence perception. Exploratory analyses revealed a link between valence perception and preferred behavioral response in women but not in men. We conclude that the display condition seems to influence automatic motivational tendencies but not higher level cognitive evaluations. Moreover, intragroup differences in explicit and implicit response behavior highlight the importance of individual factors beyond gender.

Keywords Emotion perception · Co-presence · Implicit response behavior · Explicit response behavior · Gender differences

Introduction

Recognizing human body expressions is vital for responding appropriately to social cues. Bodily expressions do not just convey a person's affective states, they also inform about action demands. For instance, a fearful face tells how someone is feeling, but it does not necessarily indicate how to respond. However, if a person reacts fearfully with their whole body by, for instance, drawing away from a potential threat, one can prepare for a concrete action. Evolutionary psychology considers emotions to be action dispositions that humans possess to navigate in the world (Bradley et al. 2001; de Gelder 2006; de Gelder et al. 2015; LeDoux 1996). Different tools have been used to study the perception of emotional body language. Some of the most common include video displays, stick figures, and point-light displays—that is, displays depicting biological motion by the kinematics of light points located on an actor's joints (see, e.g., Atkinson et al. 2004; Kaletsch et al. 2014a, b; Krüger et al. 2018;

Communicated by Francesco Lacquaniti.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00221-020-06020-5.

✉ Julia Bachmann
julia.bachmann@sport.uni-giessen.de

1 NeuroMotor Behavior Lab, Department of Psychology and Sport Science, Justus-Liebig-University, Giessen, Germany

2 Machine Learning and Data Analysis Lab, Faculty of Engineering, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

3 Center for Mind, Brain and Behavior (CMBB), Philipps University of Marburg and Justus Liebig University, Giessen, Germany

4 BioMotionLab, Department of Biology and Centre for Vision Research, York University Toronto, Toronto, Canada

Published online: 11 January 2021

Lorey et al. 2012). There is evidence that humans are able to quickly identify a person's affective state even from highly impoverished stimulus material, thus demonstrating the importance of top-down knowledge in enabling people to quickly make sense of what they are seeing.

This process of recognizing emotions is thought to be followed by a motivation to respond to them that generates so-called avoid-approach behavior (Bradley et al. 2001; Stins et al. 2011). This is based on the idea that positively valenced stimuli elicit approach tendencies, whereas negative stimuli elicit avoidance tendencies (e.g., Chen and Bargh 1999; Seidel et al. 2010b). These basal motivational tendencies can be separated into implicit and explicit stages. From an evolutionary perspective, implicit response tendencies are characterized by an automatic evaluation of incoming stimuli without conscious effort to quickly generate behavioral responses (Elliot and Covington 2001). Explicit response tendencies, in contrast, may represent more complex cognitive evaluations of a stimulus that occur at a later stage, and may potentially be affected more by social learning and reinforcement. Implicit response tendencies can, for instance, be operationalized via postural displacement. For example, Stins and Beek (2007) observed an increase in anterior center of pressure (COP) displacement in response to unpleasant images. Other researchers, in contrast, found a decrease in body sway during the presentation of unpleasant pictures. They interpreted this as freezing behavior, highlighting that the defensive system presents two basal dispositions: freezing or action (Azevedo et al. 2005; Facchinetti et al. 2006).

Many of these paradigms, however, are based on static or dynamic picture and video displays. Such "pictorial designs" separate observers physically through a screen without them being able to experience another's presence—something that may be crucial for social cognition. Virtual environments offer promising tools with which to create vivid and realistic perceptual experiences by allowing for a sense of presence within virtual space (Seidel et al. 2010b; Slater 2009). Creating a sense of presence depends on valid sensorimotor contingencies between own body movements and the resulting changes in visual input. These sensorimotor contingencies are then thought to create the embodiment in space or "place illusion" (Slater 2009). Then, one is in "presence mode" compared to "picture mode" (Troje 2019). If the space in which one is embodied is also shared by other (virtual) agents, it is called "co-presence." When in presence mode, the visual system seems to be in a similar state to when it is when interacting with the real world. This leads to an activation of the perceptual, vestibular, proprioceptive, and autonomic nervous systems in ways similar to those in a real-life situation (Slater 2003, 2009; Troje 2019).

Whereas the experimental setting may influence how we perceive emotions and how we respond to them, a considerable amount of research has demonstrated that emotion perception and response behavior can also be modulated by characteristics of the individual such as gender (Alaerts et al. 2011; Bradley et al. 2001; Hillman et al. 2004; Hoffmann et al. 2010). Although evidence is highly heterogeneous, studies suggest that women recognize emotions better than men (Alaerts et al. 2011; Hoffmann et al. 2010; Thayer and Johnsen 2000). More specifically, it has been shown that women detect emotions both more quickly and more accurately from facial or bodily expressions (Alaerts et al. 2011; Hampson et al. 2006). However, differences in classification accuracy between males and females have been found only during the display of subtle emotional expressions (Hoffmann et al. 2010; Montagne et al. 2005). Gender has also been shown to mediate response tendencies. Although the effect on explicit behavioral tendencies has largely been neglected, implicit paradigms have shown that women exhibit increased backward movement, as measured by COP displacement, in response to unpleasant stimuli as well as greater defensive reactivity to aversive pictures, as measured by a deceleration in heart rate. Men, in contrast, exhibit increased anterior movement in response to unpleasant stimuli and less defensive reactivity (Bradley et al. 2001; Hillman et al. 2004).

In this vein, it is important to implement more ecologically valid paradigms that investigate such phenomena reliably. So far, there have been no systematic investigations of whether emotion perception and response behavior are, indeed, modified depending on whether stimuli are presented in the "pictorial" space of screens or in a "visual" space shared with other agents that achieves a feeling of co-presence. A virtual reality (VR) paradigm presents a suitable method to investigate emotion perception, perceived emotional intensity, and the associated action tendencies under different conditions (Seidel et al. 2010b; Visch et al. 2010).

The research aim

The present study follows two central goals: by taking a multisystem approach, it aims to shed light on how different virtual environments affect the perception of emotional interactions (i.e., anger, sadness, affection, happiness) and trigger action tendencies in response to emotional interactions. More specifically, we tested whether co-presence modulates perception and associated response behavior toward emotional body language by creating two display conditions within a virtual reality: a pictorial display condition that simulates a conventional experiment conducted on a computer monitor and a visual display condition containing valid sensorimotor contingencies that enhance the participant's sense of co-presence and agency (for the conceptual distinction between pictorial and visual spaces, see Koenderink and van Doorn 2012; Troje 2019). We then compared the explicit rating of emotional valence as well as the explicit and implicit tendency to act in the two display conditions. Implementing these measures, we cannot only evaluate a possible influence of the virtual condition on each of them but we can also further elucidate the linkage between perception and behavioral action. Thus, we asked participants to observe emotional interactions within the two conditions and explicitly evaluate (a) the behavioral tendency to approach or avoid the observed scene, and (b) the emotional valence (i.e., positive or negative) of the scene. These explicit ratings reflect a conscious cognitive judgment of the observed stimulus. Moreover, we measured implicit bodily response behavior toward the stimulus via COP displacement, allowing us to unmask automatic response behavior that takes place at earlier stages of emotion perception processing. To assess whether subjective presence was associated with autonomic arousal, we applied a continuous skin conductance measure.

Finally, we explored whether and how emotion perception and response tendencies were modulated by the observer's gender. More specifically, we asked whether women and men differed in judging emotional valence and in their bodily response to emotional interactions both explicitly and implicitly.

Materials and methods

Participants

A total of 76 healthy adults, including 39 women ($M_{age} = 23.21$, SD = 2.58) and 37 men ($M_{age} = 25.86$, SD = 4.64) with normal or corrected-to-normal vision participated in the experiment. All participants gave written informed consent to take part in this study. None of the participants reported any history of psychiatric or neurological disorders and no current abuse of drugs or any psychoactive medication. The protocol was approved by the Institutional Ethics Committee and was conducted according to the principles of the Declaration of Helsinki.

We assessed aspects of personality with the emotional competence questionnaire (EKF, Rindermann 2009) and controlled for each participant's affective state with Beck's Depression Inventory (BDI-II; Beck et al. 1996) and the State-Trait Anxiety Inventory (STAI; Spielberger et al. 1983). With regard to state anxiety, participants' average scores ranged from 23 to 57 ($M_{female} = 34.56$; $SD_{female} = 7.56$; $M_{male} = 35.62$, $SD_{male} = 8.01$). Scores on the trait anxiety questionnaire ranged from 21 to 59 ($M_{female} = 37.10$; $SD_{female} = 9.18$; $M_{male} = 35.73$, $SD_{male} = 8.42$), with higher scores indicating greater anxiety. BDI scores ranged from 0 to 22 ($M_{female} = 5.68$; $SD_{female} = 5.18$; $M_{male} = 4.38$, $SD_{male} = 3.65$). Female and male participants did not differ in their mean STAI or BDI-II scores as shown by nonsignificant Mann–Whitney *U* tests (all *p*s > 0.05, see supplementary material S1). Furthermore, no correlation was found with our dependent variables, see supplementary material table S2–S4.

Stimuli

Creating the stimulus set

Stimuli were created with a motion capture system (VICON, Oxford, UK) that recorded the position of 41 markers attached to predefined anatomical landmarks. Eight pairs of nonprofessional actors were asked to portray one of the following four emotional states within a dialog: anger, sadness, affection, or happiness. Due to its strong interpersonal component, we included affection even though it is not considered to be a basic emotion (see Clarke et al. 2005). To increase the variability of the movements, each emotional state was portrayed in three intensities: low, medium and high. Actors were specifically asked to act out the same emotion. To facilitate a symmetric behavioral pattern, all actors received scripts of emotional situations that they were instructed to perform. They were asked to act intuitively within the context of the given situation, allowing for freedom in their expressions. In the next step, we postprocessed the motion capture data and edited the scenes into 3-s interaction sequences.

Stimulus selection

Finally, we randomly selected 96 point-light stimuli (24 × 4 emotions) and tested their recognizability in a separate pilot study ($n = 36$). Scenes had to meet two main criteria: first, mean valence ratings had to reflect the displayed emotion (i.e., negative ratings for anger and sadness and positive ratings for affection and happiness). Second, mean valence ratings between − 1 and 1 were excluded due to their ambiguity or lack of emotionality. Finally, the remaining trails were balanced (i.e., 12 sequences per emotion), leading to a final stimulus set of 48 emotional scenes. For the present study, we applied a MoSh algorithm (Loper et al. 2014) to the motion capture data (i.e., point-light displays) to create avatars for the virtual environment (for exemplary movies, see supplementary material Video S1–S8).

Materials and apparatus

Virtual environment

Using the Unity3D game engine by Unity Technologies (http://unity3d.com), we created two display conditions that we presented to participants in VR. The visual condition depicts a three-dimensional space in which participants share a common space with the dynamic avatars who are engaging in social interaction. The visual scene is presented stereoscopically and also responds with motion parallax contingent to the participant's body movements. In the pictorial condition, participants are standing in front of a virtual computer screen placed on a virtual table. In this case, stereopsis and motion parallax indicate to the participant that the screen is flat and that the visual scene being presented is projected onto it. Therefore, it is thought to elicit a lower subjective feeling of co-presence than in the visual condition (see Fig. 1a, b). Using the HTC Vive Headset, the stimulus material was presented via SteamVR software (http://steamvr.com). In both cases, VR was rendered at 90 Hz with a display resolution of 2160 × 1200 pixels and a field of view of about 110°. Stimuli were presented in front of the participants who were placed at the same location at the beginning of each sequence.

Explicit ratings

Rating responses were collected with the HTC Vive controller placed in the participant's dominant hand. After stimulus presentation, participants were shown a virtual rating scale and asked to judge (a) their tendency to respond to the stimulus (i.e., whether they would like to approach or avoid the scene) and (b) the valence of the stimulus (i.e., how positively or negatively they perceived it).

Implicit measurements

Participants stood on a mobile force plate (Accu Gait System, AMTI Force and Motion, Watertown, MA) throughout the experiment. Changes in force and torque in the x, y, and z dimensions were recorded during each stimulus presentation (i.e., for 4 s at a rate of 50 Hz). Continuous EDA measurement was applied to control for sympathetic nervous activity. We did this by tying two Ag–AgCl electrodes filled with an electrolytic gel mixture (GEL101) to the distal phalanges of the nondominant hand. The signal was recorded at a sampling rate of 1000 Hz. Physiological responses were registered through a physiological amplifier BIOPAC MP36 (BIOPAC Systems, Inc., Goleta, CA).

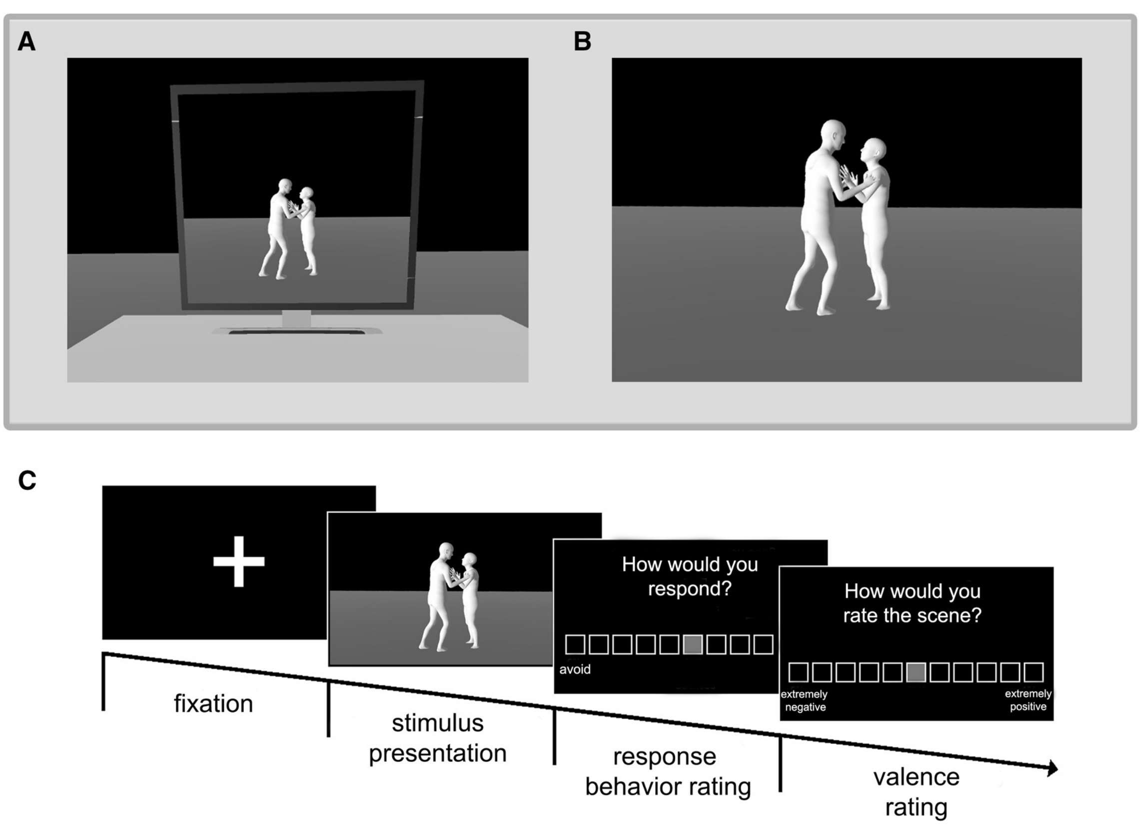

Fig. 1 Display conditions and experimental timeline. The stimulus sequences are presented within **a** the pictorial condition (i.e., a virtual computer monitor) and **b** the visual condition (i.e., the observer shares a common space with the stimuli). **c** Temporal structure of one trial

Experimental design

In the present study, we implemented a crossover design in which all participants underwent both conditions (i.e. visual condition, pictorial condition). The conditions were separated into two blocks. The first block that was presented was alternated between participants. More specifically, if one participant started with the visual condition, the next participant would start with the pictorial condition. This was done to prevent a systematic error that may occur due to exhaustion or habituation.

Procedure

Prior to the actual experiment, participants were asked to fill out a self-administered battery of questionnaires assessing emotional competencies and personality aspects (see Participants section above). The experiment started with a 2-min baseline recording of the EDA while participants looked at a gray screen while standing on the force plate. Next, they were instructed to carry out a test version of the experiment to familiarize themselves with the task. During the test version, emotional sequences appeared in the same order for all participants. These sequences were not shown in the main experiment. Subsequently, the actual experiment started. While standing on a force plate, two blocks of trials were shown, each containing all 48 emotional interactions presented in a pseudorandomized order. In one of the blocks, the trials were presented on a virtual screen (pictorial condition, Fig. 1a) and in the other they were presented in the virtual, open 3D space (visual condition, Fig. 1b).

Force plate data (i.e., implicit response behavior) were collected during the presentation of a sequence. Following each stimulus presentation, participants were instructed to make two explicit judgments (see Fig. 1c): first, they explicitly indicated their preferred behavioral response (i.e., explicit response behavior). More specifically, they were asked to observe the interactive avatars while standing in front of them and explicitly rate their tendency to approach or avoid the avatars displaying their particular emotional body expressions on an 11-point scale ranging from − 5 (avoid) to + 5 (approach) with 0 (neither) marking the center of the scale. Second, participants were asked to judge the valence of the interaction on a 11-point scale ranging from − 5 (extremely negative) to + 5 (extremely positive) with 0 (neutral) marking the center of the scale. After one half of each block, as well as at the end of each block, participants were asked to indicate how present they felt within the scene (‘How present did you feel within the scene?’) using a 7-point scale ranging from 1 (not at all) to 7 (extremely present).

Data analysis and statistics

Electrodermal activity

In the first step, the EDA signal was downsampled to 10 Hz and filtered using a fourth-order Butterworth low-pass filter with a cutoff frequency of 5 Hz. Afterwards, the signal was smoothed using a first-pass boxcar and second-pass Parzen window with a length of 150 samples. Skin conductance response (SCR) peaks were then determined according to Kim et al. (2004) and refined using a forward-ascending nearest-maximum search. An SCR was accepted as valid when above 0.02 μS and below 1 μS. To extract all SCRs, phasic component analysis was conducted using a trough-to-peak analysis within a 1–6 s delayed window after stimulus onset (Dawson et al. 2017). In the next step, a square-root transformation was applied to normalize the data. Finally, we calculated mean values for each emotional category (anger, sadness, affection, happiness) per display condition (pictorial, visual).

Center of pressure displacement

Force plate data were preprocessed and analyzed using MATLAB 2018a (MathWorks, Inc., Natick, MA). In the first step, data were low-pass filtered with a cutoff frequency of 8 Hz. For each trial, we calculated the mean COP displacement in the anteroposterior direction (COP-AP, in mm, corresponding to the COP position during the stimulus presentation relative to the COP position at the first frame of the stimulus). Next, we identified outliers and excluded trials from further analysis when the range of mediolateral COP displacement exceeded 8 cm, indicating movement of the feet. Finally, we calculated the postural responses for each emotional sequence within both display conditions per participant.

Presence

All ratings were preprocessed using MATLAB 2018a (MathWorks, Inc., Natick, MA). Statistical analyses were performed with IBM SPSS Statistics (version 25, IBM Cor, Armonk, NY). Mean values of subjective presence were calculated per participant and per display condition (i.e., pictorial, visual). A Wilcoxon signed-rank test was then conducted per group (i.e., males and females) to compare presence ratings within the pictorial and the visual condition. Following Rosenthal (1994), effect sizes r were calculated as Z statistic divided by square root of the sample size N with N being the number of total observations. To assess whether greater subjective presence was associated with higher autonomic arousal, we calculated Spearman correlation coefficients between subjective presence and skin conductance responses toward the emotional scenes.

Explicit and implicit measurements

We tested the influence of display condition and emotional category on the following dependent variables: (1) valence judgment, (2) explicit response behavior, (3) implicit response behavior. More specifically, three separate 2 (display condition: visual, pictorial) × 4 (emotion: anger, happiness, affection, sadness) × 2 (gender: male, female) repeated measures ANOVAs with the between-subject factor "participant gender" were conducted. With respect to implicit response behavior, we used COP mean displacement in the anterior–posterior direction as the dependent variable.

Relationship between valence judgment and response behavior

Finally, we aimed to explore whether the perceived valence of an emotional scene is correlated with the tendency to respond with approach or avoidance. To do so, we calculated correlation coefficients per participant for explicit judgments—that is, reported valence (positive/negative) and response behavior (avoid/approach). We did the same for valence judgment and implicit response behavior (COP displacement). In the next step, we applied a Fisher's Z transformation. We conducted one-sample t tests per group (male, female) and for each condition (pictorial, visual) to analyze whether the correlation was significant. Last, we conducted an independent-samples t test of the mean correlations of women versus men.

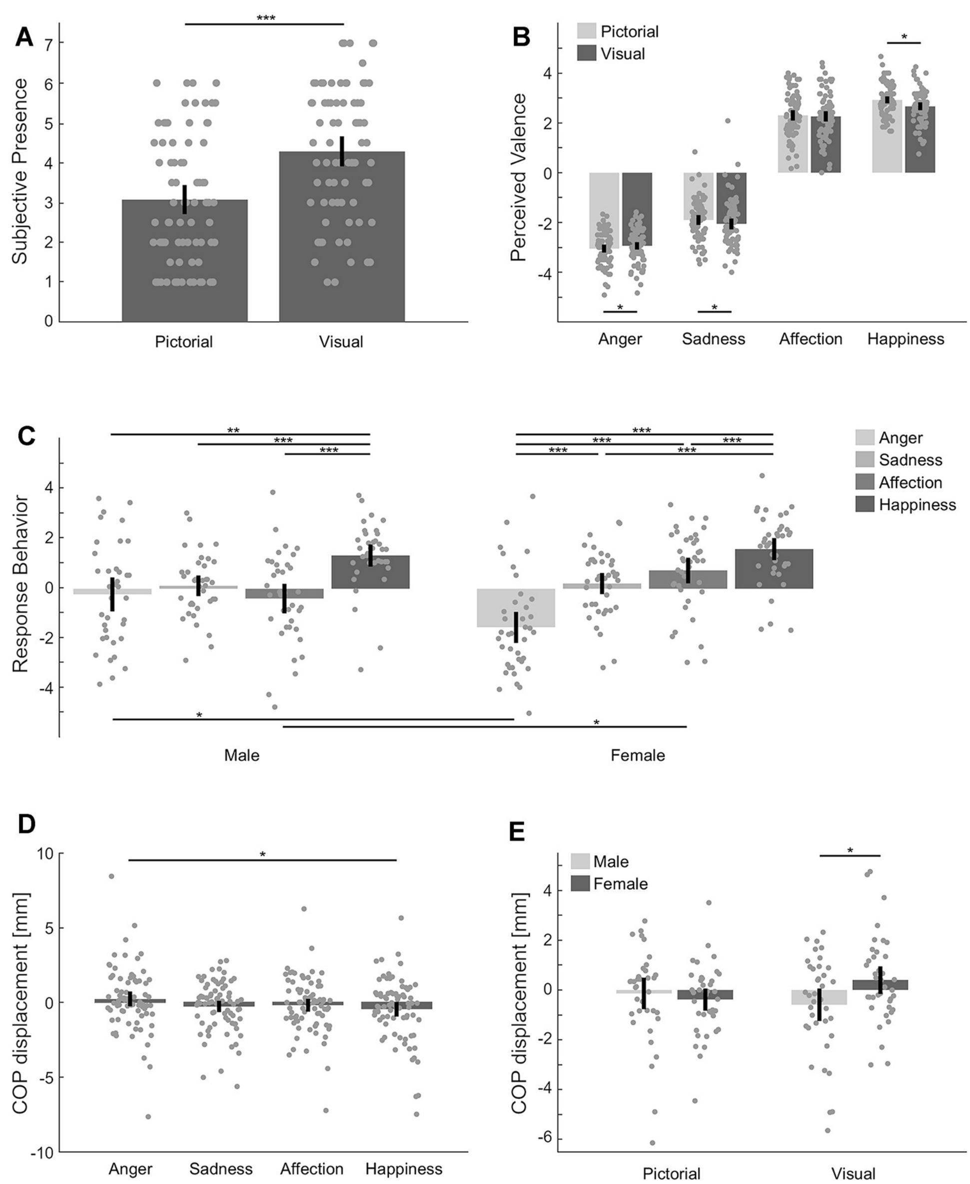

Fig. 2 Bars (including their standard errors and individual data points) showing **a** the mean subjective presence ratings per display condition (i.e., pictorial, visual) and **b** mean valence ratings per emotion separated by display condition (pictorial, visual) Only significant differences between the conditions are indicated. Post hoc tests indicated that within both conditions, all emotions differed significantly from each other (all $ps < .05$). **c** Mean explicit response behavior ratings for each emotion per group (i.e. male, female); note that significance bars above indicate differences between emotions for each gender, whereas significance bars below indicate differences between men and women for each emotion. **d** Mean COP displacement in anterior (i.e., positive values) and posterior (i.e., negative values) direction per emotion. **e** Mean COP displacement in anterior–posterior direction for each condition, separated by gender. Only significant differences between men and women are indicated. For men, COP displacement did not differ between the pictorial and visual condition ($p > 0.05$), whereas for women, COP displacement differed significantly between conditions ($p < 0.05$). Significance level is indicated by asterisks (*$p < 0.05$; ***$p < 0.001$)

Results

Subjective presence

First, we evaluated whether the perceived subjective presence differed between the two display conditions. Descriptive statistics revealed higher mean values of subjective presence in the visual condition ($M = 4.28$, $SD = 1.65$) compared to the pictorial condition ($M = 3.07$, $SD = 1.60$), see Fig. 2a. We applied a Wilcoxon signed-rank test separately for males and females. Data showed that both males ($Z = -4.98$, $p < 0.001$, $r = 0.58$) and females ($Z = -4.59$, $p < 0.001$, $r = 0.52$) felt significantly more present in the visual condition, confirming that subjective experience differed between the two conditions. To investigate whether increased subjective presence resulted in increased physiological responses toward the emotional sequences, we calculated Spearman correlation coefficients between the mean phasic electrodermal response and subjective presence, both per display condition (pictorial, visual). We found no significant correlations (all ps > 0.05), indicating that the subjective presence did not relate to the physiological responses elicited within either of the display conditions.

Does the display condition influence emotional valence perception?

In the next step, we calculated a 2 (display condition: visual, pictorial) × 4 (emotion: anger, sadness, affection, happiness) × 2 (gender: male, female) repeated measures ANOVA with valence as a dependent variable and gender as a between-subject factor. We found a main effect of display condition, $F(1, 74) = 6.94$, $p = 0.01$, $\eta_p^2 = 0.09$, indicating that, overall, stimuli were perceived slightly more negatively in the visual compared to the pictorial condition ($M_{diff} = -0.09$, $p = 0.01$). Furthermore, we found a main effect of emotion, $F(1, 74) = 1174.43$, $p < 0.001$, $\eta_p^2 = 0.94$. Post hoc analyses showed that all valence ratings differed significantly from each other (all ps < 0.001). As expected, descriptive statistics indicated negative mean valence ratings for anger and sadness, whereas affection and happiness sequences, on average, were rated positively, see Fig. 2b.

Further, we found a significant interaction between display condition and emotion category, $F(3, 222) = 7.76$, $p < 0.001$, $\eta_p^2 = 0.10$, indicating that the display condition did not affect valence perception of all emotions equally. Post hoc analyses revealed that anger ($M_{diff} = -0.12$, $p = 0.02$) was perceived slightly more negatively and happiness ($M_{diff} = 0.26$, $p < 0.01$) slightly more positively within the pictorial condition, whereas sadness was perceived more negatively within the visual condition ($M_{diff} = 0.16$, $p = 0.01$), see Fig. 2b. Overall, after Bonferroni corrections, these effects were only marginal. Affection was not perceived differently between display conditions ($p > 0.05$). This interaction implies that the observed main effect of viewing condition is mainly carried by the sequences that are displaying sadness and happiness. With regard to gender effects, mean valence ratings of each emotion did not differ between women and men, $F(3, 222) = 0.06$, $p > 0.05$, $\eta_p^2 = 0.001$. Moreover, the interaction between display condition and gender was not significant, $F(1, 74) = 0.04$, $p > 0.05$, $\eta_p^2 = 0.10$, indicating that the display condition did not affect valence perception of women and men differentially.

Does the display condition influence explicit response tendencies?

To explore the effects of the display condition and emotion category on reported response behavior, we calculated a 2 (display condition: visual, pictorial) × 4 (emotion: anger, sadness, affection, happiness) × 2 (gender: male, female) repeated measures ANOVA with explicit response behavior rating as a dependent variable and gender as a between-subject factor. We found no mediation by the display condition for this analysis as reflected by a nonsignificant main effect, $F(1, 74) = 0.64$, $p > 0.05$, $\eta_p^2 = 0.01$. We found a main effect of emotion, $F(3, 222) = 29.26$, $p < 0.001$, $\eta_p^2 = 0.28$, indicating that anger elicited greater avoidance tendencies than all other emotions (all ps < 0.05). In contrast, happiness elicited greater approach tendencies than all other emotions (all ps < 0.001). Sadness and affection did not elicit differential response tendencies ($p > 0.05$).

Furthermore, we found an interaction with the gender of a person ($F(3, 222) = 8.17$, $p < 0.001$, $\eta_p^2 = 0.09$). Post hoc analyses showed that for men, happiness elicited positive and significantly higher approach tendencies than all other emotions (all ps < 0.01). For women, we found the same effect as well as negative and significantly higher avoidance tendencies for anger sequences as compared to all other emotions (all ps < 0.001), see Fig. 2c.

Testing for gender-specific response tendencies with respect to different emotion categories, we found greater avoidance tendencies in response to anger sequences [$M_{diff} = -1.32$, $p < 0.01$, 95% CI (-2.24, -0.41)] as well as higher approach tendencies toward affection sequences in women as compared to men [$M_{diff} = 1.13$, $p < 0.01$, 95% CI (0.35, 1.92)], see also Fig. 2c. The display condition in which emotional interactions were presented, however, did not exert differential effects on men and women as reflected by a nonsignificant interaction, $F(1, 74) = 0.32$, $p > 0.05$, $\eta_p^2 = 0.004$.

Does the display condition affect implicit response tendencies?

To explore the effect of the display condition on automatic reaction tendencies toward the emotional scenes, we calculated a 2 (display condition: visual, pictorial) × 4 (emotion: anger, sadness, affection, happiness) × 2 (gender: male, female) repeated measures ANOVA with COP displacement in anterior–posterior direction as a dependent variable and gender as a between-subject factor. Results revealed no main effect of the display condition [$F(1, 74) = 0.61$, $p = 0.44$, $\eta_p^2 = 0.01$], but a main effect of emotion [$F(3, 222) = 3.05$, $p < 0.05$, $\eta_p^2 = 0.04$, see Fig. 2e]. More specifically, anger elicited significantly higher anterior movement than happiness scenes [$M_{diff} = 0.69$ mm, $p < 0.05$, 95% CI (0.004, 1.37)]. This effect was not mediated by a person's gender [$F(3, 222) = 0.98$, $p = 0.40$, $\eta_p^2 = 0.01$]. However, we did find an interaction effect between display condition and gender [$F(1, 74) = 8.54$, $p < 0.01$, $\eta_p^2 = 0.10$]. Post hoc analyses showed that women and men, on average, presented implicit avoidance behavior (i.e., increased posterior movement) in the pictorial condition. However, women displayed increased anterior movement within the visual condition, indicating approach behavior, whereas men, on average, displayed posterior movement ($M_{diff} = 0.99$ mm, $p < 0.05$, 95% CI (0.14, 1.85)], see Fig. 2f.

Is there a link between valence judgment and response behavior?

Due to the gender-mediated discrepancies between valence judgment and explicit response tendencies we observed in anger and affection sequences (compare Fig. 2b, c), we explored the phenomenon in more detail. Although men tended to judge anger sequences just as negatively as women did, about half of all male participants (i.e., 51.35% in the pictorial condition and 48.65% within the visual condition) indicated that they would like to approach the situation. Figure 3 illustrates how often specific valence and (explicit) response behavior rating combinations were observed for anger and affection sequences as well as within each condition (i.e., pictorial and visual).

For each participant as well as for both display conditions, we calculated correlation coefficients between explicit valence and response behavior ratings. With respect to anger sequences, correlation coefficients did not differ significantly from zero in either condition for the male population (all $ps > 0.05$). For women, the Bonferroni-corrected one-sample t tests indicated that correlation coefficients differed significantly from zero within the pictorial [$M = 0.42$, $t(37) = 3.34$, $p < 0.01$, $d = 0.54$, 95% CI (0.17, 0.68)], and the visual condition [$M = 0.36$, $t(37) = 2.75$, $p < 0.05$, $d = 0.45$, 95% CI (0.09, 0.62)].

We found the same effect for affection sequences. Whereas correlation coefficients found for men did not differ significantly from zero in either condition (all $ps > 0.05$), Bonferroni-corrected one-sample t tests indicated that

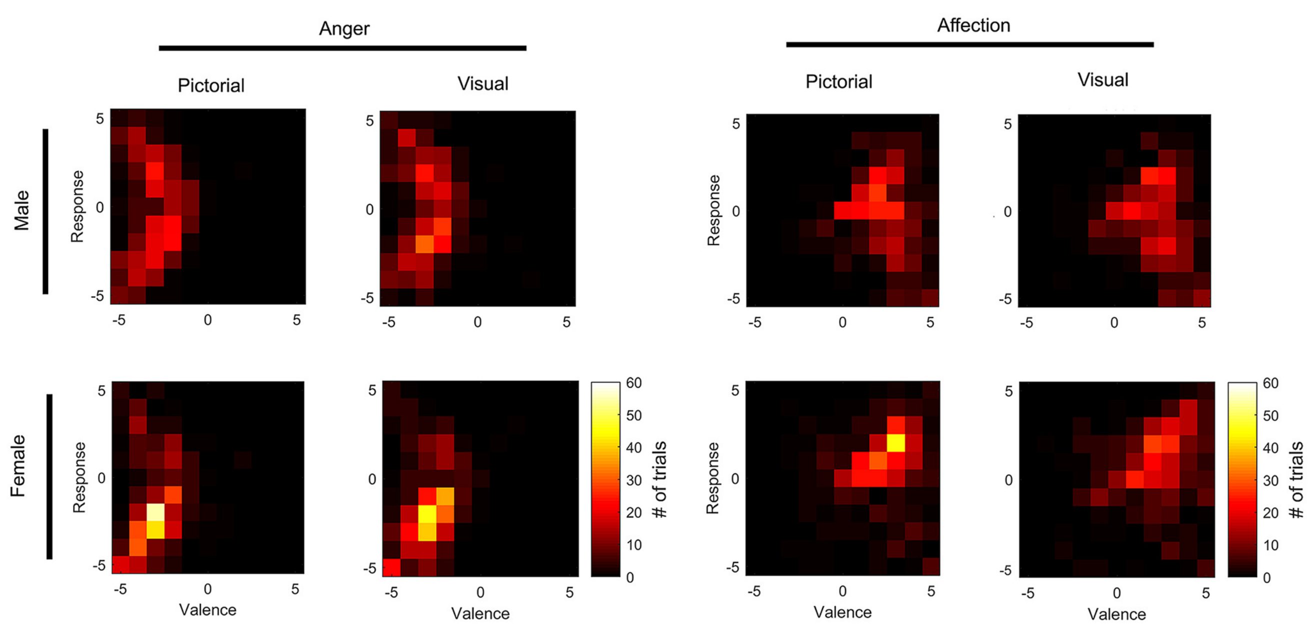

Fig. 3 Distribution of trials indicating how often specific valence × response behavior rating combinations were observed. All heat maps separated for **a** anger and **b** affection sequences in each space (pictorial, visual) and gender (female, male)

correlation coefficients for women differed significantly from zero within both the pictorial [$M = 0.44$, $t(38) = 3.97$, $p < 0.01$, $d = 0.64$, 95% CI (0.21, 0.66)], and the visual condition [$M = 0.38$, $t(36) = 3.29$, $p < 0.01$, $d = 0.54$, 95% CI (0.15, 0.61)]. Although women and men did not seem to differ in their judgment of emotional valence, there was a congruent relationship with reported response tendencies for women (i.e., avoidance in response to anger and approach in response to affection), whereas there seemed to be no relationship between these variables for men. Moreover, Bonferroni-corrected independent-samples t tests indicated that mean correlations between valence judgment and explicit response behavior rating for anger sequences differed significantly between women and men in the pictorial condition [$M = -0.58$, $t(72) = -3.12$, $p < 0.05$, $d = 0.73$, 95% CI (-0.95, -0.21)], but not in the visual condition [$t(73) = -2.31$, $p = 0.10$]. For affection, mean correlations between women and men did not differ in either of the display conditions [pictorial: $t(74) = -2.45$, $p = 0.07$, visual: $t(72) = -2.32$, $p = 0.09$].

With respect to valence judgment and implicit response tendencies (i.e., COP displacement), results showed that correlations were not significant for either women or men, indicating that there was no relationship between the explicit valence judgment and the automatic tendency to avoid or approach (all ps > 0.05).

Discussion

Using a multisystem approach, we aimed to shed light on how different virtual environments influence the perception of emotional body language and associated action tendencies. We did this by comparing two virtual display conditions: a pair of interacting partners was presented within either a pictorial condition in which participants viewed the emotional stimuli depicted on a screen or a visual condition in which participants shared the same virtual room with the interactive partners. We asked whether display conditions would affect explicit judgments (i.e., valence perception and response behavior) as well as implicit response behavior differentially. Overall, our results indicate that the visual condition in which the observer is co-present with other (virtual) agents elicited a stronger feeling of presence in the virtual world as indicated by subjective ratings. However, the display condition exerted only marginal effects on valence judgments and no effects on explicit response behavior. Implicit response behavior (i.e., COP displacement) was affected differentially by the display conditions—however, in a gender-specific manner. Our results showed that valence perception (i.e., the impression of how positively or negatively a scene is perceived) does not differ between women and men. In contrast, gender seems to mediate how individuals respond toward a given stimulus—both explicitly and implicitly. In the following sections, we shall discuss these findings in greater detail.

Valence perception

With respect to valence perception, the display condition exerted only marginal effects on how positively or negatively an emotional interaction was perceived depending on the emotional category presented. More specifically, anger and happiness were judged to be more intense in the pictorial condition, meaning that anger was perceived more negatively and happiness was perceived more positively. Sadness, in contrast, was judged more negatively in the visual condition. Affection was not perceived differently between display conditions.

Prior studies have demonstrated that immersive virtual environments (i.e., an environment in which the participant is in “presence mode”) intensify the emotional response of the viewer as measured by the subjective experience of arousal. This may lead, in turn, to an intensified perception of the content being viewed (Estupiñán et al. 2014; Visch et al. 2010). However, our data revealed that this is not necessarily the case. Instead, our results point toward a differential effect: sadness was judged to be only slightly negative, whereas anger and happiness were judged to contain greater emotional valence. Whereas sadness sequences were rated to contain less emotional valence, happiness and anger were rated as being more intense. Thus, whether a stimulus is perceived more or less intensely within a more immersive environment may depend on its valence intensity. Slater (2009) has discussed this phenomenon in terms of realistic physiological, emotional, and behavioral responses due to place illusion (i.e., a sense of being present in the virtual environment despite knowing at a higher cognitive level that the situation is not real). According to his view, more immersive environments may change our conscious cognitive perception into a more realistic, instead of a generally more intense experience. Note that valence judgments, overall, differed only slightly between the pictorial and the visual condition in our study, suggesting that the conscious evaluation of an observed emotional interaction does not seem to be influenced that strongly by the display condition. This lack of a relevant difference in valence perception is also in line with the results of a pilot study carried out by Estupiñán et al. (2014), who found that mean valence values for images of human concerns (i.e., scenes violating human rights) displayed within a VR head-mounted display did not differ significantly from the reference of the Geneva Affective Picture Database (GAPED) displayed on a computer screen.

Turning to the lack of relevant differences in valence judgements, it has to be noted that the increased sense of subjective presence was not accompanied by increased physiological arousal (i.e., increased skin conductance response) in the present study. However, previous studies have proposed that greater presence evokes greater autonomic responses (e.g., increased heart rate measures and, to a lesser degree, increased skin conductance response; see Meehan et al. 2002), especially when participants were exposed to stressful virtual environments. Freeman et al. (2005) pointed out that presence and emotion perception are related only for arousing stimuli, and that increased arousal can lead to an intensified perception of the emotional content being viewed. Hence, it seems plausible that the perception of emotional valence within highly immersive environments may be altered especially when stressful or highly arousing stimuli are displayed (Diemer et al. 2015; Freeman et al. 2005; Visch et al. 2010). In the present study, however, the stimulus material was not created to induce stress or other emotional distress in the observer, but rather to reflect real-world emotional interactions that would allow us to investigate the perception of affective states in others. Thus, it may be possible that changes in valence perception between the display conditions may occur when highly arousing stimuli are used.

Turning to the influence of gender on valence perception, we found that women and men did not differ. Our data indicate no gender differences in valence judgments for either more intense emotions such as happiness and anger or less intense emotions such as sadness and affection. Although many studies have suggested a female superiority in emotion recognition (e.g., Alaerts et al. 2011; Hall and Matsumoto 2004; Hampson et al. 2006), other studies have found that gender effects occur only when highly subtle emotions are presented compared to more intense, prototypical emotional displays (Hoffmann et al. 2010; Montagne et al. 2005). The lack of agreement on gender differences between the current and prior studies may be due to methodological differences. Whereas most studies assessed performance (i.e., accuracy) or reaction times when assigning emotional displays to their respective category (e.g., anger, happiness), we assessed how strongly the valence of a given stimulus is perceived for different emotional displays (Alaerts et al. 2011; Hampson et al. 2006) and, therefore, focused on a more subjective dimension of perceiving emotional body language.

Response behavior

Explicit response behavior

We found that explicit response behavior is not influenced differentially by the display condition, meaning that the urge to approach or avoid an emotional scene is virtually the same. Similar to valence judgments, explicit response behavior is thought to represent complex cognitive evaluations (in contrast to automatic response behavior) that may be shaped by individual factors, such as personality traits. Although we can only speculate about why co-presence does not affect explicit response tendencies, it seems plausible that participants reflected actively on how they would react if this scenario were to occur in a "real-life" situation. Applying such a strategy could certainly dampen the experience and break the illusion of being inside the virtual world despite knowing that it is not real (Slater 2009). As a result, the cognitive evaluation to avoid or approach the situation will not be affected differently by virtual co-presence.

Interestingly, we found that explicit response behavior is mediated by the observer's gender. Whereas women tend to report greater avoidance tendencies toward anger than men, they also report greater approach tendencies toward affection than men. Explicit response tendencies represent cognitive evaluations of a stimulus which may be affected by social learning and reinforcement. It is assumed that men learn to hide helplessness and express aggression related emotions such as anger more openly, whereas women are reinforced when showing helplessness and suppressing anger (Fischer 1993). Thus, with respect to anger sequences, it seems plausible that women tend to respond with avoidance while men respond with approach. However, there are notable within group inconsistencies in the tendency to avoid or approach. Thus, the impact of different personality traits on helping behavior e.g., altruism (Wang and Wang 2008) and agreeableness (Graziano et al. 2007), might further elucidate those inconsistencies. Moreover, Seidel et al. (2010b) suggest that the gender of the interacting agents might also be an influencing factor on conscious as well as automatic behavioral tendencies. The authors suggest that social learning possibly turned male faces expressing negative emotions to salient social cues, which communicate the message to respond with avoidance.

Implicit response behavior

Whereas explicit response behavior has been largely neglected in the literature, implicit response behavior has been studied quite extensively (e.g., Bradley et al. 2001; Hillman et al. 2004; Horslen and Carpenter 2011; Seidel et al. 2010b). We found that automatic response tendencies depend on the emotion being displayed. On average, anger elicited greater forward movement than happiness. Although this may initially seem counterintuitive, because anger has also been shown to elicit avoidance tendencies (e.g., Chen and Bargh 1999; Marsh et al. 2005), other studies suggest that humans can be motivated to approach, confront, and overcome the social challenge posed by angry expressions (Carver and Harmon-Jones 2009; Wilkowski and Meier

2010). This is in line with findings by Stins and Beek (2007) who reported that unpleasant images elicit a forward COP displacement (of about 1 mm). It is noteworthy, however, that we observed high variances in the implicit behavior to approach or avoid. This indicates that other factors such as learning experiences or certain personality factors may play a mediating role in the behavioral tendency to respond.

Moreover, we found an interaction between the display condition and the gender of a person when it came to implicit responses. Whereas women and men generally tended to display backward movements in the pictorial condition, women displayed approach tendencies in the visual condition. This is an interesting finding, because it indicates that the display condition may indeed differentially affect automatic responses toward emotional stimuli—at least in women.

However, it should be pointed out that, overall, the mean displacement is extremely small (less than 1 mm) and should therefore not be overinterpreted. Interestingly this “lack” of implicit behavioral tendencies toward emotional displays has also been reported by Seidel et al. (2010a), who showed that only depressive patients—but not healthy controls—displayed implicit avoidance behavior toward angry faces. Moreover, healthy controls also did not display significant behavioral tendencies toward happy faces. This is in line with findings by Stins and Beek (2007) reporting that viewing emotion eliciting pictures had little effect on body sway. The authors suggested that passive viewing might be coupled only weakly to posture, and that an effective way to probe the emotion-posture system would be to induce emotional states that are relevant for postural control, such as anxiety.

The influence of perspective

With respect to both, implicit and explicit response tendencies, an important aspect that should be highlighted is the influence of viewing perspective. Within both conditions, participants remain in an observing position, or third-person perspective, in which the observed actions are not directed at the viewer. That means, the viewer passively observes the actors’ interaction. An actor, or first-person perspective, in contrast, is created when the viewer is transported into the perspective of an actor. Research shows that spatial presence is increased in the actor perspective as compared to the observer perspective (van den Boom et al. 2015). Although speculative, it may be possible that a condition in which the viewer is put in an actor perspective, in which the observed actions are directed towards the participant, stronger approach- or avoidance tendencies could be observed.

The link between valence and explicit response behavior

Finally, we conducted an exploratory analysis of gender-mediated discrepancies between valence judgment and associated response tendencies toward anger and affection sequences. Although some studies have reported that negatively valenced stimuli such as angry faces signal the request to go away (i.e., avoidance), others have demonstrated that a negative evaluation of stimuli does not necessarily evoke avoidance behavior (Carver and Harmon-Jones 2009; Horstmann 2003; Seidel et al. 2010b; Wilkowski and Meier 2010). Wilkowski and Meier (2010), for instance, have argued that angry facial expressions communicate the intention to confront a person aggressively. Thus, they pose an important social challenge that should predispose individuals to engage in approach-motivated behavior to confront or overcome them.

Our data showed that while there is a positive link between valence judgment and explicit response tendency toward anger in women (i.e., negative valence judgment and the behavioral tendency to avoid), this link is absent in men. This is due to a high variability in explicit response behavior toward anger within the male population. Whereas some participants indicated they would like to approach the situation, others reported feeling the tendency to withdraw. Although a person’s gender does seem to play a role in the relationship between valence and explicit response behavior for anger, it is not the sole contributor. Personality models, for instance, suggest that certain traits foster or inhibit response behavior (see Carver 2005, for a review). We consider these to be interesting results, because they highlight the complexity of the phenomenon.

Conclusion

The present study showed that virtual environments in which the observer shares a common space with virtual agents increase the observer’s subjectively reported presence when observing emotional interactions. It further showed that the cognitive evaluation of stimulus valence is largely independent of the display condition (i.e., visual vs. pictorial) and that it does not differ between women and men. These findings support the notion that humans are highly adept at recognizing emotional stimuli from emotional body language, even within more abstract settings such as pictorial designs.

Moreover, we did not observe differential effects of the display condition on explicit response behavior. However, we did observe a gender-mediated effect of display condition on implicit response behavior. Although the mechanism is still unclear, it seems that being inside a “visual” space may act more strongly on automatic motivational tendencies but not on higher level cognitive evaluations. We speculated that this might be due to a break in place illusion caused by active cognition evaluation of the scene.

Interestingly, a person's gender also mediated explicit response behavior. Women indicated a tendency to explicitly avoid anger scenes and approach affection scenes more strongly than men, suggesting that they react with a greater defensive motivation toward anger as well as greater appetitive motivation toward affection than men. Interindividual differences in explicit rating behavior indicated that a proportion of the male participants explicitly wanted to approach anger scenes. Perhaps this was due to a motivation to confront or overcome a social challenge posed by angry expressions (Carver and Harmon-Jones 2009; Wilkowski and Meier 2010). Moreover, we found that the display condition did not affect men and women differently with respect to valence or explicit response behavior ratings.

Finally, we discovered a closed link between valence perception and congruent explicit response behavior in women. However, this link was absent in men.

Limitations and future implications

Virtual reality offers many options as well as challenges. Whereas it can serve as a way to manipulate experimental settings in a highly controlled manner, its complexity and the effects it has on human perception are not well understood. Here, we applied a virtual reality paradigm comparing a "pictorial" condition (i.e., a space that simulates a computer experiment on a computer screen) to a "visual" condition (i.e., a condition with a shared space between the observer and virtual agents). Although this allowed us to modify the subjective experience of presence within the virtual environment, we did not exhaust its full potential. More specifically, while the difference in subjective presence can be assessed quantitatively, it remains open whether it is a meaningful difference. Across both conditions, the viewer remains in an observer perspective and is not able to experience the scene from an actor perspective or influence the course of the emotional interaction. Recent studies have proposed that the environment is thought to be understood and perceived through our interaction with it. In other words, perceptual experience is no longer thought to result from passive information processing, but is "enacted" via regulation of sensorimotor loops and active exploration of the environment (Engel et al. 2013; Froese et al. 2014). Thus, future studies should tackle the role of an actor perspective, potentially within an interactive VR setting. In this regard, response behavior tendencies, such as approach and avoidance, can be measured by instructing the participant to physically walk towards (approach) or away from a scene (avoid). Thereby, the rating continuum could be replaced with a more ecologically valid measure.

Moreover, additional cardiovascular measures would be suited to assess possible mediator effects that have been found to influence emotion perception within VR settings, such as arousal (Diemer et al. 2015). Here, we used EDA as an implicit autonomic measure, however, Meehan et al. (2002) suggest that greater autonomic responses due to greater presence is reflected predominantly in increased heart rate and, to a lesser degree, in skin conductance.

Another important factor that should be considered is the restricted nature of our sample. Especially when making group comparisons, in which the sample is split in half (i.e. male, female), the results do not necessarily allow for generalizations that apply to other cohorts. Especially with regard to older populations. A meta-analytic review has concluded that, overall, older adults are less accurate than young adults at recognizing emotions (Ruffman et al. 2008). However, one has to consider that many tasks used static, stereotyped photographs of emotional expression, limiting ecological validity of the task (Barrett et al. 2011). In the present study, we chose to implement a setting that induces a sense of presence in the observer, making it more similar to real-life encounters with others. However, the highly homogenous sample warrant caution when generalizing the results to wider context.

With respect to gender differences, pronounced intragroup differences in explicit as well as implicit response behavior suggest that there are individual factors that may play a more critical role in the cognitive mediation of response behavior than mere gender. Thus, future studies should aim at considering the role of individual factors, such as personality, in emotion perception studies.

Author contributions Conceived and designed the experiments: JB, AZ, JK, JM, BK, NT. Performed the experiments and analyzed the data: JB, AZ, SG. Discussed the results and wrote the paper: JB, AZ, JM, BK.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by DFG IRTG 1901 "The Brain in Action" under a grant for JB. NFT's contributions to this project were funded by a NSERC Discovery grant and by the CFREF program "Vision: Science to Application."

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability statement The data that support the findings of this study are available from the corresponding author JB upon reasonable request.

References

Alaerts K, Nackaerts E, Meyns P, Swinnen SP, Wenderoth N (2011) Action and emotion recognition from point light displays: an investigation of gender differences. PLoS ONE 6(6):e20989. https://doi.org/10.1371/journal.pone.0020989

Atkinson AP, Dittrich WH, Gemmell AJ, Young AW (2004) Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33(6):717–746. https://doi.org/10.1068/p5096

Azevedo TM, Volchan E, Imbiriba LA, Rodrigues EC et al (2005) A freezing-like posture to pictures of mutilation. Psychophysiology 42(3):255–260. https://doi.org/10.1111/j.1469-8986.2005.00287.x

Barrett LF, Mesquita B, Gendron M (2011) Context in emotion perception. Curr Dir Psychol Sci 20(5):286–290. https://doi.org/10.1177/0963721411422522

Beck AT, Steer RA, Brown GK (1996) Beck Depression Inventory-II. Psychological Corporation, San Antonio

Bradley MM, Codispoti M, Cuthbert BN, Lang PJ (2001) Emotion and motivation I: defensive and appetitive reactions in picture processing. Emotion 1(3):276–298. https://doi.org/10.1037/1528-3542.1.3.276

Carver CS (2005) Impulse and constraint: perspectives from personality psychology, convergence with theory in other areas, and potential for integration. Pers Soc Psychol Rev 9(4):312–333

Carver CS, Harmon-Jones E (2009) Anger is an approach-related affect: evidence and implications. Psychol Bull 135:183–204

Chen M, Bargh JA (1999) Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. Pers Soc Psychol Rev 25(2):215–224. https://doi.org/10.1177/0146167299025002007

Clarke TJ, Bradshaw MF, Field DT, Hampson SE, Rose D (2005) The perception of emotion from body movement in point-light displays of interpersonal dialogue. Perception 34(10):1171–1180. https://doi.org/10.1068/p5203

Dawson ME, Schell AM, Filion DL (2017) The electrodermal system. In: Cacioppo JT, Tassinary LG, Berntson GG (eds) Cambridge handbooks in psychology. Handbook of psychophysiology. Cambridge University Press, Cambridge, pp 217–243

de Gelder B (2006) Towards the neurobiology of emotional body language. Nat Rev Neurosci 7:242–249. https://doi.org/10.1038/nrn1872

de Gelder B, De Borst AW, Watson R (2015) The perception of emotion in body expressions. Wiley Interdiscip Rev Cogn Sci 6(2):149–158. https://doi.org/10.1002/wcs.1335

Diemer J, Alpers GW, Peperkorn HM, Shiban Y, Mühlberger A (2015) The impact of perception and presence on emotional reactions: a review of research in virtual reality. Front Psychol 6:26. https://doi.org/10.3389/fpsyg.2015.00026

Elliot AJ, Covington MV (2001) Approach and avoidance motivation. Educ Psychol Rev 13:73–92. https://doi.org/10.1023/A:1009009018235

Engel AK, Maye A, Kurthen M, König P (2013) Where's the action? The pragmatic turn in cognitive science. Trends Cogn Sci 17(5):202–209. https://doi.org/10.1016/j.tics.2013.03.006

Estupiñán S, Rebelo F, Noriega P, Ferreira C, Duarte E (2014) Can virtual reality increase emotional responses (Arousal and Valence)? A pilot study. In: Marcus A (ed) Design, user experience, and usability. User experience design for diverse interaction platforms and environments. Springer, New York, pp 541–549

Facchinetti LD, Imbiriba LA, Azevedo TM, Vargas CD, Volchan E (2006) Postural modulation induced by pictures depicting prosocial or dangerous contexts. Neurosci Lett 410(1):52–56. https://doi.org/10.1016/j.neulet.2006.09.063

Fischer AH (1993) Sex differences in emotionality: fact or stereotype? Fem Psychol 3(3):303–318. https://doi.org/10.1177/0959353593033002

Freeman J, Lessiter J, Pugh K, Keogh E (2005) When presence and emotion are related, and when they are not. In: 8th annual international workshop on presence, September, pp 213–219

Froese T, Iizuka H, Ikegami T (2014) Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. Sci Rep 4:3672. https://doi.org/10.1038/srep03672

Graziano WG, Habashi MM, Sheese BE, Tobin RM (2007) Agreeableness, empathy, and helping: a person × situation perspective. J Pers Soc Psychol 93(4):583–599. https://doi.org/10.1037/0022-3514.93.4.583

Hall JA, Matsumoto D (2004) Gender differences in judgments of multiple emotions from facial expressions. Emotion 4(2):201–206. https://doi.org/10.1037/1528-3542.4.2.201

Hampson E, van Anders SM, Mullin LI (2006) A female advantage in the recognition of emotional facial expressions: test of an evolutionary hypothesis. Evol Hum Behav 27(6):401–416. https://doi.org/10.1016/j.evolhumbehav.2006.05.002

Hillman CH, Rosengren KS, Smith DP (2004) Emotion and motivated behavior: postural adjustments to affective picture viewing. Biol Psychol 66(1):51–62. https://doi.org/10.1016/j.biopsycho.2003.07.005

Hoffmann H, Kessler H, Eppel T, Rukavina S, Traue HC (2010) Expression intensity, gender and facial emotion recognition: women recognize only subtle facial emotions better than men. Acta Psychol 135(3):278–283. https://doi.org/10.1016/j.actpsy.2010.07.012

Horstmann G (2003) What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? Emotion 3(2):150–166. https://doi.org/10.1037/1528-3542.3.2.150

Horslen BC, Carpenter MG (2011) Arousal, valence and their relative effects on postural control. Exp Brain Res 215(1):27–34. https://doi.org/10.1007/s00221-011-2867-9

Kaletsch M, Pilgramm S, Bischoff M, Kindermann S, Sauerbier I, Stark R, Lis S, Gallhofer B, Sammer G, Zentgraf K, Munzert J, Lorey B (2014b) Major depressive disorder alters perception of emotional body movements. Front Psychiatry 5:4. https://doi.org/10.3389/fpsyt.2014.00004

Kaletsch M, Krüger B, Pilgramm S, Stark R, Lis S, Gallhofer B, Zentgraf K, Munzert J, Sammer G (2014a) Borderline personality disorder is associated with lower confidence in perception of emotional body movements. Front Psychol. https://doi.org/10.3389/fpsyg.2014.01262

Kim KH, Bang SW, Kim SR (2004) Emotion recognition system using short-term monitoring of physiological signals. Med Biol Eng Comput 42(3):419–427. https://doi.org/10.1007/BF02344719

Koenderink J, van Doorn A (2012) Gauge fields in pictorial space. SIAM J Imaging Sci 5:1213–1233. https://doi.org/10.1137/120861151

Krüger B, Kaletsch M, Pilgramm S, Schwippert SS, Hennig J, Stark R, Munzert J (2018) Perceived intensity of emotional point-light displays is reduced in subjects with ASD. J Autism Dev Disord 48(1):1–11. https://doi.org/10.1007/s10803-017-3286-y

LeDoux JE (1996) The emotional brain. Simon & Schuster, New York

Loper M, Mahmood N, Black MJ (2014) MoSh: motion and shape capture from sparse markers. ACM T Graphic (TOG) 33(6):1–13. https://doi.org/10.1145/2661229.2661273

Lorey B, Kaletsch M, Pilgramm S, Bischoff M, Kindermann S, Sauerbier I, Stark R, Zentgraf K, Munzert J (2012) Confidence in emotion perception in point-light displays varies with the ability to perceive own emotions. PLoS ONE 7(8):e42169. https://doi.org/10.1371/journal.pone.0042169

Marsh AA, Ambady N, Kleck RE (2005) The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. Emotion 5(1):119–124. https://doi.org/10.1037/1528-3542.5.1.119

Meehan M, Insko B, Whitton M, Brooks FP Jr (2002) Physiological measures of presence in stressful virtual environments. ACM T Graphics (TOG) 21(3):645–652. https://doi.org/10.1145/566654.566630

Montagne B, Kessels RP, Frigerio E, de Haan EH, Perrett DI (2005) Sex differences in the perception of affective facial expressions: do men really lack emotional sensitivity? Cogn Process 6(2):136–141. https://doi.org/10.1007/s10339-005-0050-6

Rindermann H (2009) Emotionale-Kompetenz-Fragebogen: EKF; Einschätzung emotionaler Kompetenzen und emotionaler Intelligenz aus Selbst-und Fremdsicht. Hogrefe, Göttingen

Rosenthal R (1994) Parametric measures of effect size. In: Cooper H, Hedges LV (eds) The hand-book of research synthesis. Sage, New York, pp 231–244

Ruffman T, Henry JD, Livingstone V, Phillips LH (2008) A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of aging. Neurosci Biobehav Rev 32(4):863–881. https://doi.org/10.1016/j.neubiorev.2008.01.001

Seidel EM, Habel U, Finkelmeyer A, Schneider F, Gur RC, Derntl B (2010a) Implicit and explicit behavioral tendencies in male and female depression. Psychiatry Res 177(1–2):124–130

Seidel EM, Habel U, Kirschner M, Gur RC, Derntl B (2010b) The impact of facial emotional expressions on behavioral tendencies in women and men. J Exp Psychol Hum Percept Perform 36(2):500–507

Slater M (2003) A note on presence terminology. Presence Connect 3(3):1–5

Slater M (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol 364(1535):3549–3557. https://doi.org/10.1098/rstb.2009.0138

Spielberger CD, Gorsuch RL, Lushene RE, Vagg PR, Jacobs GA (1983) Manual for the State-Trait Anxiety Inventory. Consulting Psychologist Press, Palo Alto

Stins JF, Beek PJ (2007) Effects of affective picture viewing on postural control. BMC Neurosci 8:83–90. https://doi.org/10.1186/1471-2202-8-83

Stins JF, Roelofs K, Villan J, Kooijman K, Hagenaars MA, Beek PJ (2011) Walk to me when I smile, step back when I'm angry: emotional faces modulate whole-body approach—avoidance behaviors. Exp Brain Res 212(4):603–611. https://doi.org/10.1007/s00221-011-2767-z

Thayer J, Johnsen BH (2000) Sex differences in judgement of facial affect: a multivariate analysis of recognition errors. Scand J Psychol 41(3):243–246. https://doi.org/10.1111/1467-9450.00193

Troje NF (2019) Reality check. Perception 48(11):1033–1038. https://doi.org/10.1177/0301006619879062

van den Boom AA, Stupar-Rutenfrans S, Bastiaens OS, van Gisbergen MM (2015) Observe or participate: the effect of point-of-view on presence and enjoyment. In: 360 degree movies for head-mounted displays. AmI-2015 adjunct proceeding. Retrieved from: http://ceur-ws.org/Vol-1528/paper13.pdf

Visch VT, Tan ES, Molenaar D (2010) The emotional and cognitive effect of immersion in film viewing. Cogn Emot 24(8):1439–1445. https://doi.org/10.1080/02699930903498186

Wang CC, Wang CH (2008) Helping others in online games: prosocial behavior in cyberspace. Cyber Psychol Behav 11(3):344–346. https://doi.org/10.1089/cpb.2007.0045

Wilkowski BM, Meier BP (2010) Bring it on: angry facial expressions potentiate approach-motivated motor behavior. J Pers Soc Psychol 98(2):201–210. https://doi.org/10.1037/a0017992

6 General discussion

In the past decade, emotion recognition research has expanded its view, using bodily displays of emotion to identify factors that modulate our perception. Different factors have been established, including available kinematic and postural information, the amount of information that is available overall (e.g. monads vs dyads) as well as individual characteristics of the observer (e.g. their age, mood, gender). Although studies have found reliable emotion recognition from single-agent displays of emotion (e.g. Atkinson et al., 2004; Alaerts et al., 2004), others have highlighted the importance of social context information as it is found, for instance, within interpersonal emotional dialogue (Clarke et al., 2005; Lorey et al, 2012). Thus, we further extended the exisiting literature using point-light displays of dynamic emotional interactions to deepen our understanding of factors that drive and modulate emotion recognition from bodies. By taking a multisystem approach, it is explored how the perception of affective states is shaped by the type information that is available as well as the coordination of kinematic information and examines the role of modulatory individual factors. In addition to commonly used paradigms, a new virtual reality method has been developed to study the impact of presence on emotion recognition. As outlined previously, investigating these mechanisms has great potential to advance the understanding of how it is possible to infer mental states by observing others and paves the way towards understanding why in some disorders, this ability is compromised (Kaletsch et al., 2014; Dalili et al. 2015; Krüger et al., 2018).

6.1 Key body parts that allow for emotion recognition

As outlined previously, emotion recognition research has demonstrated repeatedly that humans have little difficulty in decoding relevant cues from body movement. One core question associated with this ability is what parts of the body carry the most salient information about the affective state.

Using an occlusion paradigm, in which either arm or trunk movements were isolated and presented as PLDs, we confirmed that emotional content can be readily recognized from dyadic interactions. Regardless of the type of stimulus that was displayed, recognition rates were well above chance level, indicating that individuals are able to identify relevant cues even when presented with highly impoverished information. However, we found an emotion specificity with respect to recognition accuracy from either arm or trunk movements. More precisely, we found that happiness and anger were recognized more reliably from arm movements as compared to trunk movements. Sadness, however, seemed to be perceived mainly via cues from the head and torso. With respect to affection, we found no differences in recognition rates for body and arm movements.

These findings are in line with previous research. Gross et al. (2012), for instance, found that body posture during walking is significantly different between emotions. Sad walkers, for instance, had a more flexed neck and thorax than joyful walkers. They also found that sad walking was also accompanied by diminished arm swing. Similar observations were made by Michalak et al. (2009), who found that sad gait is characterized by reduced arm swing, vertical head movements, larger lateral swaying movements and a more slumped posture. These observations have been replicated many times, with the most common kinematic features including a forward head bend, forward chest bend, little arm movement and very little overall movement (for an overview, see Kleinsmith and Bianchi-Berthouze, 2013). As the main changes in posture and movement seem to be focused around the trunk and head, it seems plausible that perceivers benefit significantly more from a display of these features as compared to arm movements. While the lack of arm movement may indicate that the mood state could be classified as sadness, a slumped posture allows for a more accurate detection of the respective emotion. However, it should be noted that both displays are sufficient for the observer to retrieve the target emotion (sadness).

Anger and happiness, in contrast, have been characterized as high energy, fast movements with large, jerky movements and expanded limbs (see Kleinsmith and Bianchi-Berthouze, 2013). Thus, as our results indicate, arms, hands and shoulders seem to be the key parts of the body that are used to form the perceptual impression of anger and happiness. For anger especially, we found a strong advantage for arm displays with respect to emotion recognition performances.

Taken together, it can be outlined that the bodily expression of emotion in dyadic interactions is visible in the whole body and can be viewed as a complex system carrying salient information. However, it becomes apparent that certain body parts carry key information for different affective states. While the most salient information about sadness is retrieved from body and head movements, anger and happiness cues can easily be retrieved from arm movements. Affection seems to be represented equally in arm vs. trunk movements.

6.2 The role of (spatio)temporal coupling

Aside from purely kinematic and postural information of each individual, interactions further allow for a retrieval of interpersonal information. Interpersonal information may include proxemic measures such as distance or orientation but also the timing and coordination of each of the agent's actions (Moreau et al., 2016, Yokozuka et al., 2018, Lahnakoski et al., 2020; Manera 2011; Neri, 2006; Keck et al., 2022).

Previous research has demonstrated that when interactive activity that requires close body contact is observed (e.g., fighting or dancing), the human visual system relies on the spatiotemporal coupling between the agents to retrieve information relating to each agent individually (Neri et al., 2006). More specifically, within meaningfully synchronized interactions, the actions of one agent serve as the predictor for the expected actions of the other agent. Thus, within the second study, the existing literature was expanded and it was determined whether the spatiotemporal coupling between two interacting agents is necessary for the recognition of affective states. To do so, we manipulated the existing point-light interactions by delaying the movement onset of one of the agents and investigated whether the recognition will be impaired as a result of this manipulation.

Our results indicate that the recognition of emotions from dyadic interactions depends on a fixed (spatio)temporal coupling of the actions: temporal decoupling of one of the agent's actions within an emotional interaction leads to a *decrease* in recognition accuracy. However, this effect occured in an emotion specific manner and only holds true for anger and affection. The recognition of happiness and sadness were not affected by temporal decoupling.

These observations are partially in line with findings by Clarke and colleagues (2005) who presented monads (i.e. single agents) taken from dyadic interactions and found that romantic love and joy are recognized more reliably when they are depicted within an interaction. Although our results confirm this effect for romantic love, or affection in our case, we did not find this true for happiness. However, one has to take the methodological differences of both studies into account. Clarke et al. (2005) presented monads (i.e. single agents) taken from the dyadic interactions. In our case, both agents were depicted, but with different temporal offsets. As a result, the observer had constant access to movement information from the second agent, which may have benefited recognition performances.

However, the question arises why there is an emotion specificity. Although there is no definite answer to this question yet, it has been argued that some emotions may have a stronger interpersonal and communicative character than others and thus depend more on reciprocal actions whereas others may be acted out more on an individual level (Clarke et al., 2005).

In summary, we found that the perception of angry and affective interactions is dependent on a (spatio)temporal coupling between the agents, highlighting the notion that not only intrapersonal parameters but also interpersonal parameters play a central role in emotion perception (Keck et al., 2022).

6.3 The influence of (co-)presence

In the third study**,** a novel multi-system virtual reality research design was implemented to investigate whether the perceptual impression of emotional interactions is shaped by the perceived presence of the interacting agents. Most of the paradigms that have been established in emotion recognition research are based on so-called ‚pictorial' designs. That is, static or dynamic sequences are displayed via screens (e.g. Atkinson et al., 2004; Alaerts et al., 2011; Ross et al., 2012). However, in classical desktop designs the observer is physically separated from the scene and thus cannot experience the presence of the interacting agents. Research suggests that when relying on classical pictorial designs, we are investigating the perceptual system in picture mode, not in presence mode, ultimately leading to differences with respect to how it is processed on a neural level (Nanay, 2011, Troje, 2019). Being able to experience another's presence, for instance, has been shown to be strongly correlated with a sense of agency, i.e. the ability to perceive and change the environment, and thus may be crucial part of social cognition (Lallart et al., 2009; Jicol et al., 2021).

Thus, using a multisystem approach, we created a perceptual experience that enabled the participant to have an increased a sense of presence within the virtual space. In this virtual environment, the observer shares a common space with the interacting agents, enabling a subjectively perceived co-presence in which perceptual and physiological responses are expected to be similar to real-life situations (Slater, 2003; Troje, 2019).

Indeed, the comparison of picture mode and presence mode revealed that the shared space (i.e. "presence mode") led to an increase in subjectively perceived presence. However, although we successfully elicited an increased sense of presence, as confirmed by participant ratings, we observed only a marginal difference in the perception of emotional valence (i.e. the subjective evaluation of positivity or negativity) between the two conditions. Specifically, we found that the emotions anger and happiness were judged to be more intense (i.e. more negative or more positive), in the pictorial condition. Sadness, however, was judged to be more negative in the presence condition. With respect to affection, there were no differences between the conditions.

Although prior studies have demonstrated that immersive virtual environments intensify emotional responses as measured by arousal (Estupiñán et al. 2014; Visch et al. 2010), our study suggests that this does not necessarily result in a more intense *perception* of all emotions. Thus, whether or not a scene is pereived more or less intense within an immersive environment may depend on the valence intensity of the stimulus. More specifically, lower valence emotions such as sadness were perceived less intense in terms of emotional valence, while happiness and anger, two higher valence emotions, were perceived even more intensely within the shared space. Thus, according to Slater (2009), virtual environments may alter our perception by making it more realistic rather than generally more intense. These observations are consistent with a study by Visch et al. (2010), who found that immersive environments enhance the emotion experienced by participants when viewing movies from various genres, yet, does not affect the classification of the respective movie genre. A similar view is also shared by Diemer et al. (2015), who concluded that when indviduals try to make sense of what they are viewing within a virtual environment, they apply the same mechanisms as they do in the real world.

It should be highlighted, however, that the differences between picture and presence mode yielded in our study were only marginal and thus suggest that our perception of emotion from interactions is not altered strongly by increased (co-)presence.

6.4 Interindividual differences

We further investigated interindividual factors that have not been evaluated in the context of perceiving emotional interactions so far. In particular, we assessed how the emotional expressivity, i.e. the ability and willingness to express one's emotions, as well as the gender of the observer, influences emotion recognition.

6.4.1 Emotional Expressivity

Several theories that seek to understand the neural processes underlying emotion recognition, such as the simulation theory, posit that the ability to recognize emotions is highly individualized and depends heavily on an individual's internal representations. More specifically, it suggests that an observer arrives at a mental attribution by simulating, replicating, or reproducing her or his own representations of being in the same state as the observed person (Barsalou et al., 2003; Gallese, 2003; Gallese & Goldman, 1998; Goldman & Sripada, 2005; Rizzolatti & Sinigaglia, 2010).

On a behavioral level, Edey et al. (2017), for instance, showed that there is a linear relationship between one's own walking kinematics and affective state judgments. More specifically, they found that faster participants rated slower emotions more intensely relative to their ratings for faster emotions. The authors suggest that the perception of affective states in others is predicted by one's own movement kinematics.

Thus, within our study, we combined self-report measures of emotional expressivity with an occlusion paradigm to assess whether one's (motor) expression of emotion is associated with the perception of emotion. Moreover, by presenting either arm- or trunk PLDs of emotional interactions, we assessed whether an observer profits differently from either information with respect to their own emotional expressivity.

Indeed, our results provide evidence that one's emotional expressivity is associated with the perception of emotion. Specifically, we found that individuals who report themselves to be more emotionally expressive (verbally and nonverbally), and are thus able and willing to express their own emotions, are better at recognizing emotions in others than individuals with lower subjective emotional expressivity.

Additionally, we found that individuals who reported lower levels of emotional expressivity did not have significantly different recognition accuracy when exposed to either arm or trunk movements. In contrast, individuals with medium or high emotional expressivity profited especially from displays presenting arm movements.

However, one has to bear in mind that the estimation of one's emotional expressivity is based on self-report measures and thus might be. Self-report measures have been critized for not representing a correct image of the mental abilities that are being assessed (e.g. Devlin et al., 2014; Dunning et al., 2003). Yet, despite of this limitation, we discovered a significant relationship between an individual's emotional expressivity and their ability to accurately recognize emotions, as well as their use of available information.

These findings are in line with previous research which shows that one's own walking kinematics influence how affective content is interpreted (Edey et al., 2017). Moreover, Lorey et al., (2012) showed that there is a negative linear relationship between alexithymia (i.e. a compromized ability to perceive and describe their emotions) and emotion recognition. More specifically, the study shows that alexithymic individuals are less confident in their ability to recognize emotions as compared to control participants who do not have this trait (Lorey et al., 2012).

Furthermore, our results align with the concept of the common-coding principle, which posits a shared mapping of emotions and subjective states across individuals. This principle assumes that the greater the similarity between an observed action and the way the observer would perform the action, the stronger the activation of action representations (Calvo-Merino et al., 2005). Therefore, perceiving emotional behavior in others may result in an internal simulation of similar behaviors in ourselves.

A different but not mutually-exclusive approach that may account for our findings is perceptual learning. This approach assumes that repeated exposure to certain movements improves recognition performances (Grossman et al., 2004). With respect to our results, it is possible that highly expressive individuals acquire visual instead of motor representations of their very own emotional expressions, which can then be retrieved to understand another person's expressive behavior.

6.4.2 Gender

A common topic of debate is whether or not gender is a modulating variable when it comes to emotion recognition. It has been suggested that women are more adept than men at recognizing emotion (e.g. Alaerts et al., 2011; Hoffmann et al., 2010; Thayer and Johnsen, 2000). More specifically, research indicates that women detect emotions more quickly and more accurately from facial or bodily expressions (Alaerts et al., 2011). However, some research also suggests that this only holds true for subtle rather than stereotypical displays of emotion (Hoffmann et al., 2010; Montagne et al., 2005).

In our third study, we addressed the issue of gender-related differences with regard to the *subjective* perception of emotion. Within the study, the gender of a person was determined based on their self-identification as male or female. We implemented an emotional intensity task, rather than a categorization task, i.e. rating the valence intensity as compared to an emotion category, allowing for a detection of more subtle differences between women and men.

Our results do not support the notion that there are differences between men and women regarding valence intensity perception of emotional interactions depicted as PLDs. More specifically, there are no gender differences in valence judgments for either more intensely perceived emotions (i.e. happiness and anger) nor for less intense, and thus perhaps more sutble emotions, such as as sadness and affection.

The lack of agreement with regard to gender differences in emotion recognition may, in part, be due to methodological differences. Many paradigms previously used categorization tasks, while in our study, we used a valence judgment task. The expected advantage was that such ratings would allow us to detect more fine-grained differences in the way men and women subjectively perceive emotional body expressions as it uses scalar ratings as opposed to a fixed choice emotion category rating (see also Hall and Matsumoto, 2004; Fischer et al., 2018). Moreover, a valence task, as compared to a categorization task, assesses the intensity of a scene rather than its respective category and is thus a subjective more than an objective measure.

Interestingly, recent neuroscientific research suggests that the functional architecture of the brain is unlikely to be conceptualized as binary (e.g male/female) but is more likely to be continously represented on a brain-gender continuum (Sanchis-Segura et al., 2022). On the behavioral level, this may implicate that traits that are traditionally viewed as typically „female“ or „male“ may not always be distributed in a binary fashion. Thus, the use of male and female categories may not be sufficient to gain a nuanced understanding of how this factor possibily interacts with emotion recognition.

7 Conclusion

The research discussed in this thesis further stresses the link between emotions and action and the importance of the body as a channel of information. While it becomes clear that the information that is retrieved from interactions in order to recognize an emotion highly depends on the type of emotion that is displayed, it also became apparent that the recognition of certain emotions (here, anger and affection) seems to operate through functional relations between two agents, that can be either complementary or congruent depending on the respective emotion. Interestingly, the studies discussed here further show that the observer uses the available information differently depending on their own characteristics (i.e. their own ability and willingness to express and recognize emotions). This is in line with the simulation theory which views social cognition as a highly individual process than depends on one’s own internal representations (Barsalou et al., 2003; Gallese, 2003; Gallese & Goldman, 1998; Goldman & Sripada, 2005; Rizzolatti & Sinigaglia, 2010). Summing up, the studies presented in this thesis contribute to the research field of emotion recognition, adding the following insights:

I. By extending established single-agent PLD paradigms using dyadic interactions, we found that key information from the body is retrieved in an emotion-specific manner. While information about certain emotion categories (here, anger and happiness) is preferably retrieved via arm movements as compared to trunk movements, the recognition of sadness is facilitated by the display of trunk as compared to arm movements.
II. Spatiotemporal coupling within emotional interactions enhances the recognition of angry and affectionate interactions, while the recognition of happiness and sadness does not depend on a time-locked interaction, suggesting that some emotions may

have a stronger interpersonal character and thus depend on highly coordinated movements.

III. While multiple studies have shown that context is essential, we found that co-presence with the interacting partners only marginally influences the emotional percept measured in terms of valence intensity.

IV. With regard to interindividual factors, we found that the emotional expressivity of the observer modulates emotion recognition accuracy and allows the individual to benefit differently from the type of information that is displayed. Observers with high emotional expressivity exhibit higher emotion recognition accuracy.

V. While research on gender differences with respect to emotion recognition from bodies is still limited, our results do not give reason to believe that there are significant differences with respect to subjetively perceived valence intensity, i.e. how negative or positive a scene is perceived to be.

8 Future implications

For future experiments, there are some factors that should be taken into consideration. Firstly, recent studies suggest methodological shortcomings with respect to the stimulus material that is used. More specifically, most studies, including ours, have used actors or students in order to create their stimulus set. A study by Schuster and colleagues (2019), however, suggests that emotion expression in body movements (i.e. gait) is different for posed as compared to induced emotions. However, this effect was only investigated for movement velocity. Thus, it needs to be further investigated whether non naturally-occuring whole body-expressions can be relied upon when investigating emotion recognition.

Secondly, studies should further investigate whether forced-choice categorization paradigms which only use a few categories are suited to capture the complexity of emotion recognition. Although categorical forced-choice paradigms, when compared to free-response paradigms, allow for the testing of recognition rather than the production of emotion concepts, they come with some pitfalls. One of which includes the portrayal of emotion categories as mutually exclusive, thus inflating the agreement across participants (e.g. Russell 1993, 1994). Although categorical forced-choice paradigms have certain advantages, dimensional paradigms may increase information that may otherwise be lost.

In this regard, the use of binary classification such as male and female to assess gender differences has also been criticized (e.g. Sanchis-Segura et al., 2022). Research suggests that gender-related behaviors are highly inconsistent, situationally variable and are often based on stereotypes (Lippa & Connelly, 1990). Thus, a binary classification may not capture the multidimensionality of these behaviors. Therefore, future studies should aim at identifying diagnostic indicators, such as personality traits, that will yield a probability of how likely a person is to belong to a certain group. This allows for a more a-priori, theory-based approach which, in turn, will allow for an easier identification of individual factors that modulate emotion recognition.

Lastly, there is one issue that relates to the use of self-report measures to capture the relationship between emotion recognition and emotional expressivity. Up to now, there is a substantial lack of quantitative measures to support the results obtained from these measures. With respect to the present investigation, further research is needed to assess the relationship between self-reported emotional expressivity and actual movement characteristics.

9 References

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and Emotion Recognition from Point Light Displays: An Investigation of Gender Differences. *PLoS ONE*, *6*(6), e20989. doi: 10.1371/journal.pone.0020989

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion Perception from Dynamic and Static Body Expressions in Point-Light and Full-Light Displays. *Perception*, *33*(6), 717–746. doi: 10.1068/p5096

Averill, J. R. (1980). A constructionist view of emotion. In R. Plutchik and H. Kellerman (Eds.), *Emotion: Theory, research, and experience* (Vol. 1, pp. 305–39). New York, NY: Academic Press. doi: 10.1016/B978-0-12-558701-3.50018-1

Bachmann, J., Munzert, J., & Krüger, B. (2018). Neural Underpinnings of the Perception of Emotional States Derived From Biological Human Motion: A Review of neuroimaging Research. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.01763

Barrett, L. F. (2006). Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. doi: 10.1207/s15327957pspr1001_2

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. In B. H. Ross (Ed.), The psychology of learning and motivation (Vol. 43, pp. 43–92). Academic Press. doi: 10.1016/S0079-7421(03)01011-9

Blair, R. J. R. (2003). Facial expressions, their communicatory functions and neuro–cognitive substrates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 561–572. doi: 10.1098/rstb.2002.1220

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276–298. doi: 10.1037/1528-3542.1.3.276

Calvo-Merino, B., Glaser, D., Grèzes, J., Passingham, R., & Haggard, P. (2004). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cerebral Cortex*, *15*(8), 1243–1249. doi: 10.1093/cercor/bhi007

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The Perception of Emotion from Body Movement in Point-Light Displays of Interpersonal Dialogue. *Perception*, *34*(10), 1171–1180. doi: 10.1068/p5203

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085–1101. doi: 10.1037/a0025737

Dalili, M. N., Penton-Voak, I. S., Harmer, C. J., & Munafò, M. R. (2014). Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychological Medicine*, *45*(6), 1135–1144. doi: 10.1017/s0033291714002591

Charles Darwin. (1859). On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life /. *Journal of the American Medical Association*. doi: 10.5962/bhl.title.59991

Darwin, C. (1872). The expression of the emotions in man and animals. *University of Chicago Press.* doi: 10.1037/10001-000

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3475–3484. doi: 10.1098/rstb.2009.0190

de Gelder, B., de Borst, A., & Watson, R. (2014). The perception of emotion in body expressions. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 149–158. doi: 10.1002/wcs.1335

de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, *13*(4), 247–268. doi: 10.1007/bf00990296 de Rosnay, M., Cooper, P. J., Tsigaras, N., & Murray, L. (2006). Transmission of social anxiety from mother to infant: An experimental study using a social referencing paradigm. *Behaviour Research and Therapy*, *44*(8), 1165–1175. doi: 10.1016/j.brat.2005.09.003

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Sciences*, *7*(12), 527–533. doi: 10.1016/j.tics.2003.10.004

Devlin, H. C., Zaki, J., Ong, D. C. & Gruber, J. (2014). Not As Good as You Think? Trait Positive Emotion Is Associated with Increased Self-Reported Empathy but Decreased Empathic Performance. PLoS ONE, 9(10), e110470. doi: 10.1371/journal.pone.0110470

Dunning, D., Johnson, K., Ehrlinger, J. & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. Current Directions in Psychological Science, 12(3), 83–87. doi: 10.1111/1467-8721.01235

Edey, R., Yon, D., Cook, J., Dumontheil, I., & Press, C. (2017). Our own action kinematics predict the perceived affective states of others. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(7), 1263–1268. doi: 10.1037/xhp0000423

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. doi: 10.1080/02699939208411068

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-Cultural Elements in Facial Displays of Emotion. *Science*, *164*(3875), 86–88. doi: 10.1126/science.164.3875.86

Estupiñán, S., Rebelo, F., Noriega, P., Ferreira, C., & Duarte, E. (2014). Can Virtual Reality Increase Emotional Responses (Arousal and Valence)? A Pilot Study. *Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments*, 541–549. doi: 10.1007/978-3-319-07626-3_51

Fischer, A. H., Kret, M. E., & Broekens, J. (2018). Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis. *PLOS ONE*, *13*(1), e0190712. doi: 10.1371/journal.pone.0190712

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–228. doi: 10.1037/0022-3514.57.2.212

Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 517–528. doi: 10.1098/rstb.2002.1234

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. doi: 10.1016/s1364-6613(98)01262-5

Geraets, C., Tuente, S. K., Lestestuiver, B., van Beilen, M., Nijman, S., Marsman, J., & Veling, W. (2021). Virtual reality facial emotion recognition in social environments: An eye-tracking study. *Internet Interventions*, *25*, 100432. doi: 10.1016/j.invent.2021.100432

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*(3), 193–213. doi: 10.1016/j.cognition.2004.01.005

Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2012). Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Human Movement Science*, *31*(1), 202–221. doi: 10.1016/j.humov.2011.05.001

Grossman, E. D., Blake, R., & Kim, C. Y. (2004). Learning to See Biological Motion: Brain Activity Parallels Behavior. *Journal of Cognitive Neuroscience*, *16*(9), 1669–1679. doi: 10.1162/0898929042568569

Hall, J. A., & Matsumoto, D. (2004). Gender Differences in Judgments of Multiple Emotions From Facial Expressions. *Emotion*, *4*(2), 201–206. doi: 10.1037/1528-3542.4.2.201

Heck, A., Chroust, A., White, H., Jubran, R., & Bhatt, R. S. (2018). Development of body emotion perception in infancy: From discrimination to recognition. *Infant Behavior and Development*, *50*, 42–51. doi: 10.1016/j.infbeh.2017.10.007

Heeter, C. (1992). *Being* There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments*, *1*(2), 262–271. doi: 10.1162/pres.1992.1.2.262

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, *135*(3), 278–283. doi: 10.1016/j.actpsy.2010.07.012

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, *99*(3), 561–565. doi: 10.1037/0033-295x.99.3.561

Jicol, C., Wan, C. H., Doling, B., Illingworth, C. H., Yoon, J., Headey, C., Lutteroth, C., Proulx, M. J., Petrini, K., & O'Neill, E. (2021). Effects of Emotion and Agency on Presence in Virtual Reality. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. doi: 10.1145/3411764.3445588

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception &Amp; Psychophysics*, *14*(2), 201–211. doi: 10.3758/bf03212378

Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., Stark, R., Lis, S., Gallhofer, B., Sammer, G., Zentgraf, K., Munzert, J., & Lorey, B. (2014). Major Depressive Disorder Alters Perception of Emotional Body Movements. *Frontiers in Psychiatry*, *5*. doi: 10.3389/fpsyt.2014.00004

Karon, B. P. & Tomkins, S. S. (2008). Affect Imagery Consciousness: The Complete Edition: Volume I: The Positive Affects. *Springer Publishing Co Inc*.

Kee, K. S., Green, M. F., Mintz, J., & Brekke, J. S. (2003). Is Emotion Processing a Predictor of Functional Outcome in Schizophrenia? *Schizophrenia Bulletin*, *29*(3), 487–497. doi: 10.1093/oxfordjournals.schbul.a007021

Keck, J., Zabicki, A., Bachmann, J., Munzert, J., & Krüger, B. (2022). Decoding spatiotemporal features of emotional body language in social interactions. *Scientific Reports*, *12*(1). doi: 10.1038/s41598-022-19267-5

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (2003). Facial expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 415–432). Oxford University Press.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2009). Facial Emotion Perception in Schizophrenia: A Meta-analytic Review. *Schizophrenia Bulletin*, *36*(5), 1009–1019. doi: 10.1093/schbul/sbn192

Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective Body Expression Perception and Recognition: A Survey. *IEEE Transactions on Affective Computing*, *4*(1), 15–33. doi: 10.1109/t-affc.2012.16

Krüger, B., Kaletsch, M., Pilgramm, S., Schwippert, S. S., Hennig, J., Stark, R., Lis, S., Gallhofer, B., Sammer, G., Zentgraf, K., & Munzert, J. (2017). Perceived Intensity of Emotional Point–Light Displays is Reduced in Subjects with ASD. *Journal of Autism and Developmental Disorders*, *48*(1), 1–11. doi: 10.1007/s10803-017-3286-y

Lahnakoski, J. M., Forbes, P. A., McCall, C. & Schilbach, L. (2020). Unobtrusive tracking of interpersonal orienting and distance predicts the subjective quality of social interactions. *Royal Society Open Science*, 7(8), 191815. doi: 10.1098/rsos.191815

Lippa, R., & Connelly, S. (1990). Gender diagnosticity: A new Bayesian approach to gender-related individual differences. *Journal of Personality and Social Psychology*, *59*(5), 1051–1065. doi: 10.1037/0022-3514.59.5.1051

Lorey, B., Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., Stark, R., Zentgraf, K., & Munzert, J. (2012). Confidence in Emotion Perception in Point-Light Displays Varies with the Ability to Perceive Own Emotions. *PLoS ONE*, *7*(8), e42169. doi: 10.1371/journal.pone.0042169

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative Interactions Improve Visual Detection of Biological Motion. *PLoS ONE*, *6*(1), e14594. doi: 10.1371/journal.pone.0014594

Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time Will Show: Real Time Predictions during Interpersonal Action Perception. *PLoS ONE*, *8*(1), e54949. doi: 10.1371/journal.pone.0054949

Marín-Morales, J., Llinares, C., Guixeres, J., & Alcañiz, M. (2020). Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing. *Sensors*, *20*(18), 5163. doi: 10.3390/s20185163

Matsumoto, D., & Hwang, H. C. (2017). Methodological Issues Regarding Cross-Cultural Studies of Judgments of Facial Expressions. *Emotion Review*, *9*(4), 375–382. doi: 10.1177/1754073916679008

Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., & Schulte, D. (2009). Embodiment of Sadness and Depression—Gait Patterns Associated With Dysphoric Mood. *Psychosomatic Medicine*, *71*(5), 580–587. doi: 10.1097/psy.0b013e3181a2515c

Montagne, B., Kessels, R. P. C., Frigerio, E., de Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, *6*(2), 136–141. doi: 10.1007/s10339-005-0050-6

Moreau, Q., Galvan, L., Nazir, T. A., & Paulignan, Y. (2016). Dynamics of Social Interaction: Kinematic Analysis of a Joint Action. *Frontiers in Psychology*, *7*. doi: 10.3389/fpsyg.2016.02016

Motta, S. C., Goto, M., Gouveia, F. V., Baldo, M. V. C., Canteras, N. S., & Swanson, L. W. (2009). Dissecting the brain's fear system reveals the hypothalamus is critical for responding in subordinate conspecific intruders. *Proceedings of the National Academy of Sciences*, *106*(12), 4870–4875. doi: 10.1073/pnas.0900939106

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*(9), 1186–1192. doi: 10.1038/nn1759

Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology*, *7*(1), 1–30. doi: 10.1111/j.1467-9450.1966.tb01334.x

Persad, S. M., & Polivy, J. (1993). Differences between depressed and nondepressed individuals in the recognition of and response to facial emotional cues. Journal of Abnormal Psychology, 102(3), 358–368. doi: 10.1037/0021-843X.102.3.358

Poyo Solanas, M., Vaessen, M. J., & de Gelder, B. (2020). The role of computational and subjective features in emotional body expressions. Scientific Reports, 10(1). doi: 10.1038/s41598-020-63125-1

Qiao-Tasserit, E., Garcia Quesada, M., Antico, L., Bavelier, D., Vuilleumier, P., & Pichon, S. (2017). Transient emotional events and individual affective traits affect emotion recognition in a perceptual decision-making task. *PLOS ONE*, *12*(2), e0171375. doi: 10.1371/journal.pone.0171375

Rindermann, H. (2009). Emotionale-Kompetenz-Fragebogen: EKF; Einschätzung emotionaler Kompetenzen und emotionaler Intelligenz aus Selbst-und Fremdsicht. *Hogrefe*.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264–274. doi: 10.1038/nrn2805

Ross, P. D., Polson, L., & Grosbras, M. H. (2012). Developmental Changes in Emotion Recognition from Full-Light and Point-Light Displays of Body Movement. *PLoS ONE*, *7*(9), e44815. doi: 10.1371/journal.pone.0044815

Russell, J. A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology*, *45*(6), 1281–1288. doi: 10.1037/0022-3514.45.6.1281

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, *17*(1), 41–51. doi: 10.1007/bf00995206

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. doi: 10.1037/0033-2909.115.1.102

Sanchis-Segura, C., Aguirre, N., Cruz-Gómez, L. J., Félix, S., & Forn, C. (2022). Beyond "sex prediction": Estimating and interpreting multivariate sex differences and similarities in the brain. *NeuroImage*, *257*, 119343. doi: 10.1016/j.neuroimage.2022.119343

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion*, *23*(7), 1307–1351. doi: 10.1080/02699930902928969

Schuster, B., Sowden, S., Abdlkarim, D., Wing, A., & Cook, J. (2019). Acting is not the same as feeling: Emotion expression in gait is different for posed and induced emotions. *Frontiers in Human Neuroscience*, *13*. doi: 10.3389/conf.fnhum.2019.229.00010

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. doi: 10.1162/pres.1997.6.6.603

Sokolov, A. A., Krüger, S., Enck, P., Krägeloh-Mann, I., & Pavlova, M. A. (2011). Gender Affects Body Language Reading. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00016

Souto, T., Silva, H., Leite, A., Baptista, A., Queirós, C., & Marques, A. (2019). Facial Emotion Recognition: Virtual Reality Program for Facial Emotion Recognition—A Trial Program Targeted at Individuals With Schizophrenia. *Rehabilitation Counseling Bulletin*, *63*(2), 79–90. doi: 10.1177/0034355219847284

Thayer, J., & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology*, *41*(3), 243–246. doi: 10.1111/1467-9450.00193

Trevisan, D. A., & Birmingham, E. (2016). Are emotion recognition abilities related to everyday social functioning in ASD? A meta-analysis. *Research in Autism Spectrum Disorders*, *32*, 24–42. doi: 10.1016/j.rasd.2016.08.004

Troje, N. F. (2019). Reality Check. *Perception*, *48*(11), 1033–1038. doi: 10.1177/0301006619879062

Tse, W. S., & Bond, A. J. (2004). The Impact of Depression on Social Skills. *Journal of Nervous and Mental Disease*, *192*(4), 260–268. doi: 10.1097/01.nmd.0000120884.60002.2b

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, *7*(3), 487–494. doi: 10.1037/1528-3542.7.3.487

van der Veen, F. M., Evers, E. A. T., Deutz, N. E. P., & Schmitt, J. A. J. (2006). Effects of Acute Tryptophan Depletion on Mood and Facial Emotion Perception Related Brain Activation and Performance in Healthy Women with and without a Family History of Depression. *Neuropsychopharmacology*, *32*(1), 216–224. doi: 10.1038/sj.npp.1301212

von der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150373. doi: 10.1098/rstb.2015.0373

Visch, V. T., Tan, E. S., & Molenaar, D. (2010). The emotional and cognitive effect of immersion in film viewing. *Cognition &Amp; Emotion*, *24*(8), 1439–1445. doi: 10.1080/02699930903498186

Harald G. Wallbott. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, *28*(6), 879–896. doi: 10.1002/(sici)1099-0992(1998110)28:6

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, *51*(4), 690–699. doi: 10.1037/0022-3514.51.4.690

Walle, E. A., Reschke, P. J., & Knothe, J. M. (2017). Social Referencing: Defining and Delineating a Basic Process of Emotion. *Emotion Review*, *9*(3), 245–252. doi: 10.1177/1754073916669594

Walker-Andrews, A. S. (1998). Emotions and Social Development: Infants' Recognition of Emotions in Others. *Pediatrics*, *102*(5), 1268–1271. doi: 10.1542/peds.102.se1.1268

Yokozuka, T., Ono, E., Inoue, Y., Ogawa, K. I., & Miyake, Y. (2018). The Relationship Between Head Motion Synchronization and Empathy in Unidirectional Face-to-Face Communication. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.01622

10 List of publications

Bachmann, J., Munzert, J., & Krüger, B. (2018). Neural underpinnings of the perception of emotional states derived from biological human motion: A review of neuroimaging research. *Frontiers in Psychology*, *9*, 1763.

Bachmann, J., Zabicki, A., Munzert, J., & Krüger, B. (2020). Emotional expressivity of the observer mediates recognition of affective states from human body movements. *Cognition & Emotion, 34(7), 1370-1381.*

Bachmann, J., Zabicki, A., Gradl, S., Kurz, J., Munzert, J., Troje, N. F., & Krueger, B. (2021). Does co-presence affect the way we perceive and respond to emotional interactions? *Experimental brain research*, 239(3), 923-936.

Bachmann, J., Krüger, B., Keck, J., Munzert, J., & Zabicki, A. (2022). When the timing is right: The link between temporal coupling in dyadic interactions and emotion recognition. *Cognition*, 229, 105267.

Keck, J., Zabicki, A., **Bachmann, J.**, Munzert, J., & Krüger, B. (2022). Decoding spatiotemporal features of emotional body language in social interactions. *Scientific Reports*, 12(1).