Observers predict actions from facial emotional expressions during real-time social interactions

Leon O. H. Kroczek[1*,] Angelika Lingnau[2], Valentin Schwind[3, 4], Christian Wolff[4], & Andreas Mühlberger[1]

[1] Department of Psychology, Clinical Psychology and Psychotherapy, University of Regensburg, Regensburg, Germany

[2] Department of Psychology, Cognitive Neuroscience, University of Regensburg, Regensburg, Germany

[3] Human Computer Interaction, University of Applied Sciences in Frankfurt a. M, Frankfurt a. M., Germany

[4] Department of Media Informatics, University of Regensburg, Regensburg, Germany

Short title: Facial emotion and action

Key words: Psychophysiology, Social Cognition, Virtual Reality, Action, Emotion, Heart rate

* Corresponding author

Mail: leon.kroczek@ur.de

Abstract

In face-to-face social interactions, emotional expressions provide insights into the mental state of an interactive partner. This information can be crucial to infer action intentions and react adaptively towards another person’s actions. Here we investigate how facial emotional expressions impact subjective experience and physiological and behavioral responses to social actions during real-time interactions. Thirty-two participants interacted with virtual agents while fully immersed in Virtual Reality. Agents displayed an angry or happy facial expression before they directed an appetitive (fist bump) or aversive (punch) social action towards the participant. Participants responded to these actions, either by reciprocating the fist bump or by defending the punch. For all interactions, subjective experience was measured using ratings. In addition, physiological responses (electrodermal activity, electrocardiogram) and participants’ response times were recorded. Aversive actions were judged to be more arousing and less pleasant relative to appetitive actions. In addition, angry expressions increased heart rate relative to happy expressions. Crucially, interaction effects between facial emotion and action were observed. Angry expressions reduced pleasantness stronger for appetitive compared to aversive actions. Furthermore, skin conductance responses to aversive actions were increased for happy compared to angry expressions and reaction times were faster to aversive compared to appetitive actions when agents showed an angry expression. These data indicate that observers used facial emotional expression to generate expectations for particular actions. Consequently, the present study demonstrates that observers integrate information from facial emotional expressions with actions during social interactions.

(247 words)

Impact statement (70 words)

The study implements a novel paradigm allowing for real-time social interaction in Virtual Reality to study interaction effects of facial emotion and action in experience, physiology, and behavior. Facial emotions affected the experience in social interactions as well as physiological and behavioral responses to actions. Our data strongly suggest that observers process facial emotional expressions to infer action intentions in order to generate fast and adaptive responses during social encounters.

Highlights:

- Studying social interaction using an ecologically valid, closed-loop interaction paradigm in Virtual Reality
- Measurement of subjective experience, physiological responses, and behavior
- Facial emotional expressions are integrated with social actions and affect the evaluation of social interactions
- Facial emotional expressions activate the sympathetic system and support adaptive behavior

1. Introduction

Social actions - like greeting someone with a handshake, congratulating with a tap on the shoulder, comforting with touch, or defending oneself with a push or a punch – are a fundamental part of real-life human interactions. Every day we experience numerous social encounters with different agents in various (emotional) contexts and with a range of different communicative goals. Importantly, in each of these encounters, people need to coordinate social actions between themselves and the interaction partner in a fast and adaptive manner (Curioni et al., 2019; Sebanz et al., 2006). Inferring another person's intention before an action has been completed may allow preparing adaptive responses. This may help to keep social interactions in synchrony, e.g. by reciprocating a handshake, or to get a time advantage in preparing a defense, e.g. when an attack needs to be parried (Sebanz & Knoblich, 2009). Typically, observers are fast and accurate in inferring other agents' intentions both for actions directed towards persons and objects (Frith & Frith, 2006; Sartori et al., 2011). This raises the question how observers infer action intentions during social interactions.

Previous research has demonstrated that observers exploit a wide range of multimodal cues to infer intentions. These cues include action kinematics (Becchio et al., 2008), preshaping of the hand (Ambrosini et al., 2011), but also body posture (De Gelder, 2006), gaze (Ambrosini et al., 2011, 2015), and facial expressions (Kroczek et al., 2021). In addition, observers use contextual information related to the situation, identity, and gender of the interactive partner (Bach & Schenke, 2017; Ferstl et al., 2017; Krüger et al., 2013). However, while there is evidence that the processing of action intentions relies on a range of different sources of information, the exact mechanisms by which social cues impact the processing of action intentions and the preparation of responses during social interactions remain elusive.

Facial emotional expressions are highly salient non-verbal communicative cues that are omnipresent in interpersonal encounters (Frith, 2009). Not only do facial emotions allow to infer the mental state of the interactive partner but they can also be predictive with respect to upcoming social actions (Jack & Schyns, 2015). As an example, Kroczek et al. (2021) recently demonstrated that observing an angry facial expression biased participants' action judgements towards aversive actions (i.e. punches), especially when actions were hard to recognize. These data suggest that observers use facial emotional expressions to infer action intentions. Such information can be beneficial in preparing adaptive responses (Csibra & Gergely, 2007). However, most previous studies investigated the processing of action intentions in the absence of interactive behavior by simply letting participants passively observe actions (cf. isolation paradigms, Becchio et al., 2010). For social-interactive actions, non-interactive paradigms may reveal an incomplete picture only. This is especially relevant as appetitive or aversive outcomes of social actions only come into play when there is an interaction between persons. In order to investigate such claims it is important to study online interactive paradigms that require an interactive partner, not only to observe but also to react upon actions of another person (X. Pan & Hamilton, 2018; Redcay & Schilbach, 2019; Schilbach et al., 2011). Thus, studying real-time social interactions may be advantageous for understanding mechanisms related to the processing of social action intentions.

The goal of the present study was to investigate whether facial emotions impact the evaluation of face-to-face social interactions and whether facial emotions bias physiological and behavioral responses to actions. For this reason, we implemented a novel Virtual Reality (VR) paradigm using a Cave automatic virtual environment (CAVE) system where participants interacted with virtual agents (one female, one male). As experimental manipulations, virtual agents first displayed a facial emotional expression (happy vs. angry) and then performed an action towards the participants (fist bump vs. punch). Participants were instructed to react to this action by using a congruent action (reciprocal fist bump vs. defend punch), thereby moving their hand to the position of the hand of the virtual agent. Continuous tracking of participants' hand movements allowed creating action-contingent reactions once participants had reached the target position, thereby increasing interactivity. We obtained ratings of arousal, valence, and realism after each interaction to characterize subjective experiences. In addition, physiological parameters (electrodermal activity, EDA, electrocardiogram, ECG) were continuously recorded during interactions and response times (RTs) of action responses were measured as a behavioral index. We expected that facial expressions would affect behavioral and physiological responses and the evaluation of social actions. Because of their negative valence and high salience, angry compared to happy facial expressions should increase unpleasantness and physiological responses in social interactions and should facilitate responses to actions by decreasing reaction times.

Materials and methods

1.1. Participants

Thirty-three healthy students participated in the study (23 female, M_{Age} = 22.20 years, SD_{Age} = 2.84, $range_{Age}$ = 18 – 30 years). All participants had normal or corrected-to-normal vision and did not report any mental or neurological disorder. One participant was excluded from analysis of HR data due to excessive artifacts in the ECG data. Another participant was excluded from the RT analysis due to technical problems with RT measurement. The study was reviewed and approved by the ethics board of the University of Regensburg and the study was conducted according to the approved procedures. The study is in line with the Declaration of Helsinki. All participants gave written informed consent.

1.2. Study design

We used an *Emotion (2)* x *Action (2)* within-subject design. Participants engaged in short face-to-face interactions with virtual agents (female or male). During these interactions, we manipulated the facial emotional expression that was displayed by the virtual agent (independent variable *Emotion*: angry vs. happy) and the action that was performed by the virtual agent (independent variable *Action*: fist bump vs. punch). Agents always displayed the emotional facial expression first and then performed the action. In order to investigate the interplay of emotion and action, we measured participants' subjective experience in terms of ratings of arousal, valence, and realism, as well as physiological responses with respect to heart rate and skin conductance and the reaction times of participants' responses towards the actions of the virtual agents.

1.3. Apparatus and stimulus material

The present experiment was conducted in Virtual Reality using a CAVE system with a size of 3.6 m x 2.4 m x 2.5 m. Participants wore 3D shutter glasses with attached motion tracker targets (Advance Realtime Tracking GmbH). Virtual Reality was projected on the four surrounding walls and the floor of the CAVE (Barco F50 WQX6A projectors with a resolution of 2560 by 1600 pixels). An additional motion tracker target (Advance Realtime Tracking GmbH) was attached to the right hand of the participants. VR was rendered using the Unreal 4 game engine (v 4.22, Epic Games Inc.) in a cluster of ten computers (i7-4790k, GeForce 1080, 16 GB RAM). Sounds were presented via a surround sound system (Yamaha HTR-3066).

The experiment included a full virtual room, two virtual agents as well as action animations of fist bump and punch actions. Video stimuli showing the same room, agents, and animations have been implemented in a previous study (Kroczek et al., 2021). An empty room with an elevator door at the front wall served as the virtual environment (see Figure 1). Two virtual agents (one male and one female) were created using Daz3D (Daz3D Inc.), based on the standard Genesis 8 models (https://www.daz3d.com/genesis8) with black clothes and standard geometry-based hair. Animated actions (fist bump or punch) were created for both agents. Animations were based on movement recordings from one male and one female actor. Fist bump and punch actions of the actors were recorded by tracking 55 optical markers (39 passive body, 16 active finger marker) on a full-body motion tracking suit using an OptiTrack motion capturing system (12 cameras: 8 PRIME 13 and 4 PRIME 13W). Movements and skeleton animation were recorded using Motive software (v 2.2) and then preprocessed in Autodesk 3ds Max (v 2019) by reducing the number of keyframes (from 240 to 30 fps) for automatic noise reduction. Animations of both actions were aligned to a common reference frame and the ten initial frames of the actions were averaged, resulting in ten similar, albeit not identical, frames for both animations. This procedure was performed separately for the male and the female animations. Finally, three different exemplars per action were created during post-processing by inducing slight variations with respect to the end position (vertical and horizontal offset) of each action. All animations were exported into the Unreal Engine. As action animations were recorded from real actions, punches were performed faster than fist bump action (movement time fist bump; M = 1.05 s, SD = 0.039 s, movement time punch: M = 0.738 s, SD = 0.023). This difference lies in the nature of the actions.

Physiological measures included electrocardiogram and electrodermal activity. For ECG recordings, three electrodes were attached to the chest of the participants with one electrode at the sternum, a reference electrode at the left, lower coastal arch and a ground electrode at the right, lower coastal arch. For EDA recordings, two 6 mm Ag/AgCl electrodes were attached to the thenar site of the palm of the left hand (Boucsein et al., 2012). All physiological data were recorded at 1000 Hz using a V-Amp amplifier (BrainProducts, Gilching, Germany) connected to a recording PC. In order to allow free movements inside the CAVE system, participants carried the amplifier inside a backpack during the experiment. Data were recorded with BrainVision Recorder software (BrainProducts, Gilching, Germany) and streamed using the Lab Streaming Layer (LSL, Kothe, 2014).

In addition, participant’s head position and the position of the hand was tracked at 60 Hz (DTrack 2 software, Advance Realtime Tracking GmbH) and recorded using LSL.

1.4. Procedure

Upon arrival at the laboratory participants were informed about experimental procedures, gave written informed consent, and filled in questionnaires to assess anxiety and social cognition and to screen participants for clinically relevant symptoms. These questionnaires included demographic information (age, sex, occupation), social anxiety (Social Phobia Inventory; Sosic et al., 2008), general anxiety (State-Trait anxiety inventory, as well as sensitivity to reward and punishment (SRSP, Torrubia et al., 2001). In addition, the “reading the mind in the eyes” test was conducted (Baron-Cohen et al., 2001).

Then, electrodes were attached at the chest (ECG) and left hand (EDA) of the participant. In addition, a motion tracker target was attached to the right hand of the participants, and they were given 3D shutter glasses. Finally, participants were led into the CAVE system where the empty virtual room was displayed.

After entering the CAVE, participants were allowed to explore the environment for 2 minutes and were then asked to stand on a marked starting position. The starting position was located at the center of the CAVE and was oriented so that participants would face the elevator door, which was projected on the long wall of the CAVE (distance from the starting position to the elevator door in VR was approx 3.5 m).

Once participants were standing correctly oriented on the starting position, the experiment was started. The experiment consisted of 24 trials with an identical trial structure (see Figure 1). Trial order was pseudo-randomized with no more than three repetitions of emotion, action, or agent gender. Every trial started with the opening of the elevator door, with the agent standing behind the door. Agents were always facing the participant with their gaze focused on the participant. After 2000 ms, the agents started to display a facial emotional expression (happy condition: smile; angry condition: frown). The agent maintained this emotional expression throughout the trial. Another 2500 ms after the onset of the facial expression, the agent moved to a position in front of the participant (0.75 m distance to participants’ starting position, walking duration 2500 ms). Agents were in a neutral body posture with both hands in a resting position, hanging loosely next to the legs. The agent remained in this position for another 2000 ms and then initiated the action with the right hand (fist bump or punch). Importantly, the action was stopped at the apex position of the movement and participants had to react towards the action by moving their right hand to the hand position of the virtual agent. Participants were instructed to show congruent actions, i.e. to respond with a fist bump when the agent performed a fist bump and to defend the punch when the agent performed a punch. No instructions were given regarding response speed and accuracy. Once participants reached the target position (defined as a sphere with a radius of 15 cm centered around the position of the agent hand), a clapping sound was played from the loudspeakers and the virtual agent retracted the hand to the resting position. The agent then turned around and left the room through the door.

After the agent had left, ratings were obtained for arousal, valence, and realism by auditory presentation of the rating questions. Participants gave an oral response which was noted by the experimenter. Arousal ratings were obtained by asking “How high was your arousal?” (0 = no arousal, 100 = very high arousal), valence ratings were obtained by asking “How unpleasant did you feel?” (0 = very pleasant, 100 = very unpleasant), and realism ratings were obtained by asking “How realistic was the situation?” (0 = completely unrealistic, 100 = completely realistic). Following the ratings, the next trial started.

The Virtual Reality experiment had a total duration of approximately 25 minutes. After the last trial had been presented, participants were led outside the CAVE and filled in further questionnaires including a second state anxiety inventory, as well as questionnaires related to presence (iGroup presence scale, Multimodal Presence Scale, Makransky et al., 2017; Schubert et al., 2001) and simulator sickness (Kennedy et al., 1993).

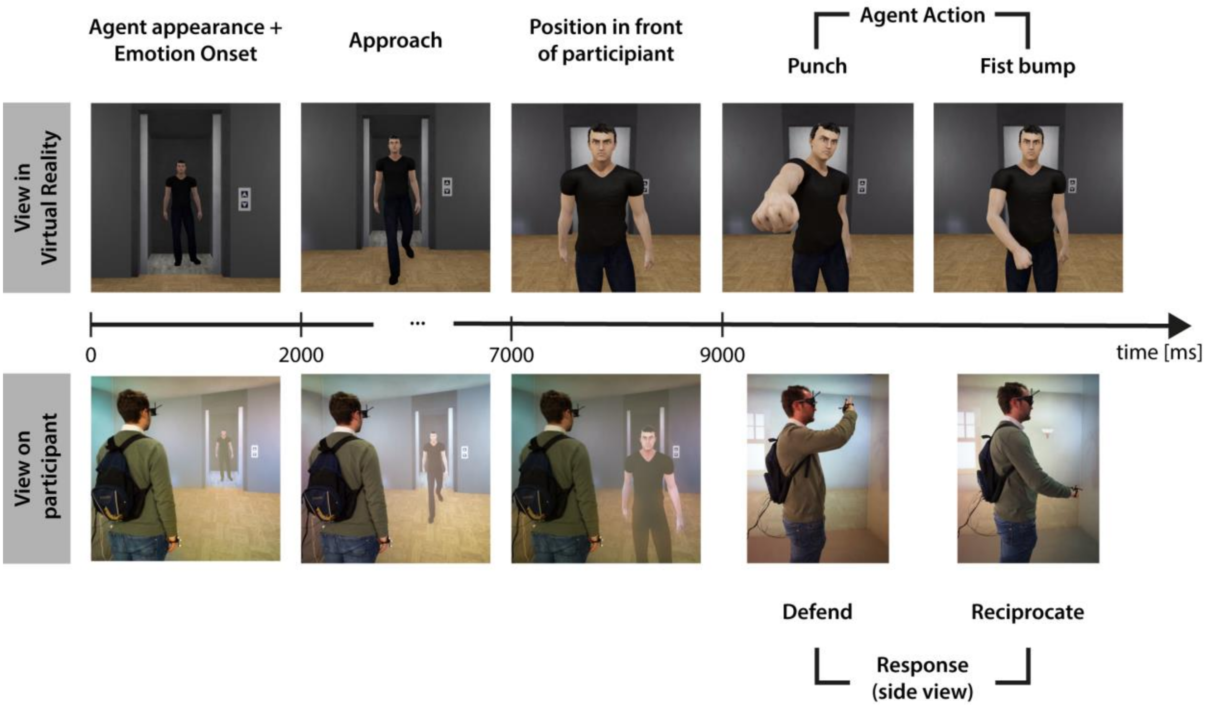

Figure 1: Schematic illustration of the experimental trial structure. Upper row shows trial procedure in Virtual Reality, bottom row shows participant inside the CAVE system at corresponding time points. At trial start, an elevator door opened and revealed a virtual agent (first column). Next, the agent displayed either a happy or angry facial emotional expression and then approached (second column) the participant until a final position was reached (third column). 2000 ms after reaching this position, the virtual agent initiated either a punch or fist bump action (fourth and fifth column respectively). Participants had to react towards this action with a congruent response, either by defending the punch (fourth column, bottom row) or by reciprocating the fist bump (fifth column, bottom row).

1.5. Data processing and statistical analyses

Physiological and behavioral data were preprocessed using custom scripts in MATLAB (v 8.6, MathWorks, Natik, USA). For ECG data, the Pan-Tompkins algorithm was applied to identify R-peaks in the continuous signal (J. Pan & Tompkins, 1985). One participant had to be excluded from analysis because R-peaks could not be identified reliably. Next, intervals between R-peaks (RR) were calculated and converted to heart rate. RR-intervals which deviated more than three standard deviations from the mean were excluded (mean percentage of excluded RR intervals = 0.93, SD = 0.64). In addition, segments of interest were manually checked for incorrectly identified R-peaks. To obtain event-related measures, heart rate was interpolated and sampled at 1000 Hz. Finally, segments with a length of 16 s timelocked to the onset of the emotion of the virtual agent including a 2 s pre-stimulus interval were extracted. Segments were baseline corrected by subtracting the average heart rate in the 2 s period before the emotion onset. For statistical analysis, HR data was averaged in segments of 1 s length.

For EDA, data were low-pass filtered using a first order butterworth filter with a cut-off frequency of 1 Hz and then log-transformed to account for the non-normal distribution (Boucsein et al., 2012). Analogous to ECG analysis, segments of 16 s length were extracted time-locked to emotion onset including a 2 s pre-stimulus interval. Segments were baseline corrected using the 2 s pre-stimulus interval. For statistical analysis, SCR amplitudes were further averaged in 16 non-overlapping time windows of 1 s length.

Reaction times (RTs) were calculated as the time difference between the time point when the virtual agents reached the apex position of the action and the time point when participants reached the hand position of the virtual agent. We chose apex position as a start of RT measurement because action animations of fist bump and punch were of different length (see above). Trials were rejected when participants responded later than 2 seconds after the agent had completed the action (mean number of rejected trials = 1.78 trials, SD = 1.86).

All data were averaged across trials into four experimental conditions (Emotion x Action: Happy – Fist bump, Happy – Punch, Angry – Fist bump, and Angry – Punch) and then exported for further analyses. Statistical analyses were conducted in the R environment (v 4.1.1, R Core Team, 2016). For rating variables and RTs, we conducted repeated-measures ANOVAs with the within-subject factors Emotion and Action. For the windowed time-series data of EDA and HR responses, we conducted repeated-measures ANOVAs with the within-subject factors Time Window, Emotion, and Action.

Post-hoc t-tests were conducted to follow up on significant effects with the Holm method (Holm, 1979) applied to correct for multiple comparisons. Assumptions of normality were assessed by Shapiro-Wilk tests ($p > .05$). All analyses were conducted with Type-I errors set to alpha = 5 %.

1.6. Open Science statement

Study procedures, hypotheses, and analyses were not pre-registered prior to data acquisition. Anonymized raw data and analysis scripts are publicly available in an online repository (https://osf.io/q4cru/).

2. Results

2.1. Experience: Ratings

Arousal

A repeated measures ANOVA with Arousal ratings as dependent variable (see Figure 2, left panel) revealed a main effect of *Emotion*, F(1,32) = 14.16, p < .001 , η_p^2 = 0.31, and a main effect of *Action* F(1,32) = 8.52, p = .006, η_p^2 = 0.21, but no interaction effect between *Emotion* and *Action*, F(1,32) = 0.36, p = .553, η_p^2 = 0.01. Arousal was rated significantly higher for angry (M = 35.0, SD = 18.4) compared to happy facial expressions (M = 28.6, SD = 15.2) as well as for punch (M = 34.8, SD = 18.0) compared to fist bump actions (M = 29.8, SD = 16.0).

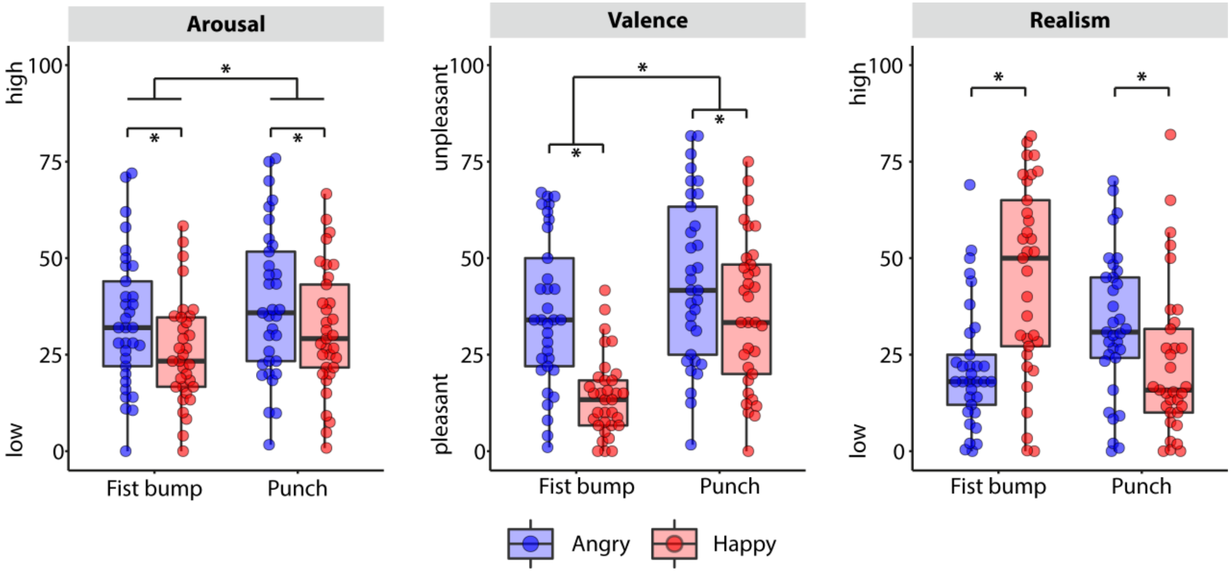

Figure 2: Subjective experience as a function of facial emotional expression and action of the virtual agent. Ratings on a scale from 0 to 100 reflect arousal (left), valence (middle), and realism (right). Box plots are superimposed with individual data points.

Valence

For valence ratings (Figure 2, middle panel), we obtained a main effect of *Emotion, F(1,32)* = 41.19, p < .001, η_p^2 = 0.56, a main effect of *Action*, F(1,32) = 50.42, p <

.001, $\eta_p^2 = 0.61$, as well as a significant interaction between *Emotion* and *Action*, $F(1,32) = 26.38$, $p < .001$, $\eta_p^2 = 0.45$. Post-hoc t-tests (Holm corrected) revealed that happy expressions with fist bump actions were rated as more pleasant compared to other combinations of facial expression and action (Angry-Fist bump: $t(32) = -7.18$, $p < .001$, $d = -1.25$; Angry-Punch: $t(32) = -8.52$, $p < .001$, $d = 1.48$: Happy-Punch: $t(32) = -7.33$, $p <.001$, $d = -1.28$), while angry expressions paired with punch actions were rated as more unpleasant compared to other combinations of facial expression and action (Angry-Fist bump: $t(32) = 4.57$, $p < .001$, $d = 0.80$; Happy-Punch: $t(32) = 3.77$, $p = .001$, $d = 0.66$). Interestingly, the effect of emotion (i.e. the difference between happy and angry expressions) was significantly stronger when agents performed fist bump actions compared to punch actions, $t(32) = -5.14$, $p < .001$, $d = -0.89$.

In sum, the analysis of valence ratings revealed that happy expressions with fist bump actions were evaluated as most pleasant, while angry expressions with punch actions were evaluated as most unpleasant. Furthermore, facial emotions had a stronger modulatory effect on appetitive fist bump actions compared to aversive punch actions.

Realism

For realism ratings (Figure 2, right panel), results showed a main effect of *Emotion*, $F(1,32) = 15.02$, $p < .001$, $\eta_p^2 = 0.32$, a main effect of *Action,* $F(1,32) = 9.93$, $p = .004$, $\eta_p^2 = 0.24$, as well as a significant interaction between *Emotion* and *Action*, $F(1,32) = 23.63$, $p < .001$, $\eta_p^2 = 0.43$. Post-hoc t-tests revealed that fist bumps with a happy expression were rated as more realistic than fist bumps with an angry expression, $t(32) = 5.85$, $p < .001$, $d= 1.02$. In contrast, punch actions were rated as more realistic when performed with an angry expression compared to a happy expression, $t(32) = 2.77$, $p =.018$, $d= 0.48$. However, fist bumps with happy expressions were rated as even more realistic than punches with angry expression, $t(32) = 4.33$, $p < .001$, $d = 0.75$. Furthermore, the effect of facial expressions was stronger for fist bump actions compared to punch actions, $t(32) = 4.86$, $p < .001$, $d = 0.85$. In other words, the combination of congruent pairs of facial expression and action was rated as more realistic than incongruent pairs of facial expression and action. However, appetitive face-action pairs were more realistic than aversive face-action pairs.

2.2. Physiology

Heart rate

Changes in heart rate following the display of the facial emotional expression (Figure 3) were analyzed throughout the interaction with the virtual agent using a 2 x 2 x 16 repeated measures ANOVA with the factors *Emotion*, *Action* and *Time Window*. There was a significant main effect of *Emotion*, $F(1,31) = 8.446$, $p = .007$, $\eta_p^2 = 0.21$, a significant main effect of *Time Window*, $F(16,465) = 17.215$, $p <.001$, , $\eta_p^2 = 0.36$ ($\varepsilon = 0.28$), and a significant interaction of *Emotion* and *Time Window, $F(15,465) = 3.65$, $p = .007$,* $\eta_p^2 = 0.11$ ($\varepsilon = 0.27$). There was no significant effect involving the factor *Action* (all $F < 1$). The main effect of *Time Window* was driven by a general heart deceleration at the onset of the action in the time window from 7 to 8 seconds compared to the preceding time window from 6 to 7 seconds, $t(32) = -3.92$, $p = .005$, and a consecutive heart acceleration in the time window from 10 to 11 seconds compared to the preceding time window from 9 to 10 seconds, $t(32) = 5.58$, $p < .001$. With respect to the interaction effect between *Emotion* and *Time Window*, a follow-up analyses revealed that angry compared to happy facial expressions increased heart rate from 5 to 9 seconds post emotion onset, $Fs(1,31) = 4.63\text{-}11.04$, $ps = .002 - .039$, $\eta s_p^2 = 0.13\text{-}0.28$.

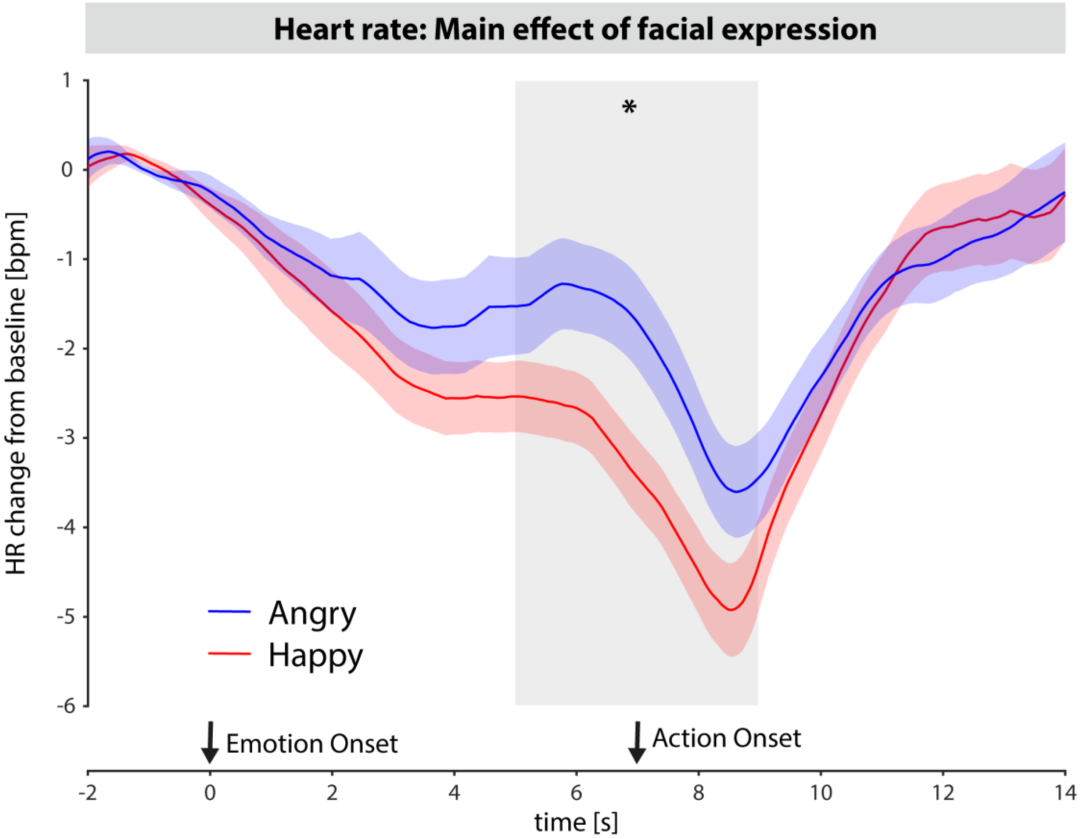

Figure 3: Effect of facial emotional expression on heart rate following the onset of the facial emotion. Time windows with significant differences between angry and happy facial expressions are highlighted in grey. Shaded areas reflect SEM.

Skin conductance responses

SCR amplitude following the onset of facial emotional expression was analyzed using a 2 x 2 x 16 repeated measures ANOVA including the factors *Emotion*, *Action,* and *Time Window*. As can be seen in Figure 4, a typical SCR peak was observed following the action of the virtual agent. The statistical analysis revealed a main effect of *Action*, F(1,32) = 5.99, p = .020, η_p^2 = 0.16, a main effect of *Time Window*, F(15,480) = 21.41, p < .091, η_p^2 = 0.40 (ε = 0.10), an interaction of *Action* and *Time Window,* F(15,480) = 15.57, p < .001, η_p^2 = 0.33 (ε = 0.15), as well as an interaction of *Emotion*, *Action* and *Time Window*, F(15,480) = 3.12, p = .049, η_p^2 = 0.09 (ε = 0.14). No other effects were significant.

Follow-up analyses revealed a main effect of action with increased SCR for punch compared to fist bump actions in time windows from 10 – 14 seconds post emotion onset, Fs(1,32) = 12.06 -*26.83* , ps < .001, ηs_p^2 = 0.27 – 0.46, and a significant interaction effect of *Emotion* and *Action* from 11-14 seconds post emotion onset, Fs(1,32) = 4.40 – 5.87, ps = .021 - .044, ηs_p^2 = 0.12 – 0.16. In the window of the significant *Emotion* by *Action* interaction *(*11-14 s post emotion onset), post-hoc tests showed an increased SCR when punch actions had been preceded by a happy compared to an angry facial expression, t(32) = 3.54, p = .005, d = 0.62, but there was no difference between angry and happy expressions for fist bump actions, t(32) = 0.76, p = .455, d = 0.13.

In summary, SCR to punch actions was increased when agents were displaying a happy compared to an angry facial expression, while SCRs to fist bump actions did not differ between facial emotions.

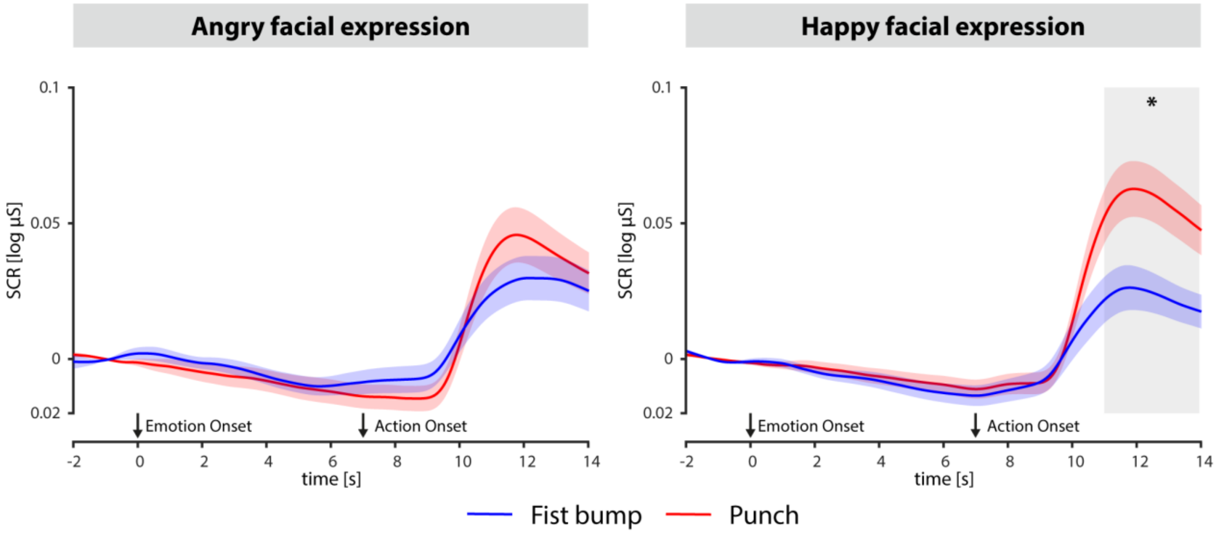

Figure 4: SCR following the onset of the emotional expression. Data from angry (left) and happy (right) facial expression conditions are shown in separate graphs. Actions of the virtual agent are color-coded (fist bump = blue, punch = red). Time windows

with significant differences between punch and fist bump conditions are highlighted in grey. Shaded areas reflect SEM.

2.3. Behavior: Reaction times

A 2 x 2 repeated-measures ANOVA on reaction times revealed a significant interaction between *Emotion* and *Action*, $F(1,31) = 9.68$, $p = .004$, $\eta_p^2 = 0.24$, but no main effect for *Emotion*, $F(1,31) = 0.09$, $p = .761$, $\eta_p^2 < 0.01$, or *Action,* $F(1,31) = 3.79$, $p = .061$, $\eta_p^2 = 0.11$. Post-hoc t-tests revealed that RTs were faster for punch compared to fist bump actions when the agents displayed an angry facial expression, $t(31) = -3.16$, $p = .021$, $d = -0.56$, but not when the agents displayed a happy facial expression, $t(31) = -0.39$, $p = .702$, $d = -0.07$. There were no significant differences between happy and angry facial expressions for fist bump actions, $t(31) = -2.60$, $p = .070$, $d = -0.46$, and punch actions, $t(31) = -2.16$, $p = .153$, $d = -0.38$.

In summary, angry facial expressions lead to faster responses to aversive punch action compared to appetitive fist bump actions, while reaction times of both actions did not differ when agents displayed a happy facial expression.

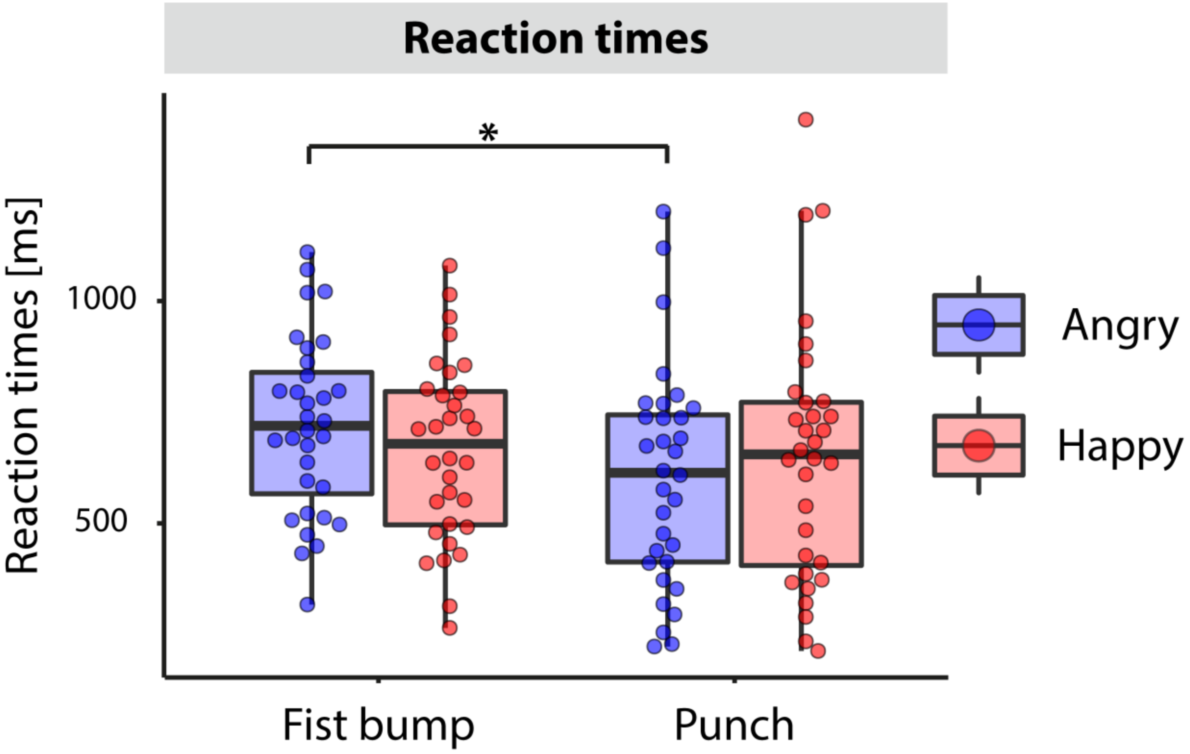

Figure 5: Reaction times of responses to the actions as a function of action and facial emotional expression in milliseconds. Box plots are superimposed with individual data points.

3. Discussion

Facial emotional expressions modulate the evaluation of face-to-face interactions and influence physiological and behavioral responses to social actions. In the present study we implemented an interactive Virtual Reality paradigm where participants responded to social actions of virtual agents. Virtual agents displayed either an angry or happy facial expression while directing aversive punch or appetitive fist bump actions towards the participant. In line with our hypotheses, angry expressions and punch actions were perceived as more arousing than happy expressions and fist bump actions. Crucially, pleasantness of an action was modulated by the accompanying facial expressions: fist bumps paired with a smile were perceived as pleasant, while the same fist bumps paired with angry expressions were perceived as unpleasant (even to a similar degree as punch actions). Furthermore, realism of interactions was evaluated on basis of the congruency between facial expression and action. Congruent expression-action pairs (both expression and action aversive or appetitive) were rated as more realistic compared to incongruent pairs (expression aversive and action appetitive or vice versa). Interestingly, physiological parameters were affected differently by facial expressions and actions. Heart rate showed a general effect of facial expression with an increase for angry compared to happy expression that was most prominent shortly before and during action initiation of the virtual agent. By contrast, skin conductance responses were affected by the interaction of facial expressions and actions. SCRs to aversive punch actions were increased when agents displayed a happy compared to an angry facial expression, while SCRs to appetitive fist bump actions did not differ between facial expressions. Finally, facial emotional expressions also influenced behavioral responses, i.e. reaction time. Participants responded faster to aversive compared to appetitive actions when actions were preceded by an angry but not when actions were preceded by a happy facial expression. Taken together, the present findings shed light on the interplay of facial emotional expressions and actions in social interactions. Our data provide evidence that observers use facial emotional expressions to generate expectations for actions and that these expectations affect the evaluation of social interactions as well as physiological and behavioral responses.

Previous studies have highlighted facial expressions as communicative cues that allow to infer mental states of others (Frith, 2009; Frith & Frith, 2012). Thereby observers can predict upcoming behavior, thus allowing for adaptive responses in social interactions (Sebanz & Knoblich, 2009). The results of the present study suggest that observers use facial emotional expression to build expectations regarding aversive or appetitive actions. In line with this notion, SCRs to aversive punch actions were increased when punches were following a happy compared to an angry facial expression. This suggests that aversive actions were unexpected for happy facial expression as indexed by a heightened physiological orienting response (Bradley, 2009). There was, however, no increased orienting response for unexpected appetitive actions (fist bump actions following angry expressions). This could be explained by differences in salience and threat imminence between fist bump and punch actions, as costs of an unexpected fist bump actions are lower than the costs of an unexpected punch action (Codispoti et al., 2001; Flykt et al., 2007). Furthermore, realism ratings regarding expression-action pairs showed that congruent pairs (both aversive/appetitive) were perceived as more realistic than incongruent pairs (one aversive, one appetitive). This suggests that observers integrated action expectations based on facial expressions with the actual performed actions. Note, however, that ratings can be only seen as indirect evidence for action expectations, as participants rated realism only after the full interaction had been presented and consequently might have based their evaluations on post-hoc processing. In sum, our results suggest an anticipatory mechanism of facial emotional expressions, where observers use facial emotional expressions to predict upcoming social actions.

The results of the present study further suggest an adaptive role of facial emotional expressions in interpersonal behavior. On a physiological level, we found increased heart rate responses for angry compared to happy facial expressions in a time-window prior and during the action initiation. Thus, observers showed increased physiological activation to the aversive facial expressions. One might speculate that the expectation of an aversive action increased sympathetic activity in order to prepare the organism towards a threatening action (Bradley et al., 2012; Löw et al., 2008). Heightened sympathetic activity during the observation of the action might be beneficial in generating an adaptive response. It should be noted, however, as we did not jitter the interval between facial expression and action, we cannot rule out the possibility that the effect of facial expression would occur with a timing of 5 – 9 seconds regardless of the onset of the action. This should be tested in future experiments. Interestingly, while we found a typical heart rate orienting response in form of a rapid deceleration following the onset of the action (Bradley, 2009; Bradley et al., 2012), we also observed increased rather than decreased heart rate for angry relative to happy facial expressions. Decelerated heart rate has been reported as a physiological response to threatening stimuli (freezing, e.g. Hagenaars et al., 2014). The observed acceleration in the present study may be explained by our study design, which allowed participants to actively react towards a threat action rather than to just passively observe it. As a consequence, given the option to react, heart rate increases might facilitate action responding in terms of a fight-or-flight response (Löw et al., 2008; Roelofs, 2017). In line with this idea, our data also show that facial emotional expressions affected behavioral response times. Importantly, angry facial expressions lead to faster responses to aversive punches than appetitive fist bumps, demonstrating a behavioral benefit of incorporating information from facial emotional expressions. In sum, our findings suggest that inferring action intentions from facial expressions increases sympathetic activity that results in a more effective response to social (threat-) actions. This may constitute an adaptive mechanism in real-time social interactions.

Finally, our data suggest that inferred intentions have a strong impact on the evaluation of social actions that may even override the valence of the actual performed action. Pleasantness of appetitive fist bump actions strongly differed between facial expressions, suggesting that the inferred intention influenced experience to a greater degree than the actual performed action. Facial expressions have been related to impression formation (Todorov et al., 2013). With respect to the current study, expectations based on facial expressions might be processed as true social intentions, thus altering the meaning of a given action. For the experience of punch actions, however, the influence of facial expressions was less prominent. Thus, angry expression can render an appetitive action as unpleasant, but happy expression cannot render an aversive action as pleasant to the same degree. One might speculate that fist bump actions may be less unequivocal with respect to being aversive/appetitive and thus be more affected by inferred intentions from facial expressions. In line with this interpretation, Kroczek et al. (2021) found that angry facial emotions biased observers towards aversive punch actions in a perception task using video clips of the same stimuli as in the present study. Importantly, this bias was strongest when actions were ambiguous. Together, these findings highlight the role of facial expressions for the inference of action intentions in social settings. Thus, in addition to cues from gaze (Ambrosini et al., 2015; Cavallo et al., 2015) and body posture (De Gelder, 2006; de Gelder & Poyo Solanas, 2021), facial expressions provide important information for upcoming actions.

In conclusion, the present study implemented a real-time interactive paradigm in Virtual Reality to investigate the influence of facial emotional expressions on social actions. Facial emotions had an impact on the evaluation of social actions and influenced physiological and behavioral responses. Consequently, facial emotional expressions are important cues in social interactions that allow to infer action intentions of an interactive partner and to generate adaptive responses.

4. Acknowledgements

We are thankful to Andreas Ruider for technical support and Julia Veitenhansl and Kevin Weber for data collection.

5. References

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437–1442. https://doi.org/10.1152/jn.00118.2011

Ambrosini, E., Pezzulo, G., & Costantini, M. (2015). The eye in hand: predicting others' behavior by integrating multiple sources of information. *Journal of Neurophysiology*, *113*(7), 2271–2279. https://doi.org/10.1152/jn.00464.2014

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), 1–17. https://doi.org/10.1111/spc3.12312

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251. https://doi.org/10.1017/S0021963001006643

Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition*, *106*(2), 894–912. https://doi.org/10.1016/j.cognition.2007.05.004

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

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology*, *46*(1), 1–11. https://doi.org/10.1111/j.1469-

8986.2008.00702.x

Bradley, M. M., Keil, A., & Lang, P. J. (2012). Orienting and emotional perception: Facilitation, attenuation, and interference. *Frontiers in Psychology*, *3*(NOV), 1–6. https://doi.org/10.3389/fpsyg.2012.00493

Cavallo, A., Lungu, O., Becchio, C., Ansuini, C., Rustichini, A., & Fadiga, L. (2015). When gaze opens the channel for communication: Integrative role of IFG and MPFC. *NeuroImage*, *119*, 63–69. https://doi.org/10.1016/j.neuroimage.2015.06.025

Codispoti, M., Bradley, M., & Lang, P. (2001). Affective reaction to briefly presented pictures. Psychophysiology. *Psychophysiology*, *38*, 474–478.

Csibra, G., & Gergely, G. (2007). ‘Obsessed with goals’: Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007

Curioni, A., Vesper, C., Knoblich, G., & Sebanz, N. (2019). Reciprocal information flow and role distribution support joint action coordination. *Cognition*, *187*(February), 21–31. https://doi.org/10.1016/j.cognition.2019.02.006

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., & Poyo Solanas, M. (2021). A computational neuroethology perspective on body and expression perception. *Trends in Cognitive Sciences*, *25*(9), 744–756. https://doi.org/10.1016/j.tics.2021.05.010

Ferstl, Y., Bülthoff, H., & Rosa, S. De. (2017). Action recognition is sensitive to the identity of the actor. *Cognition*, *166*, 201–206. https://doi.org/10.1016/j.cognition.2017.05.036

Flykt, A., Esteves, F., & Öhman, A. (2007). Skin conductance responses to masked conditioned stimuli: Phylogenetic/ontogenetic factors versus direction of threat?

Biological Psychology, *74*(3), 328–336. https://doi.org/10.1016/j.biopsycho.2006.08.004

Frith, C. D. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3453–3458. https://doi.org/10.1098/rstb.2009.0142

Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research*, *1079*(1), 36–46. https://doi.org/10.1016/j.brainres.2005.12.126

Frith, C. D., & Frith, U. (2012). Mechanisms of Social Cognition. *Annual Review of Psychology*, *63*, 287–313. https://doi.org/10.1146/annurev-psych-120710-100449

Hagenaars, M. A., Oitzl, M., & Roelofs, K. (2014). Updating freeze: Aligning animal and human research. *Neuroscience and Biobehavioral Reviews*, *47*, 165–176. https://doi.org/10.1016/j.neubiorev.2014.07.021

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Jack, R. E., & Schyns, P. G. (2015). The Human Face as a Dynamic Tool for Social Communication. *Current Biology*, *25*(14), R621–R634. https://doi.org/10.1016/j.cub.2015.05.052

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Kothe, C. (2014). *Lab Streaming Layer (LSL)*.

Kroczek, L. O. H., Lingnau, A., Schwind, V., Wolff, C., & Mühlberger, A. (2021). Angry facial expressions bias towards aversive actions. *PLOS ONE*, *16*(9), e0256912. https://doi.org/10.1371/journal.pone.0256912

Krüger, S., Sokolov, A. N., Enck, P., Krägeloh-Mann, I., & Pavlova, M. A. (2013). Emotion through locomotion: Gender impact. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0081716

Löw, A., Lang, P. J., Smith, J. C., & Bradley, M. M. (2008). Both predator and prey: Emotional arousal in threat and reward. *Psychological Science*, *19*(9), 865–873. https://doi.org/10.1111/j.1467-9280.2008.02170.x

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Pan, J., & Tompkins, W. J. (1985). A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, *BME-32*(3), 230–236. https://doi.org/10.1109/TBME.1985.325532

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

R Core Team. (2016). *R: A Language and Environment for Statistical Computing* (3.3.2). https://www.r-project.org/

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Roelofs, K. (2017). Freeze for action: Neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718). https://doi.org/10.1098/rstb.2016.0206

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition*, *119*(2), 242–252. https://doi.org/10.1016/j.cognition.2011.01.014

Schilbach, L., Eickhoff, S. B., Cieslik, E., Shah, N. J., Fink, G. R., & Vogeley, K. (2011). Eyes on me: An fMRI study of the effects of social gaze on action control. *Social Cognitive and Affective Neuroscience*, *6*(4), 393–403. https://doi.org/10.1093/scan/nsq067

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

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

Sosic, Z., Gieler, U., & Stangier, U. (2008). Screening for social phobia in medical in- and outpatients with the German version of the Social Phobia Inventory (SPIN). *Journal of Anxiety Disorders*, *22*(5), 849–859. https://doi.org/10.1016/j.janxdis.2007.08.011

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology*, *23*(3), 373–380. https://doi.org/10.1016/j.conb.2012.12.010

Torrubia, R., Avila, C., Molto, J., & Caseras, X. (2001). The sensitivity to punishment and sensitivity to reward questionnaire. *Personality and Individual Differences*, *31*(1), 837–862. http://linkinghub.elsevier.com/retrieve/pii/S0191886900001835%0Apapers3://pu blication/uuid/6653053D-BEAD-45B6-B558-11772D74A4AE

Author contributions:

LK, AL, CW, and AM designed the study, LK programmed the experiment, supervised data acquisition and analyzed data. VS and CW conducted motion tracking and supervised creation of the virtual environment. LK, AL, and AM wrote the paper.

Additional Information

Conflict of Interest: The authors declare no competing interests.

Funding: This study was supported by the “UR Fellows” program of the University of Regensburg. AL is supported by a Heisenberg Professorship (Li 2840/2-1) from the German Research Foundation.

Ethics statement: The study was reviewed and approved by the ethics board of the University of Regensburg and the study was conducted according to the approved procedures. The study is in line with the Declaration of Helsinki. All participants gave written informed consent.

Figure Captions

Figure 1: Schematic illustration of the experimental trial structure. Upper row shows trial procedure in Virtual Reality, bottom row shows participant inside the CAVE system at corresponding time points. At trial start, an elevator door opened and revealed a virtual agent (first column). Next, the agent displayed either a happy or angry facial emotional expression and then approached (second column) the participant until a final position was reached (third column). 2000 ms after reaching this position, the virtual agent initiated either a punch or fist bump action (fourth and fifth column respectively). Participants had to react towards this action with a congruent response, either by defending the punch (fourth column, bottom row) or by reciprocating the fist bump (fifth column, bottom row).

Figure 2: Subjective experience as a function of facial emotional expression and action of the virtual agent. Ratings on a scale from 0 to 100 reflect arousal (left), valence (middle), and realism (right). Box plots are superimposed with individual data points.

Figure 3: Effect of facial emotional expression on heart rate following the onset of the facial emotion. Time windows with significant differences between angry and happy facial expressions are highlighted in grey. Shaded areas reflect SEM.

Figure 4: SCR following the onset of the emotional expression. Data from angry (left) and happy (right) facial expression conditions are shown in separate graphs. Actions of the virtual agent are color-coded (fist bump = blue, punch = red). Time windows with significant differences between punch and fist bump conditions are highlighted in grey. Shaded areas reflect SEM.

Figure 5: Reaction times of responses to the actions as a function of action and facial emotional expression in milliseconds. Box plots are superimposed with individual data points.