ORIGINAL PAPER

Time to Smile: How Onset Asynchronies Between Reciprocal Facial Expressions Influence the Experience of Responsiveness of a Virtual Agent

Leon O. H. Kroczek[1] · Andreas Mühlberger[1]

Abstract

Human social interactions are highly coordinated in time and involve the reciprocal exchange of facial emotional expressions. While timing has been identified as an important factor in social communication via gaze or speech, so far there has been no systematic investigation on how the temporal delays in the exchange of facial emotional expressions between interactive partners influence the experience of responsiveness. In the present study, 40 participants were cued to direct a facial emotional expression (angry or happy) towards a virtual agent in front of them and the virtual agent would then react with a congruent facial expression. The temporal delay between the cue and the reaction was manipulated as an independent variable. Exact delays between facial expressions were determined based on the onset latencies of participants' facial expressions measured via facial EMG (*M. Zygomaticus* and *M. Corrugator)*. Ratings of experienced responsiveness of the virtual agent were collected as a dependent measure. Using a linear mixed effect model in a growth curve analysis revealed that experienced responsiveness peaked at delays around 700 ms. Moreover, experienced responsiveness at shorter temporal delays was higher when happy versus angry facial expressions were exchanged, while the reversed pattern was found at longer delays. Our results demonstrate a crucial role of timing in non-verbal communication, suggesting that temporal relations between facial expressions are processed as social information. Furthermore, our results can inform the implementation of virtual social interactions.

✉ Leon O. H. Kroczek
   leon.kroczek@ur.de

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

⚫ Springer

Introduction

Social interactions can be described as the mutual exchange between interactive partners and involve a large number of non-verbal behaviors, including facial expressions, eye gaze, and body movements (Frith & Frith, 2012). In real-time, face-to-face interactions these behaviors do not occur in isolation but always emerge in the dyadic exchange between interactive partners, for example by gaze following or by reacting to an emotional facial expression (Kroczek & Mühlberger, 2022; Pfeiffer et al., 2012). This reciprocity, where the behavior of one person results in changes in the behavior of another person, has been described as the defining feature of social interactions (Gallotti et al., 2017). The exchange of social behavior has also been discussed in terms of social agency, which describes the feeling of being the cause of another person's actions (Brandi et al., 2020; Silver et al., 2021). The experience of social agency in interactions has been linked to the responsivity of an interaction partner, i.e. the latency of another person's response, as well as the congruency between actions (Brandi et al., 2019). While reciprocity does not necessarily mean that the exchanged behavior is congruent, interactive behavior has been typically studied in terms of synchrony or mimicry (Chartrand & Lakin, 2013; Fischer & Hess, 2017). Synchronized non-verbal behavior also has implications on how we evaluate other persons and has been shown to result in a more positive social evaluation of an interactive partner (Lakin et al., 2003; Tarr et al., 2018). This highlights the role of reciprocity in defining interpersonal relations and coordinating social interactive behavior.

Despite the importance of reciprocal behavior in social interaction, the underlying mechanism remain largely unknown. This can be explained by the fact that social processes have been predominantly studied in settings where participants are passive observers of standardized social stimuli and a reciprocal exchange of behavior is not possible (Becchio et al., 2010; Redcay & Schilbach, 2019). Fortunately, an increasing number of studies have been conducted to investigate real-time face-to-face interactions in dyads (Heerey & Crossley, 2013; Hess & Bourgeois, 2010; Lahnakoski et al., 2020; Riehle et al., 2017). These studies have a high ecological validity and confirm that the experience of social interactions and the evaluation of interactive partners is closely linked to non-verbal communicative behavior. Furthermore, new paradigms have been established that combine Virtual Reality with the online measurement of behavior in a closed-loop so that participants' behavior can be used to elicit behavior in virtual agents (Kroczek et al., 2020; Pfeiffer et al., 2012; Tarr et al., 2018; Wilms et al., 2010). These interactive virtual settings provide high experimental control and allow the systematic manipulation of social interactions in order to reveal underlying mechanism (Hadley et al., 2022).

Previous studies in interactive settings have highlighted the role of temporal dynamics in social interactions. Heerey and Crossley (2013) examined the temporal delay in the exchange of smiles in real dyads engaging in natural conversations. They found a median temporal delay of 780 ms for genuine smiles with a large proportion of temporal delays being shorter than 200 ms. Interestingly, a different pattern was observed for polite smiles for which temporal delays were in general later and which were also less frequently shorter than 200 ms. The fast response times for genuine smiles have been interpreted in terms of an anticipatory processing that is based on the content of the interaction. Similar findings were reported in another study where synchrony in corresponding facial EMG signals (Zygomaticus and Corrugator) was measured between two interactive partners engaging in a conversa-

tion task (Riehle et al., 2017). In line with previous results the authors found synchrony for smiles for time lags within 1000 ms with a major proportion of synchronization below 200 ms. Moreover, smiling lead to larger synchrony between interactive partners than frowning which might be related to the affiliative function of smiling. Temporal dynamics of facial emotional expressions have also been studied in children with autism spectrum disorder. In these children, emotional mimicry was found to be temporally delayed compared to a control group (Oberman et al., 2009). While temporal effects have not been studied in social anxiety, there is evidence that emotional mimicry in general might be altered. For instance high compared to low social anxious participants were found to show increased mimicry for polite but not genuine smiles (Dijk et al., 2018; Heerey & Kring, 2007), suggesting that social anxious persons might use mimicry in order to avoid conflicts. Overall, previous findings suggest an important role of temporal dynamics in the reciprocal exchange of facial emotional expressions.

It should be noted, however, that previous studies mostly used descriptive approaches to characterize temporal dynamics in the exchange of facial emotional expressions, but there was no manipulation of temporal delay that would allow to investigate the influence of temporal delay on the experience of social interactions. The latter approach, however, has been used in a previous study relating to gaze following (Pfeiffer et al., 2012). Here, the temporal delay between participants' eye gaze and the subsequent gaze following of a virtual agent was manipulated and participants were asked to rate the degree of the relatedness of the agent's gaze response. Interestingly, experience of relatedness peaked at temporal delays between 400 and 800 ms while immediate responses (no delay) were experienced as less related. This finding suggests that observers have clear expectations about the temporal dynamics of social reciprocal behavior and deviations from these predictions affect the interpretation of social signals. It remains unknown, however, whether similar mechanisms take place in the processing of facial emotional expressions and whether this is influenced by the valence of a facial emotional expression.

Therefore, the current study was conducted with the goal to investigate whether the temporal delay between one's own facial emotional expression and the facial emotional expression of a virtual agent influences the degree to which the agent's facial emotional expression is perceived as a reaction to oneself. This measure was implemented to quantify how strongly participants related an expression of the virtual agent to their own expression. The paradigm presented facial expressions with differing valence, namely smiles and frowns. Study procedures and hypotheses were pre-registered before the start of data acquisition (https://osf.io/7yzb4/). First, we expected to find a main effect of temporal delay between facial expressions on the experience of responsiveness of a virtual agent. More specifically, we hypothesized that intermediate delays (500–1000ms) would lead to higher ratings of experienced responsiveness compared to short (<500ms) or longer (>1000ms) delays. In addition, we hypothesized an interaction effect between temporal delay and valence of facial emotional expressions (angry or happy). On the one hand, happy expressions are more common in social interactions (Hess & Bourgeois, 2010) and are typically reciprocated within very short delays (Heerey & Crossley, 2013; Riehle et al., 2017) due to anticipatory processing. On the other hand, angry facial expressions, while less common, are salient social cues that signal threat and missing such a cue might lead to aversive consequences (Kroczek et al., 2021). Therefore, we expected that due to participants experience with timing of happy facial expressions in real-life social interactions, happy compared to

angry facial expressions would result in an increased experience of responsiveness of the virtual agent for short temporal delays (< 500ms), whereas angry compared to happy facial expressions would result in an increased experience of responsiveness of the virtual agent for longer temporal delays (> 1000 ms). In addition, an exploratory analysis was conducted to test for correlations between individual differences in the relation of temporal delays and experienced responsiveness of the virtual agent and symptoms related to social anxiety and autism.

Materials and Methods

Participants

Forty healthy volunteers participated in the study (33 female, M_{Age} = 21.75 years, SD_{Age} = 2.44, range$_{Age}$ = 18–28 years, 95% university students). Participants were recruited at Regensburg University and via social media. All participants had normal or corrected-to-normal vision and did not report any mental or neurological disorder. Experimental procedures were 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. Students enrolled in the Regensburg University psychology program were given course credit for compensation.

Study Design

The study was implemented as a within-subject design with the experimental factors *facial emotional expression* and *temporal delay.* The *experience of responsiveness of the virtual agent* was measured as dependent variable via ratings. Participants were asked to rate how strong they felt that a virtual agent reacted towards them, i.e. how strongly they relate the agents expression to their own expression. *Facial emotional expression* was manipulated as the exchange of either angry or happy facial expressions between the participant and the agent. Note, that facial emotional expressions were always exchanged in a congruent manner (happy expressions followed happy expressions and vice versa for angry expressions). *Temporal delay* was manipulated as the time interval between the onset of the cue that prompted participants to display the facial emotional expression and the onset of the facial emotional expression of the virtual agents. Five different temporal delays were implemented: no delay (0 ms), 500 ms delay, 1000 ms delay, 1500 ms delay, and 2000 ms delay. Importantly, the goal of the experimental manipulation of temporal delay as described above was to introduce variability in time differences between the onsets of participants' facial expressions and the onset of agents' facial expressions that was approximately related to the five different delays. However, we expected that the onset of a participant's facial expression would not be exactly synchronized to the onset of the cue (e.g. due to processing times and attentional effects). Therefore, we identified the actual onset of the participants' facial expressions by analyzing the continuously recorded EMG (see below) and then used these onsets to determine the actual temporal delay between facial expressions that was entered in the statistical analyses (see Fig. 1). This procedure was successful in introducing a wide range of temporal delays between onsets of facial expressions (distribution of temporal delays are summarized

in the supplementary materials, Table S1). Please note, that a late response of the participant could result in a trial where the expression of the virtual agent preceded the expression of the participant. There was no significant difference between participant's average response times for happy ($M = 234$ ms, $SD = 164$ ms) and angry expressions ($M = 219$ ms, $SD = 175$ ms), $t(39) = -0.57$, $p = .570$. Finally, as an attentional control condition, we also implemented trials where the agents did not respond with a facial emotional expression. Responsiveness ratings in these trials were near zero ($M = 3.23$, $SD = 8.75$) suggesting that participants paid attention to the facial expression of the virtual agents.

Stimulus Material

Short video clips of four different virtual agents were presented as stimulus material. Virtual agents (two females, two males) were created using MakeHuman (v 1.1.1, www.make-human.org). These agents were then animated using Blender (v2.79, Blender Foundation, Amsterdam, Netherlands). Two emotional expressions, happy and angry, were implemented in accordance to the facial action coding system (Ekman & Friesen, 1978). Expressions were identical across all virtual agents. In order to increase liveliness and naturalness, virtual agents were animated to show eye blinks and slight head motion. Animations of eye blinks and head motion differed between virtual agents but were identical across emotional expressions and have been used in a previous study (Kroczek & Mühlberger, 2022). Video stimuli were rendered with 60 fps with different lengths: 3000 ms, 3500 ms, 4000 ms, 4500 ms and 5000 ms. In all video clips, agents displayed a neutral facial expression in the initial 1000 ms. The onset of the facial expression was initiated after another delay of 0 ms, 500 ms, 1000 ms, 1500 ms, or 2000 ms (relating to the different levels of manipulated temporal delay). The neutral expression changed within 500 ms to an emotional expression (happy or angry) and remained in that expression for another 1500 ms. In the control condition the agents remained in with a neutral expression for a total of 3000 ms. In sum, a set of 2 (emotion: angry, happy) x 4 (agents: 2 male, 2 female) x 5 (length: 3000, 3500, 4000, 4500, 5000 ms) + 4 baseline = 64 video clips was used in the study.

Procedure

Before the start of the experiment, participants received instructions about the procedure of the experiment. They were instructed to interact with the virtual agents in front of them by directing facial emotional expressions at the agents once a cue was presented on the screen and that the agents would then react to them. For EMG measurements, electrodes were attached to the face (see below). Participants were seated in front of a 21.5-inch LCD-screen (HP E221c, 1920×1080 resolution, 60 Hz) with a distance of 50 cm.

Stimulus presentation was controlled using Psychtoolbox-3 (Pelli, 1997) implemented in Matlab 8.6 (MathWorks, Natick, MA, USA). A schematic overview of the trial structure is displayed in Fig. 1. Trials started with the presentation of a fixation cross for 1000 ms. Next, participants were instructed about the emotional expression they had to direct at the agent. For that reason, the emotion noun was presented on the screen for 2000 ms (i.e. *Happiness* or *Anger*). After another fixation cross had been displayed for 1000 ms, the video clip was presented in the center of the screen (video size on screen: 1519×854). Video clips started with the display of a virtual agent showing a neutral facial expression for 1000 ms. Then,
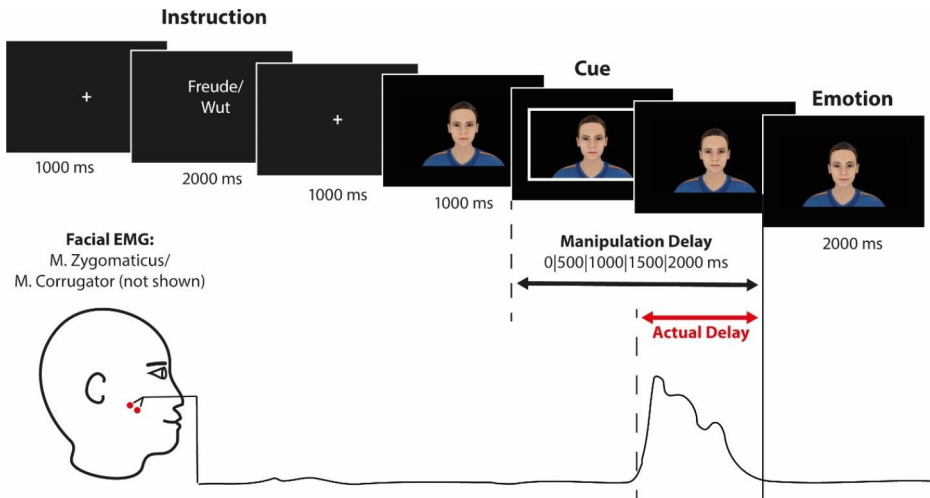
Fig. 1 Experimental procedure. Top row shows the trial structure. Participants were first instructed about the facial emotional expression they had to display later in the trial (happy or angry). Then a virtual agent appeared on the screen displaying a neutral expression. After 1000 ms, a white rectangular frame appeared around the agent, which served as a cue for participants to now direct the facial expression at the virtual agent. Following a variable delay (experimental manipulation) between 0 and 2000 ms the agent would then return a congruent facial expression. Bottom row demonstrates how the exact temporal delay between facial expressions of the participant and facial expression of the agent was measured. Continuous EMG was recorded at M. Zygomaticus and M. Corrugator. Data were offline preprocessed and the onset of participants' smiles and frowns were identified (dashed line). These onsets were then used to calculate the actual temporal delay between facial expressions. In summary, while the time period between the cue (rectangular frame) and the onset of the expression of the virtual agent was experimentally manipulated, we defined the actual temporal delay as the time period between the onset of participant's facial expression (measured vie EMG) and the onset of the expression of the virtual agent

a white rectangular frame appeared around the video, serving as a cue for participants to direct the instructed emotional expression at the virtual agent. The cue had a duration of 500 ms. Depending on the experimental condition, agents' facial expressions would change to an emotional expression after 0 ms, 500 ms, 1000 ms, 1500 ms or 2000 ms post cue onset or would remain in a neutral expression (control condition). The transition from neutral to an emotional expression had a duration of 500 ms. Agents then displayed the emotional expression for another 1500 ms until the end of the video clip. In the baseline conditions the facial expression of the agents did not change but remained neutral for a total of 2000 ms following the onset of the cue.

In every trial participants were asked to rate how strongly they felt that the agents reacted towards them on a scale form 0 (not at all) to 100 (very strongly). Participants had to enter the rating on a visual analog scale using the computer mouse. There was no time limit for the response.

In total 160 trials were presented, including 20 trials (5 per agent) for each combination of *facial emotional expression* and *manipulated temporal delay* and 20 baseline trials (5 per agent). Trial order was pseudorandomized with no more than 3 repetitions of facial emotional expression or temporal delay.

Data Acquisition and Processing

Questionaires

Participants filled in the German version of the social phobia inventory (Connor et al., 2000; Sosic et al., 2008) and the short version of the German adaptation of the Autism Quotient (Baron-Cohen et al., 2001; Freitag et al., 2007). The social phobia inventory (SPIN) includes 17 items that assess symptoms of social anxiety in the previous week. Answers are given on a five-point Likert scale ranging from 0 (not at all) to 4 (extremely). A cut-off score of 19 has been suggested to be indicative of social anxiety (Connor et al., 2000). The short version of the Autism Quotient (AQ-k) includes 33 items assessing agreement with statements describing autistic traits. Answers are given on a four point Likert scale ranging from 0 (no agreement) to 4 (full agreement). A score of 17 has been suggested as a cut-off value for potentially clinically significant symptoms (Freitag et al., 2007). Sum scores were calculated for each questionnaire and entered into statistical analysis.

Physiology: Electromyography

Facial EMG was measured at the *M. zygomaticus major* (Zygomaticus) and the *M. corrugator supercilii* (Corrugator). For each muscle, two 8 mm Ag/AgCl electrodes were attached to the surface of the skin. Before electrode attachment, skin was prepared using alcohol and an abrasive paste (Skin-Pure, Nihon Kohden, Tokio, Japan). Impedances were kept below 50 kOhm. Electrode positions followed the guidelines by Fridlund and Cacioppo (1986) with the ground electrode on the center of the forehead (Fridlund & Cacioppo, 1986). Data was sampled with 1000 Hz using a V-Amp amplifier (BrainProducts, Gilching, Germany).

Data preprocessing was conducted in Matlab 8.6 (MathWorks, Natik, MA; USA). First, the two electrodes of each muscle were re-referenced to each other. Next, a bandpass filter between 30 and 500 Hz and a notch filter of 50 Hz were applied. Data were then rectified and integrated using a moving average with a window size of 125 ms. Data were then segmented around the onset of the facial emotional expression of the virtual agent (5000 ms pre onset, 2000 ms post onset). In a next step, an experimenter (blinded to the experimental condition) manually marked the onset of the EMG response by identifying the point where a steep rise of the EMG signal could be observed that indicated the onset of a peak in EMG activity. Onsets of happy expressions were defined on basis of the signal in the Zygomaticus and onsets of angry expressions were defined on basis of the signal in the Corrugator. For a subset of 576 trials (~ 10%) we obtained additional onset markings of a second independent rater and found that onset markings between raters differed with a root mean square of 33 ms ($SD = 88$ ms), suggesting a high degree of interrater agreement.

Finally, the temporal difference between the onset of the facial expression of the participant and the onset of the facial expression of the virtual agent were exported for each trial and participant. Trials where no EMG activation peak could be visually detected or the wrong muscle was activated were excluded from further analysis (mean number of trials rejected = 4.46, $SD = 7.83$). Finally, we removed latencies that were below the 1% or above the 99% quantiles of all data points (94 trials removed). Note, that this was a deviation from the pre-registered analysis plan that was implemented to reduce the influence of extreme

data points in the growth curve analysis (Dedrick et al., 2009). The full dataset and the corresponding model are plotted in the supplementary material (Figure S5).

Statistical Analyses

Statistical analysis was conducted in the R environment (v 4.1.1). Data were analyzed using linear mixed effect model as implemented in the lme4 package (Bates et al., 2015). Responsiveness ratings were modelled by including the main effect of *facial emotional expression* as a fixed effect (coding: angry = 0, happy = 1). In addition, main effects of *temporal delay* between facial expressions of participants and agents were analyzed using a growth curve approach with fixed effects being modelled as orthogonal, third-order polynomials relating to linear, quadratic, and cubic effects (Mirman et al., 2008). Finally, interaction effects between *facial emotional expression* and the linear, quadratic, and cubic effects of *temporal delay* were also entered as fixed effects in the model. Likelihood ratio tests were conducted to determine the random effects structure. The final model included random intercepts per participant and random slopes per participant for the facial emotional expression, as well as linear, quadratic, and cubic effects of temporal delay and the interaction between the linear effect of temporal delay and facial emotional expression. Main effects and interactions were evaluated using F-tests with Satterthwaite approximations for degrees of freedom (Luke, 2016).

An exploratory analysis was conducted to test for the association between the individual relation of temporal delay and experienced interactivity and scores in questionnaires relating to social phobia (SPIN) and autism (AQ-k). For that reason, we used the linear mixed-effect model that resulted of the main analysis to describe the individual relation between temporal delays and the experienced responsiveness of the virtual agent for each participant (see Supplementary Material Figure S1). Based on these individual models we extracted latencies at which experienced responsiveness of the virtual agent was maximal and used Pearson's correlations to test the associations between extracted latencies and SPIN or AQ-k scores respectively. Tests were corrected for multiple comparisons according to Holm (Holm, 1979).

Open Science Statement

Study procedures, hypotheses, and statistical analyses were pre-registered prior to data acquisition. Note, that EMG data analysis as well as the marking of the onset of the EMG response was not pre-registered. All study materials including anonymized raw data, analysis scripts, and stimulus materials are publicly available in an online repository (https://osf.io/7yzb4/).

Results

Modelling Experience of Responsiveness of the Virtual Agent

Participants' experience of responsiveness of the virtual agent was modelled using predictors for *facial emotional expression*, the linear, quadratic and cubic effect of *temporal delay*,

Table 1 Model summary for linear mixed effect model with formula: Responsiveness Rating~Facial Emotional Expression* (Delay+Delay^2+Delay^3)+(1+Facial Emotional Expression * Delay+Delay^2 +Delay^3|Subject). Model included trials with negative temporal delays (see Supplementary Material Table S2 for model without negative temporal delay trials). Facial emotional expression contrast coding: angry=0, happy=1

Predictors	Estimates	SE	95% CI
Intercept	63.04	1.80	59.40–66.69
Facial Expression	-0.17	0.86	-1.92–1.57
Delay	69.29	39.92	-11.40–149.98
Delay^2	-209.78	36.08	-282.54 – -137.02
Delay^3	130.84	23.68	83.28–178.39
Facial Expression x Delay	-48.76	21.55	-91.93 – -5.59
Facial Expression x Delay^2	-1.64	15.16	-31.36–28.07
Facial Expression x Delay^3	0.61	15.23	-29.25–30.47

as well as all interactions between *facial emotional expression* and *temporal delay* (linear, quadratic, cubic). The linear mixed effect model (Table 1; Fig. 2) revealed a significant quadratic effect of temporal delay, $F(1,38.8)=35.93$, $p <. 001$, a significant cubic effect of temporal delay, $F(1,39.8)=34.54$, $p<.001$, and a significant interaction between facial emotional expression and the linear effect of temporal delay, $F(1,55.8)=5.12$, $p=.028$. The model revealed that the experience of responsiveness peaked at a temporal delay of 705 ms when happy expressions were exchanged and 752 ms when angry expressions were
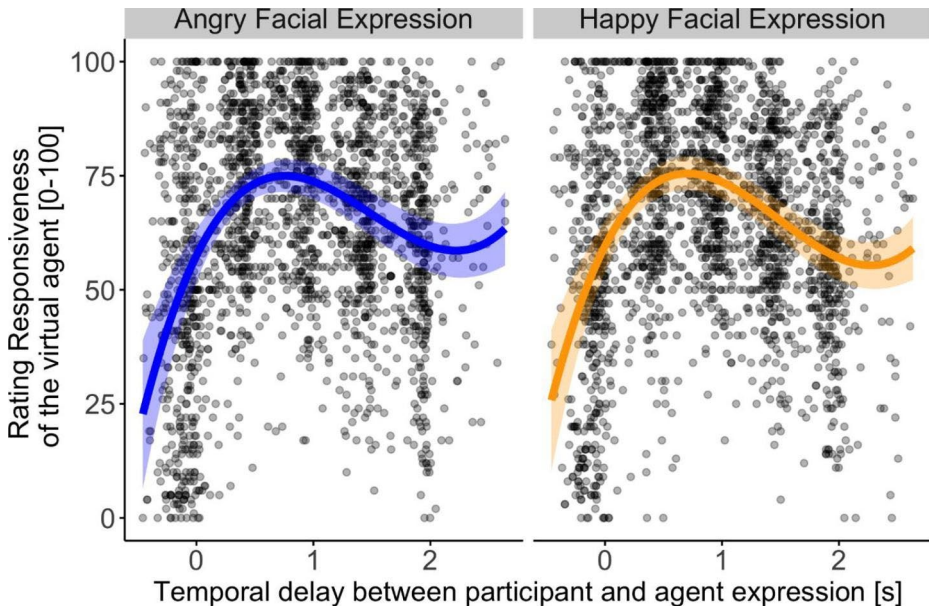

Fig. 2 Model fit on individual trial data. Relation of temporal delay between facial expressions (x-axis) and ratings of responsiveness of the virtual agent (y-axis) for the exchange of angry (blue) and happy facial expressions (orange). Single trial data points are overlaid with the model fit of the linear mixed effect models. Shaded areas reflect the 95% confidence interval. A direct comparison of model fits is presented in the supplementary material Figure S4

exchanged. The quadratic effect (b = -209.78, $SE=36.08$) demonstrates that experience of responsiveness increased with longer temporal delays until the peak was reached and then decreased again with longer temporal delays. In addition, the cubic effect (b=130.84, $SE=23.68$) of temporal delays demonstrates that experience of responsiveness did not return to zero at longer temporal delays.

A post-hoc analysis was conducted to follow-up on the interaction effect (b= -48.76, $SE=21.55$). The linear effect of temporal delay on experienced responsiveness was compared between angry (b=69.3, $SE=40.2$) and happy (b=20.5, $SE=35.6$) facial emotional expressions. Contrasting the model slopes showed a more positive linear effect of temporal delay in the angry compared to the happy condition, $t(54.9)=2.215$, $p=.031$. This difference in slopes demonstrates increased experienced responsiveness for the exchange of happy compared to angry facial expressions at short temporal delays and increased experience of responsiveness for the exchange of angry compared to happy facial expressions at long temporal delays (see Supplementary Material Figure S3 for an illustration of the interaction effect). This finding was supported by an additional analysis of the individual latencies where responsiveness ratings were maximal (peak latencies). A paired t-test (one-sided) showed that peak latencies for happy facial expressions were significantly earlier than for angry facial expressions, $t(39) = -1.718$, $p=.047$, d = -0.27.

Please note that the pre-registered analysis reported above also included trials with negative temporal delays between the expression of the participant and the expression of the virtual agent, meaning that in these trials, the facial expression of the virtual agent preceded the expression of the participant. As participants were asked to rate how strong they felt that the agents reacted to them, these trials might have induced a different evaluation of responsiveness of the virtual agent. Therefore, we conducted an additional (not pre-registered) analysis by testing the same model as specified above on a dataset where negative temporal delays were excluded. Importantly, the linear mixed effect model revealed similar results. There was a significant quadratic effect of temporal delay, $F(1,36.61)=11.044$, $p=.002$, a significant cubic effect of temporal delay, $F(1,36.39)=41.70$, $p<.001$, and a significant interaction between facial emotional expression and the linear effect of temporal delay, $F(1,60.88)=8.033$, $p=.006$. In contrast to the model including negative temporal delays, the new model also revealed a significant linear effect of temporal delay, $F(1,39.49)=13.13$, $p<.001$, and a significant interaction between facial emotional expression and the cubic effect of temporal delay, $F(1,865.33)=6.57$, $p=.011$. The model identified peaks of experienced responsiveness of the virtual agent at a temporal delay of 679 ms for happy expressions and at 734 ms for angry expressions. In line with the previous model that included all temporal delays (see above) these results confirm an inverted U-shape relation between temporal delay and experienced responsiveness of the virtual agent and show that this influence is not driven by negative delays (see also Table S2 and Figure S4 for a complete model summary). Excluding negative temporal delays, however, revealed that short temporal delays were generally rated with higher responsiveness than longer temporal delays.

Overall, the present data provide evidence that the experience of responsiveness of the virtual agent can be modelled as a quadratic and cubic function of the temporal delay between facial expressions. Furthermore, differential effects for the exchange of happy or angry facial expressions were found, with happy expressions relating to higher responsiveness at shorter, and angry expression relating to higher responsiveness at longer temporal
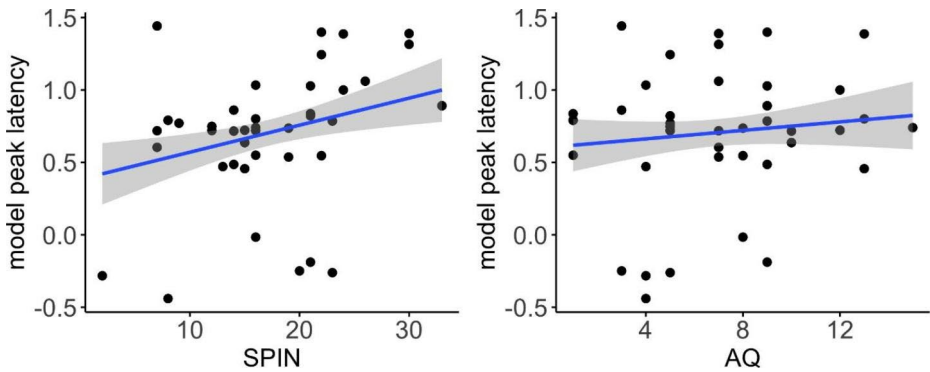
Fig. 3 Association between Social Phobia Inventory (left) and Autism Quotient (right) scores and latencies relating to the individual model peak in experienced responsiveness of the virtual agent. Blue line shows linear fit to the data

delays. These results were also observed when trials where the virtual agent's expression preceded the expression of the participant were excluded.

Correlation Between Individual Model Parameters and Social Anxiety and Autism

Additional exploratory analyses were conducted to investigate whether individual model parameters, i.e. latencies at which the individual models showed maximal ratings of responsiveness of the virtual agent (model peak latency), were correlated to scores in questionnaires assessing symptoms of social anxiety and autism (Fig. 3). These analyses revealed a marginal significant positive association between individual peak latencies and SPIN scores, $r(38) = 0.35$, $p = .055$. There was no significant correlation between individual peak latencies and AQ-k scores, $r(38) = 0.16$, $p = .316$.

Discussion

The present study investigated the influence of temporal delay between sending and receiving a facial emotional expression on the experience of responsiveness of a virtual agent in a virtual face-to-face interaction. In line with our first hypothesis, we found that the temporal delay between facial expressions influenced experienced responsiveness of the virtual agent. Ratings of responsiveness peaked at latencies between 700 and 750 ms with both shorter and longer delays than peak latency being experienced as less responsive to one's own facial expression. Furthermore, we found that the relation between temporal delay and experienced responsiveness of the virtual agent differed according to the valence of facial emotional expression. The exchange of happy compared to angry facial emotional expressions resulted in an increased experience of responsiveness at shorter temporal delays, while the reversed pattern was observed at longer temporal delays, confirming our second hypothesis.

The observed effects of temporal delay on the experience of responsiveness of the virtual agent in reciprocal social interactions most likely reflects participants experiences in real-time face-to-face interactions. The present study revealed that the experience of responsiveness peaked around 700–750 ms, which is very close to the median temporal delay in

the exchange of facial expressions reported for real-life, face-to-face dyadic interactions (Heerey & Crossley, 2013). This suggests that a persons' experience in social interactions is used as an expectation for temporal effects in the exchange of facial expressions and that these expectations do also hold for interactions with virtual agents. Similar effects have been found for temporal delays in gaze following (Pfeiffer et al., 2012). In contrast to the present results, previous studies (Heerey & Crossley, 2013; Riehle et al., 2017) found that a large proportion of temporal delays occurred within 200 ms. It should be noted, however, that these studies measured facial expressions during real-time conversation between two interactive partners. Information within the conversations might have promoted anticipatory processes that resulted in short temporal delays, while no such information was available in the present experimental paradigm. Instead, when no anticipatory processing is possible, a reciprocal reaction requires additional processing time starting from the actual onset of the initial expression. A reaction that occurs before or during this processing time window, may be seen as unlikely as a response to one's own expression. Future studies should include contextual information to investigate temporal dynamics in the exchange of facial expressions with respect to anticipatory processing (that is attributed to the interactive partner). Overall, the current findings demonstrate that persons are sensitive to temporal information in the exchange of facial emotional expressions. Temporal delays around 700 ms evoked the highest degree of responsiveness in face-to-face interactive setting with a virtual agent conforming data from real interacting dyads. Temporal information in the reciprocal exchange of facial expression therefore might be evaluated with respect to the probability of observing such a delay in an everyday interaction.

In the present study we manipulated the temporal delay between expressions by cueing the participant to show an expression and then varying the time between the cue and the expression of the agent. The onset of the participant's expression, however, was defined on basis of the EMG signal. In some trials this resulted in "negative temporal delays", i.e., when the expression of the virtual agent occurred before the expression of the participant. As participants were instructed to rate how strong they felt that the agent's reacted towards them, such expressions of the virtual agents without an initial expression of the participant, might have influenced the results. Importantly, however, even when these negative temporal delays were excluded from the analysis, we observed the same inverted U-shape relation between temporal delay and experienced responsiveness of the virtual agent. This suggests that even though negative temporal delays led to a reduced experience of responsiveness this did not influence the general relation of timing between expressions and perceived responsiveness. It should be noted, however, that the present paradigm used a pre-defined experimental procedure with a focus on initial expressions of the participants. Future studies should investigate timing in a more flexible trial structure to allow for initial responses of the virtual agent and test whether the same effects of temporal delay can be observed.

In addition to the main effect of temporal delay we observed a differential effect of temporal delay for the experience of responsiveness of the virtual agent in the exchange of happy versus angry facial emotional expressions. For happy expressions responsiveness was experienced higher at short temporal delays, whereas for angry expressions responsiveness was experienced higher at longer temporal delays. On the one hand, this might reflect the different communicative functions of happy and angry facial expressions. Happy facial expression are likely to indicate affiliative intent (Hess et al., 2000), while angry facial expressions indicate threat (Lundqvist et al., 1999). Consequently, it might be adaptive to

relate an angry expression of another person to oneself even when there is a longer delay between expressions in order to prepare for a potential attack. This is in line with the *anger superiority effect* that describes processing advantages for angry faces (Gong & Smart, 2021; Hansen & Hansen, 1988). Preparing for an upcoming threat might be especially important in an interactive setting. On the other hand, given that the exchange of smiles has been observed at very short temporal delays (Heerey & Crossley, 2013; Riehle et al., 2017), it might be the case that participants experienced temporal delays as more natural for happy compared to angry facial expressions which resulted in increased ratings of responsiveness of the virtual agent. It should be noted, however, that there was no general difference in the level of responsiveness between happy and angry emotional expressions. Overall, these data show that emotional content and motivational tendencies seem to affect the influence of temporal delays in reciprocal non-verbal behavior during social interactions.

We also conducted an exploratory analysis to investigate whether symptoms related to autism or social anxiety might influence the relation of temporal delays and the experienced responsiveness of the virtual agent. There was no significant relation between AQ-k scores and the temporal delay at which experienced responsiveness peaked. However, this result should be treated with caution, as there was only small variability of AQ-k scores in our sample and no score exceeded the cut-off of 17 (Freitag et al., 2007). With respect to social anxiety our sample included more variability (Connor et al., 2000). Here, we observed a non-significant trend towards a positive relation between SPIN scores and temporal delay of experienced peak responsiveness. In line with previous findings on the processing of facial emotions in social anxiety disorder (Dijk et al., 2018; Mühlberger et al., 2009; Staugaard, 2010), but needed to be confirmed in a replication trial, this might indicate altered processing of the temporal dynamics in social interactions in persons with high social anxiety. This could be related to attentional biases (McTeague et al., 2018) or increased self-referential processing in social anxiety (Abraham et al., 2013). Overall, the present data point towards altered processing of temporal dynamics of social interactions in social anxiety. Future studies using clinical samples are required to confirm such effects for social anxiety disorder.

To our knowledge this study is the first to systematically investigate the influence of temporal delay in reciprocal facial expressions on the experience of responsiveness. However, there are some limitations that need to be mentioned. First, there was no real social interaction between participants and virtual agents. Participants were cued to show a facial expression in a trial-wise manner and the responses of the virtual agents were pre-defined in the experimental procedure and did not depend on the actual behavior of the participants. There were also no other forms of interaction or communication. Interestingly, despite this lack of real interaction, temporal delay still modulated the experience of responsiveness of the virtual agent, suggesting that even the pseudo-interactive setting allowed participants to relate an expression of the virtual agents to their own expression. However, the lack of additional context cues might have prevented anticipatory processing of facial expressions (Heerey & Crossley, 2013). It should also be noted that participants were confronted with virtual agents rather than real persons. While virtual agents have been found to elicit similar social responses as have been observed for real persons (Weyers et al., 2006), this might have affected participants belief in intentionality of the virtual agents (Brandi et al., 2019). Future studies should implement more natural social settings, for instance in Virtual Reality, and provide additional contextual cues. Another limitation of our study is that the exchange of facial expressions was not related to any behavioral consequences. In real life, facial

expressions might signal actions that require the preparation of adaptive responses (Kroczek et al., 2021). Without behavioral consequences, facial expressions might be less relevant. This could be investigated by measuring the effects of temporal dynamics when behavioral consequences of differing valence are coupled to the exchange of facial expressions. Finally, it should be acknowledged that the sample acquired in the present study mostly consisted of young adult, female academics. Future studies should include more diverse samples to increase generalizability.

Social interactions are highly coordinated in time. By investigating face-to-face interactions with a virtual agent, we found that temporal dynamics in the exchange of facial expressions influenced the experienced responsiveness of the virtual agent and that this effect was modulated by valence of the facial expression. These results highlight temporal dynamics as important information during face-to-face social interactions. Finally, our results can be taken as a reference to optimize the experience of social interactions with virtual agents in the field of human-computer interaction.

Declarations

Ethics approval and consent to participate Experimental procedures were 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.

Competing interests The authors declare no competing interests.

References

Abraham, A., Kaufmann, C., Redlich, R., Hermann, A., Stark, R., Stevens, S., & Hermann, C. (2013). Self-referential and anxiety-relevant information processing in subclinical social anxiety: An fMRI study. *Brain Imaging and Behavior*, *7*(1), 35–48. https://doi.org/10.1007/s11682-012-9188-x.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger Syndrome/High-Functioning autism, males and females, scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

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.

Brandi, M. L., Kaifel, D., Bolis, D., & Schilbach, L. (2019). The interactive self – a review on simulating social interactions to Understand the Mechanisms of Social Agency. *I-Com*, *18*(1), 17–31. https://doi.org/10.1515/icom-2018-0018.

Brandi, M. L., Kaifel, D., Lahnakoski, J. M., & Schilbach, L. (2020). A naturalistic paradigm simulating gaze-based social interactions for the investigation of social agency. *Behavior Research Methods*, *52*(3), 1044–1055. https://doi.org/10.3758/s13428-019-01299-x.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology*, *64*(1), 285–308. https://doi.org/10.1146/annurev-psych-113011-143754.

Connor, K. M., Davidson, J. R. T., Churchill, E., Sherwood, L., Foa, A., E., & Weisler, R. H. (2000). Psychometric properties of the social phobia inventory (SPIN). New self-rating scale. *British Journal of Psychiatry*, *176*(APR.), 379–386. https://doi.org/10.1192/bjp.176.4.379

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of Methodological issues and applications. *Review of Educational Research*, *79*(1), 69–102. https://doi.org/10.3102/0034654308325581.

Dijk, C., Fischer, A. H., Morina, N., van Eeuwijk, C., & van Kleef, G. A. (2018). Effects of social anxiety on emotional mimicry and contagion: Feeling negative, but smiling politely. *Journal of Nonverbal Behavior*, *42*(1), 81–99. https://doi.org/10.1007/s10919-017-0266-z.

Ekman, P., Friesen, W. V. Facial action coding system: A technique for the measurement of facial movement. CA: Consulting Psychologists Press., Ellsworth, P. C., & Smith, C. A. (1978). (1988). From appraisal to emotion: Differences among unpleasant feelings. Motivation and Emotion, 12, 271–302. https://doi.org/10.1007/s10751-008-9818-2

Fischer, A., & Hess, U. (2017). Mimicking emotions. *Current Opinion in Psychology*, *17*(August), 151–155. https://doi.org/10.1016/j.copsyc.2017.07.008.

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ)—Die kurzversion AQ-k. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *36*(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280.

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

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.

Gallotti, M., Fairhurst, M. T., & Frith, C. D. (2017). Alignment in social interactions. *Consciousness and Cognition*, *48*, 253–261. https://doi.org/10.1016/j.concog.2016.12.002.

Gong, M., & Smart, L. J. (2021). The anger superiority effect revisited: A visual crowding task. *Cognition and Emotion*, *35*(2), 214–224. https://doi.org/10.1080/02699931.2020.1818552.

Hadley, L. V., Naylor, G., & Hamilton, A. F. D. C. (2022). A review of theories and methods in the science of face-to-face social interaction. *Nature Reviews Psychology,* *1*(1), 42-54. https://doi.org/10.1038/s44159-021-00008-w.

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

Heerey, E. A., & Crossley, H. M. (2013). Predictive and reactive mechanisms in smile reciprocity. *Psychological Science*, *24*(8), 1446–1455. https://doi.org/10.1177/0956797612472203.

Heerey, E. A., & Kring, A. M. (2007). Interpersonal consequences of social anxiety. *Journal of Abnormal Psychology*, *116*(1), 125–134. https://doi.org/10.1037/0021-843X.116.1.125.

Hess, U., & Bourgeois, P. (2010). You smile-I smile: Emotion expression in social interaction. *Biological Psychology*, *84*(3), 514–520. https://doi.org/10.1016/j.biopsycho.2009.11.001.

Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior*, *24*(4), 265–283. https://doi.org/10.1023/A:1006623213355.

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

Kroczek, L. O. H., & Mühlberger, A. (2022). Returning a smile: Initiating a social interaction with a facial emotional expression influences the evaluation of the expression received in return. *Biological Psychology*, *175*, 108453. https://doi.org/10.1016/j.biopsycho.2022.108453.

Kroczek, L. O., Pfaller, M., Lange, B., Müller, M., & Mühlberger, A. (2020). Interpersonal distance during real-time social interaction: Insights from subjective experience, behavior, and physiology. *Frontiers in Psychiatry*, *11*, 561. https://doi.org/10.3389/fpsyt.2020.00561

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.

Lahnakoski, J. M., Forbes, P. A. G., McCall, C., & Schilbach, L. (2020). Unobtrusive tracking of interpersonal orienting and distance predicts the subjective quality of social interactions: Predicting quality of social interaction. *Royal Society Open Science*, *7*(8), https://doi.org/10.1098/rsos.191815rsos191815.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*(3), 145–162.

Luke, S. G. (2016). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 1–9. https://doi.org/10.3758/s13428-016-0809-y.

Lundqvist, D., Esteves, F., & Öhman, A. (1999). The face of wrath: Critical features for conveying facial threat. *Cognition and Emotion*, *13*(6), 691–711. https://doi.org/10.1080/026999399379041.

McTeague, L. M., Laplante, M. C., Bulls, H. W., Shumen, J. R., Lang, P. J., & Keil, A. (2018). Face perception in social anxiety: Visuocortical dynamics reveal propensities for hypervigilance or avoidance. *Biological Psychiatry*, *83*(7), 618–628. https://doi.org/10.1016/j.biopsych.2017.10.004.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006.

Mühlberger, A., Wieser, M. J., Herrmann, M. J., Weyers, P., Tröger, C., & Pauli, P. (2009). Early cortical processing of natural and artificial emotional faces differs between lower and higher socially anxious persons. *Journal of Neural Transmission*, *116*(6), 735–746. https://doi.org/10.1007/s00702-008-0108-6.

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2009). Slow echo: Facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Developmental Science*, *12*(4), 510–520. https://doi.org/10.1111/j.1467-7687.2008.00796.x.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., & Vogeley, K. (2012). Eyes on the mind: investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Frontiers in Psychology, 3*, 537. https://doi.org/10.3389/fpsyg.2012.00537.

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.

Riehle, M., Kempkensteffen, J., & Lincoln, T. M. (2017). Quantifying facial expression synchrony in face-to-face dyadic interactions: Temporal dynamics of simultaneously recorded facial EMG signals. *Journal of Nonverbal Behavior*, *41*(2), 85–102. https://doi.org/10.1007/s10919-016-0246-8.

Silver, C. A., Tatler, B. W., Chakravarthi, R., & Timmermans, B. (2021). Social agency as a continuum. *Psychonomic Bulletin & Review*, *28*(2), 434–453. https://doi.org/10.3758/s13423-020-01845-1.

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.

Staugaard, S. R. (2010). Threatening faces and social anxiety: A literature review. *Clinical Psychology Review*, *30*(6), 669–690. https://doi.org/10.1016/j.cpr.2010.05.001.

Tarr, B., Slater, M., & Cohen, E. (2018). Synchrony and social connection in immersive virtual reality. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-21765-4.

Weyers, P., Mühlberger, A., Hefele, C., & Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology*, *43*(5), 450–453. https://doi.org/10.1111/j.1469-8986.2006.00451.x.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes-using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(1), 98–107. https://doi.org/10.1093/scan/nsq024.