Why we interact: On the Functional Role of the Striatum in the Subjective Experience of Social Interaction

Short title

Why we interact

Authors

Ulrich J. Pfeiffer[1,2]*, Leonhard Schilbach[1], Bert Timmermans[3], Bojana Kuzmanovic[1,4],

Alexandra L. Georgescu[1], Gary Bente[5], Kai Vogeley[1,2]

Affiliations

[1]Neuroimaging Lab, Department of Psychiatry, University Hospital Cologne, Kerpener Strasse 62, 50937 Cologne, Germany

[2]Institute of Neuroscience and Medicine – Cognitive Neuroscience (INM-3), Research Center Juelich, Leo-Brandt Strasse, 52428 Juelich, Germany

[3]School of Psychology, University of Aberdeen, Aberdeen AB24 3FX, Scotland, UK

[4]Institute of Neuroscience and Medicine – Ethics in the Neurosciences (INM-8), Research Center Juelich, Leo-Brandt-Strasse, 52428 Juelich, Germany

[5]Department of Psychology, Faculty of Human Sciences, University of Cologne, Richard-Strauss-Strasse 2, 50931 Cologne, Germany

*Corresponding author (u.pfeiffer@fz-juelich.de)

Abstract

There is ample evidence that human primates strive for social contact and experience interactions with conspecifics as intrinsically rewarding. Focusing on gaze behavior as a crucial means of human interaction, this study employed a unique combination of neuroimaging, eye-tracking, and computer-animated virtual agents to assess the neural mechanisms underlying this component of behavior. In the interaction task, participants believed that during each interaction the agent's gaze behavior could either be controlled by another participant or by a computer program. Their task was to indicate whether they experienced a given interaction as an interaction with another human participant or the computer program based on the agent's reaction. Unbeknownst to them, the agent was always controlled by a computer to enable a systematic manipulation of gaze reactions by varying the degree to which the agent engaged in joint attention. This allowed creating a tool to distinguish neural activity underlying the subjective experience of being in engaged in social and non-social interaction. In contrast to previous research, this allows measuring neural activity while participants experience active engagement in real-time social interactions. Results demonstrate that gaze-based interactions with a perceived human partner are associated with activity in the ventral striatum, a core component of reward-related neurocircuitry. In contrast, interactions with a computer-driven agent activate attention networks. Comparisons of neural activity during interaction with behaviorally naïve and explicitly cooperative partners demonstrate different temporal dynamics of the reward system and indicate that the mere experience of engagement in social interaction is sufficient to recruit this system.

1. Introduction

In the hierarchy of human needs, the need to affiliate with others has been located directly after physiological and prior to egoistic needs related to self-actualization and esteem (Maslow, 1943). Accordingly, an intrinsic motivation for social interaction unique to the human species has been proposed (Baumeister and Leary, 1995; Tomasello, 2009). Over the last decade, multiple neuroeconomic studies have indeed found reward-related brain activity during social interactions (Rilling and Sanfey, 2011). Two key regions of the reward system are the ventral striatum (VS) and the medial orbitofrontal cortex (mOFC) which have been implicated in the anticipation and consumption of rewards (Berridge et al., 2009). While the VS has been specifically linked to the anticipation of rewards and the computation of reward prediction errors (Báez-Mendoza and Schultz, 2013; Daniel and Pollmann, 2014), the mOFC appears to be involved in the subjective experience of reward (Peters and Büchel, 2010) as well as value-guided decision making (Noonan et al., 2012). While many studies indicate a link between social interaction and the reward system (Krach et al., 2010; Rilling and Sanfey, 2011), , the application of economic games to study social interaction typically involves high-level concepts such as trust, fairness, cooperation, or competition (Fehr and Camerer, 2007). As a consequence, the claim that experiencing engagement in interaction with others *per se* is rewarding has never been put to the test.

An understanding of the neural mechanisms underlying human sociality has recently been argued to require measurements of brain activity during active participation in naturalistic social interactions rather than detached observation of social stimuli (Hari and Kujala, 2009; Schilbach et al., 2013). Accordingly, there is growing consensus that "it is in engagement with

other people rather than in thought that people normally and fundamentally know other people" (Reddy and Morris, 2004, p. 657). The relative paucity of studies of naturalistic social interactions can be explained by the difficulty of designing experimental paradigms which allow experimental control while participants subjectively experience engagement in social interaction. Up to date, the most natural social interactions have been studied using EEG hyperscanning while participants perform spontaneous motor coordination tasks (e.g. Tognoli et al., 2007; Dumas et al., 2010), engage in joint attention (Lachat et al., 2012) or play games together (Babiloni et al., 2007; Astolfi et al., 2010). However, the types of interactions are too complex for application in fMRI studies – either due to the involvement of excessive movements or due to the inherent complexity in the case of spontaneous motor coordination tasks (Pfeiffer et al., 2013; Schilbach et al., 2013).

The aim of the present study was to investigate the function of the reward-system during naturalistic interactions. To this end, we addressed the neural mechanisms supporting the subjective experience of being engaged in social interaction by examining neural activity while participants actively participated in gaze-based interactions. Gaze was selected because it constitutes a crucial domain of everyday social encounters and has the advantage that it can be implemented inside an MRI scanner due to the minimal involvement of body movements (Pfeiffer et al., 2013). Gaze behavior was visualized via computer-animated agents in real-time (e.g. Fox et al., 2009). The combination of neuroimaging, eye-tracking and virtual reality techniques allowed implementing realistic but basic social interactions while maintaining experimental control (Bohil et al., 2011; Pfeiffer et al., 2013).

The interaction task applied in the present study was designed to create situations in which the gaze-based interaction with a virtual agent either induced the subjective experience of

4

being engaged in human social interaction or the subjective experience of being in a non-social interaction – i.e. with a computer program. To this end, each block of the interaction task comprised five trials in which the agent would either engage in joint or non-joint attention with the participant (Figure 1A/B). Joint attention was chosen as a building block of the interaction task because it is a core component of naturalistic social interactions (Mundy and Newell, 2007). Participants believed that during each block the agent was either controlled by a computer algorithm or a human interaction partner. In fact, the interaction partner was a confederate and the agent's gaze behavior was always controlled by the algorithm to permit systematic manipulation. This was accomplished by varying the proportion of joint attention trials from zero to five out of five, thereby modifying behavioral contingency over a block. Participants' task was to decide on the nature of their interaction partner based on the agent's reactions during each block. Thereby, the decision between human and computer emerged during the course of the interaction, while other studies explicated this distinction a priori as an independent variable (McCabe et al., 2001; Gallagher et al., 2002; Sanfey et al., 2003). This allowed assessing the neural mechanisms underlying the subjective experience of being engaged in human social interaction (Pfeiffer et al., 2011).

Unconstrained as well as cooperative interaction contexts were established in two phases in which the interaction partner was either introduced as naïve to participants' task, or as an explicit cooperator (e.g. Taborsky, 2007) helping them to identify human interactions. Based on the claim that social interaction is *per se* rewarding, we hypothesized that the reward component inherent to cooperative contexts would already be present in unconstrained interactions. Furthermore, we predicted that the striatum would encode reward components related to a

motivation to interact, whereas the orbitofrontal cortex was expected to encode the rewarding experience.

2. Materials and Methods

2.1 Participants

32 right-handed volunteers participated in the study, which was approved by the ethics committee of the Medical Faculty of the University of Cologne. 12 participants were excluded due to excessive movements (n = 4), technical problems with the eye-tracker (n = 5), and disbelief in the cover story (n = 3). 20 participants (9 female/11 male, $M = 27.75$ years, $SD \pm 6.44$) were included in the analyses.

2.2 Visual stimulation and eye-tracking

The paradigm is an fMRI adaptation of the 'non-verbal Turing test' which has recently been validated behaviorally (Pfeiffer et al., 2011) and used the same male virtual agent displaying a neutral facial expression as has been used in previous studies to ensure comparability (Pfeiffer et al., 2011, 2012; Schilbach et al., 2010). Stimuli were presented using a thin-film-transistor liquid crystal display (TFT-LCD) screen attached at a distance of 100 cm from the end of the scanner (viewing angle: 14° x 18° horizontal x vertical). They were displayed to participants via a mirror on the head coil. Participants' gaze was monitored via the same mirror using an EyeLink 1000 eye-tracking system (SR Research, Missisauga, Canada). Gaze data were collected at a sampling rate of 500 Hz on the EyeLink host computer and made available to Presentation™

(Neurobehavioral Systems, Albany, CA) for interactive stimulus presentation (Wilms et al., 2010).

2.3 Interaction task

Participants interacted with a virtual agent in a series of interaction blocks (Figure 1A). They believed that in each block the gaze reactions of the agent were either controlled by another participant or a computer algorithm. In fact, the other participant was a confederate and the agent's reactions were always computer-controlled to allow for systematic variation. Each block (Figure 1A) comprised five trials (Figure 1B) in which the agent engaged in joint (JA) or non-joint attention (NJA). Systematic variation resulted in six experimental conditions (0/5, 1/5, 2/5, 3/5, 4/5, and 5/5 times of JA) and a control condition in which the agent closed its eyes on each trial. Each of these seven conditions was repeated six times in each interaction context, thus yielding 42 blocks per context. In the control condition, participants disengaged from the actual task and only were required to watch the face and press one of the two buttons after the end of the block. The rationale was to improve the model in the fMRI analyses by including a regressor relating for task-unspecific perceptual and motor aspects. Similar control conditions were used in previous studies of live interactions (e.g. Redcay et al., 2010, 2012; Schilbach et al., 2010). The trial-by-trial variation of gaze behavior thereby induced a variation of behavioral contingency, which increased the more often the agent displayed the same reaction during one block. An agent engaging in JA in each trial (positive contingency) therefore behaved as contingently as an agent engaging in NJA in each trial (negative contingency).
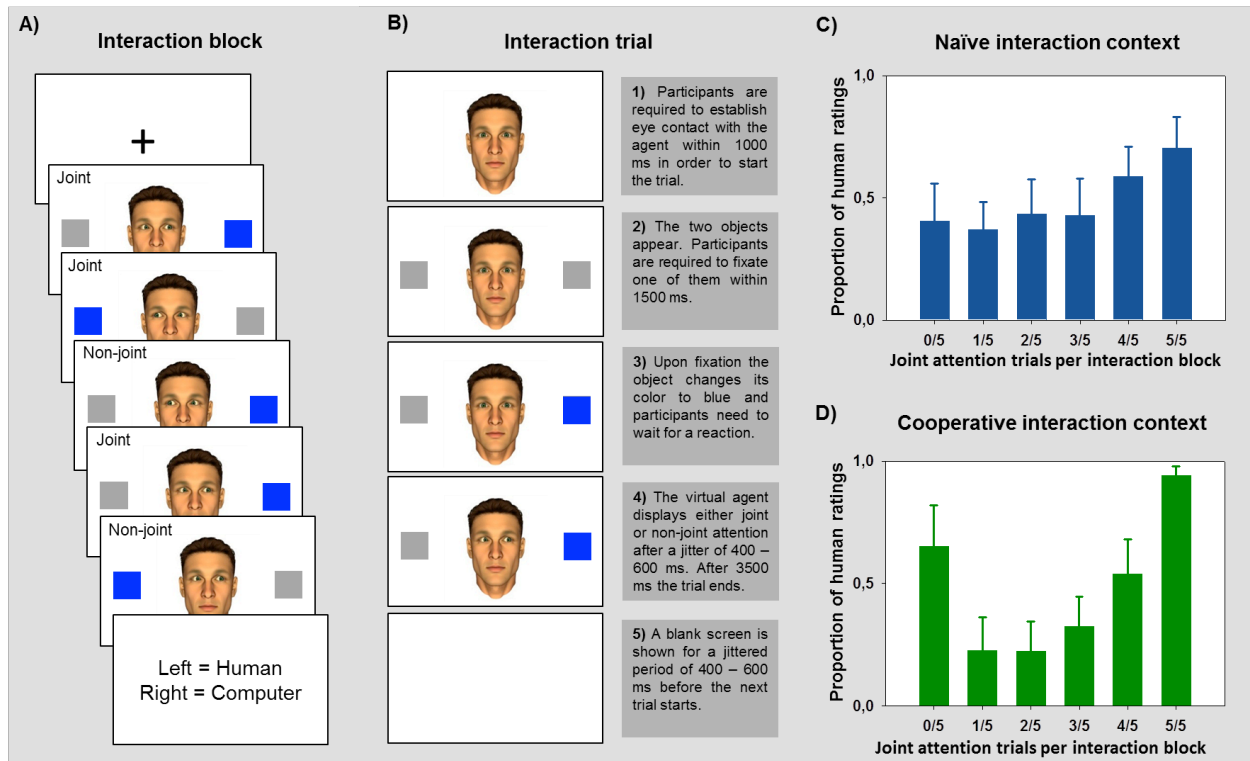
Figure 1: Task structure and behavioral results

(A) Each interaction block comprises five gaze trials. At the end of each block participants indicate whether they experienced this interaction as social ('human') or non-social ('computer'). This block exemplifies a 3/5 condition in which the agent engages in joint attention three out of five possible times. (B) In each of five trials of an interaction block, participants initiate an exchange of gaze shifts. (C) In the naïve context, the mean proportion of 'human' ratings correlates with increased congruency of gaze reactions. (D) In the cooperative context, the mean proportion of 'human' ratings correlates with the mere contingency of the agent's gaze reactions. Error bars indicate 95% confidence intervals.

At the beginning of a trial, participants established eye contact with the agent within 1000 ms. Upon eye contact, two objects appeared, one of which participants had to fixate within 1500 ms. If participants fixated an object earlier than 1500 ms, the remaining time was added to the jittered break at the end of the trial in order to keep average trial length at 4000 ms. If there was no fixation within these time limits, trials were aborted and the block registered as invalid.

Selected objects were marked in blue to provide participants with feedback about successful gaze registration. They were informed that their initial gaze shift to the object was transmitted to an agent on the other participant's screen in real-time, and that they would likewise see the other's gaze reaction as visualized on their screen. With a jittered latency of 400 – 600 ms (Pfeiffer et al., 2012), the agent followed participants' gaze or looked to the other object, thereby establishing JA or NJA. This reaction was displayed until trial duration of 3500 ms was reached. Before the next trial started, a blank screen was presented for 500 - 1000 ms plus the remaining time of the object fixation phase.

After each block, participants indicated whether they believed they had been interacting with the other participant or a computer algorithm by button press within a response window of 1500 ms. The response window was followed by a jittered break of 5000 to 7000 ms before the next block began.

2.4 Interaction contexts

The experiment consisted of two phases to distinguish unconstrained from cooperative interactions. In the first phase, the confederate was introduced as *naïve* to participants' task. Participants were told that their interaction partner had been instructed to react to each of their gaze shifts by looking at one of the objects without any additional information. This was supposed to provide an unconstrained interaction context in which participants had no a priori assumptions about the other's behavior. In the second phase, the other was introduced as *cooperative* in order to assess neural activity while participants engage in the same type of interaction in a cooperative context. They were told that the other's task was now to react to

them in such a way that would facilitate the distinction between human- and computer-controlled interactions. It was not specified in which exact way the interaction partner would do this.

The order of the two phases was not randomized because the naïve condition required participants to assume that the other did not have any knowledge about their task and reacted in an unconstrained fashion. Specifically, we wanted to prevent activation of the high-level concept of cooperation *before* participants engage in naïve interactions because this could have primed expectations leveling out effects of mere interactivity in the second phase. To prevent habituation and novelty effects, participants engaged in a practice session of 5 minutes before the first run. During this session, the experimenter monitored their behavior to provide additional instruction if necessary. As noted above, the design is a within-subject adaptation of a behavioral study (Pfeiffer et al., 2011). Behavioral results replicate those of the between-subject version (see Figure 2 of Pfeiffer et al., 2011). Moreover, participants' responses in the debriefing questionnaires did not indicate differences between the within- and between-subject version (compare Figure 2 of the present study and Figure S1 of Pfeiffer et al., 2011).
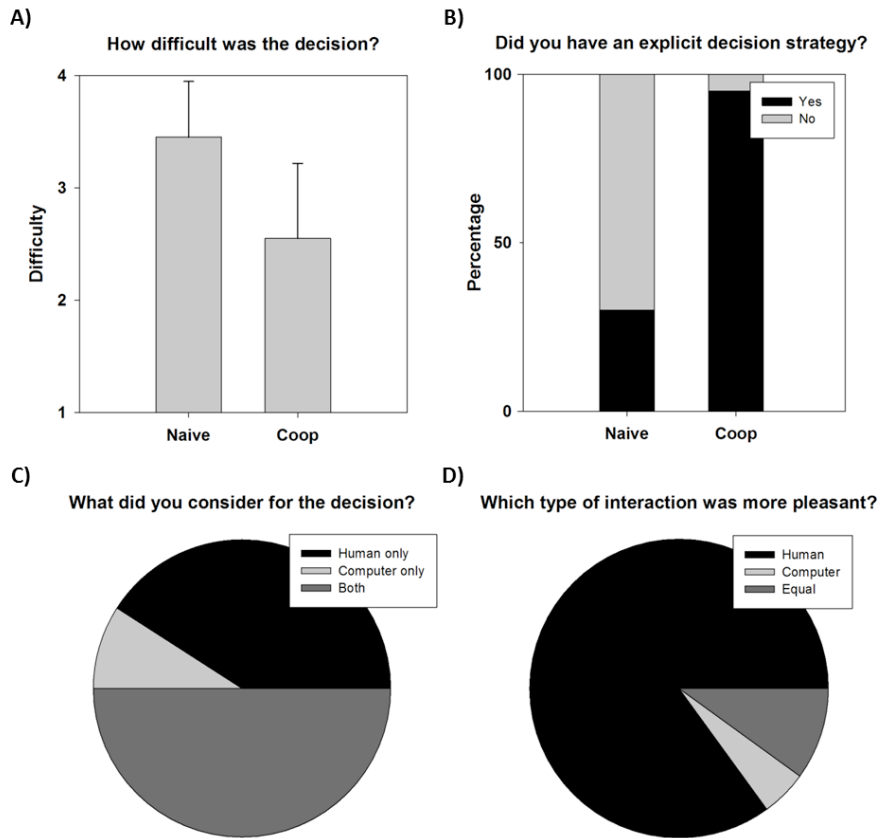
Figure 2: Participants' responses to post-experiment questionnaires

(A) Results indicate that the decision was more difficult in the naïve than in the cooperative context. (B) In the naïve context, the decision was not based on strategic thinking and explicit decision criteria, but on intuitive processes. (C) Participants preferred grounding their decision on considerations about human behavior rather than about the functionality of a computer program. (D) Interactions experienced as social interactions with another human participant were experienced as more enjoyable than non-social interactions.

2.5 Procedure

Participants received detailed written instructions covering the interaction task. Before they were led to the scanner, they were briefly introduced to the confederate, who was seemingly being instructed at the same time. Upon calibration of the eye-tracker, scanning commenced with a

practice session of eight blocks. The ensuing experiment had two phases, each consisting of two 10-minute runs that were preceded by a re-calibration of the eye-tracker. The experiment started with the naïve phase. After the second run, there was a break of 3 minutes. During this break, participants were instructed that their interaction partner received additional instructions and that a cooperative phase was now beginning. After the first and third run there was a short break. In each run, all conditions were repeated thrice in a randomized fashion. After the experiment, participants completed a questionnaire (Figure 2). Among other questions, this questionnaire contained the question, how pleasant participants experienced interactions with humans and computers, respectively. Following this question, they had the chance to note down any comments they had with respect to the study. People who uttered disbelief in the cover story were later excluded from the analysis. Upon completion of data acquisition, participants received an email debriefing them in detail regarding the cover story and the experiment. They were then asked explicitly whether they had believed to be interacting with another human participant in some of the interaction blocks. This served as the final manipulation check and determined whether participants entered data analysis or not.

2.6 Behavioral data analysis

Prior to analysis, an arcsine transformation was performed to correct for violations of normality (McDonald, 2009). The effects of the factors *Context* (naïve vs. cooperative) and *Gaze contingency* (JA trials/block) on participants' ratings were analyzed using repeated-measures ANOVAs. Planned polynomial contrasts were applied for trend analyses. All results were Greenhouse-Geisser corrected. Effect size was calculated as ω^2 (Cohen, 1988): small effects: $\omega^2 < .006$; moderate effects: $\omega^2 < 0.15$; large effects: $\omega^2 > 0.15$.

To examine how participants' decision-making process unfolded over time, we assessed how the blocks' rating was influenced by the occurrence of JA or NJA on each of its five trials. The rationale behind these analyses is that we sought to obtain an implicit measure of how trial information was integrated in the two different interaction contexts. This was necessary because post-experiment questionnaires can only provide explicit, introspective and therefore subjectively biased information about decision-making. Initially, we performed three logistic regressions with *Rating* (computer = 0, human = 1) as dependent variable using the 'Forward' method, in which predictors are added consecutively – starting with the strongest, and adding predictors with decreasing strength until adding a new predictor fails to explain more variance. The first analysis was performed over both contexts (naïve and cooperative) and included ten predictors: five *Trial* predictors for the agent's reactions on each of the five trials (NJA = 0, JA = 1), and five *Trial* x *Context* (naive = 0, cooperative = 1) predictors. Subsequently, we performed two separate logistic regressions for each context (naive vs. cooperative), in which only the agent's reactions on the five trials were entered as predictors.

2.7 Functional data acquisition

Scanning was performed on a Siemens Trio 3-T scanner (Siemens Medical Solutions, Erlangen, Germany). A T2*-weighted gradient echo planar imaging (EPI) sequence with the following parameters was used: TR = 2200 ms, TE = 30 ms, 36 axial slices, slice thickness = 3.0 mm, in-plane resolution = 3.0 x 3.0 mm, field of view = 200.0 x 200.0 mm. In each run, 280 images were acquired. The first five images of each run were discarded to eliminate saturation effects.

2.8 Functional data preprocessing

Images were processed using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). Motion correction was completed by an affine registration procedure (Ashburner and Friston, 1999). Images were then realigned to the first image of the time series and subsequently to the mean of these images. Next, the mean EPI was computed for each participant and spatially normalized to the MNI single subject template (Collins et al., 1994) using the unified segmentation function of SPM8 with a 2x2x2 mm isotropic resolution. The ensuing deformation was applied to individual EPI volumes. Spatial smoothing of the normalized images was performed using an 8-mm FWHM Gaussian kernel.

2.9 Functional data analysis

Data were analyzed using a General Linear Model as implemented in SPM8. The following general specifications apply to all conducted analyses, each of which will be described in detail thereafter. Low-frequency signal drifts were removed using a high-pass filter with a cutoff of 128 seconds (Macey et al., 2004). At subject level, experimental conditions were modeled by a boxcar reference vector convolved with the canonical hemodynamic response function. Invalid blocks were modeled on a distinct regressor. Each experimental condition was contrasted against the implicit baseline by weighting the regressor of interest with 1 and the remaining regressors with 0, and the resulting contrasts were fed into a flexible factorial design (as provided by SPM8) with factors *Subject* and *Condition* using a random-effects model for group level comparisons (Worsley et al., 1996). Here, all effects were thresholded at $p < .05$ at cluster-level, family-wise-error-corrected for multiple comparisons ($p_{FWE-corr} < .05$), with an underlying voxel-level threshold of $p < .001$, uncorrected. Version 1.8 of the SPM anatomy toolbox (Eickhoff et

al., 2005) and the brain atlas of Duvernoy (Duvernoy, 1999) were used for anatomical localization. Activation maps were superimposed on an SPM canonical T1-weighted image. Error bars of the parameter estimates indicate 90% confidence intervals.

The present study design allows multiple different analyses, which made it necessary to select a strategy for data analysis which allows addressing the main questions of our study. The first major question concerned the neural mechanisms of the experience of being actively engaged in social interaction (i.e. 'human'- versus 'computer'-rated blocks) and a possible modulation of these mechanisms by interaction context. The second key question addressed the temporal integration of behavioral cues (i.e. gaze reactions) depending on interaction context. Study design therefore justified three major lines of analysis:

(1) Analysis based on participants' ratings (analyses 1 – 3). The first three analyses were guided by participants' responses to examine activity differences between interactions rated as 'human' (hum) and 'computer' (com). At the subject level of the first analysis, blocks rated as 'human' and 'computer' were modeled on distinct regressors, separately for the two contexts. Stimulus events were defined from block onset to block end (defined by the response window, total duration of 20000 ms). At group level, we tested for effects of the rating both collapsed across the two contexts ($hum_{all\ block} > com_{all\ block}$ and $com_{all\ block} > hum_{all\ block}$), as well separately for the naïve and the cooperative context ($hum_{naïve\ block} > com_{naïve\ block}$, $com_{naïve\ block} > hum_{coop\ block}$, $hum_{coop\ block} > com_{coop\ block}$, $com_{coop\ block} > hum_{coop\ block}$). Two further analyses were informed by the output of the regressions performed on the behavioral data. In the second analysis, early components of decision-making in blocks rated as 'human' and as 'computer' were analyzed by confining stimulus events to the first two trials (8000 ms) of a block. Group level contrasts were: $hum_{naïve\_early} > com_{naïve\_early}$, $com_{naïve\_early} > hum_{naïve\_early}$, $hum_{coop\_early} > com_{coop\_early}$, and

com$_{\text{coop\_first}}$>hum$_{\text{coop\_first}}$. In the third analysis, the effect of trial progression was modeled by using linear parametric modulation of the BOLD response by trial position, separately for blocks rated as 'human' and 'computer' and for naïve and cooperative contexts. By doing so, we isolated brain regions in which activity increased with increasing trial position. This resulted in the following group level comparisons: hum$_{\text{naïve\_param}}$>com$_{\text{naïve\_param}}$, com$_{\text{naïve\_param}}$>hum$_{\text{naïve\_param}}$, hum$_{\text{coop\_param}}$> com$_{\text{coop\_param}}$, and com$_{\text{coop\_param}}$>hum$_{\text{coop\_param}}$.

(2) Analysis based on gaze contingency (analysis 4). Driven by the manipulation of gaze contingency across proceeding trials within each block, the fourth analysis tested for linear increases of neural activity with increasing positive and negative contingency. At subject level, blocks with different conditions of contingency (0/5, 1/5, 2/5, 3/5, 4/5, and 5/5 times of JA) were modeled as distinct regressors, separately for the naïve and the cooperative context. At group level, we applied differentially weighted contrasts in order to test for effects of increasing positive contingency (naïve$_{\text{increase\_JA}}$ and coop$_{\text{increase\_JA}}$: -3 -2 -1 1 2 3), and for effects of increasing negative contingency (naïve$_{\text{increase\_NJA}}$ and coop$_{\text{increase\_NJA}}$: 3 2 1 -1 -2 -3).

(3) Event-related analysis of joint attention (analysis 5). In the final analysis we compared how JA and NJA trials were processed in the naïve and cooperative contexts irrespective of the experimental condition in which they occurred. Notably, JA and NJA constitute complex events comprising multiple steps such as the establishment of mutual gaze, a gaze shift to one of the objects and a gaze reaction by the agent. Nonetheless, they can be used in such a fashion because they only differ in one crucial aspect – i.e. the congruency of the agent's gaze reaction which either follows participants' gaze or averts its gaze to the other object, thereby engaging in JA or NJA, respectively (e.g. Materna et al., 2008; Schilbach et al., 2010; Redcay et al., 2012). For this analysis, stimulus events were defined at subject level from trial

onset to the trial end defined by the appearance of the blank screen, thereby amounting to a length of 3500 ms. JA and NJA trials were modeled on distinct regressors, separately for the naïve and the cooperative contexts. At group level, the following contrasts were computed: $JA_{naïve} > NJA_{naïve}$, $NJA_{naïve} > JA_{naïve}$, $JA_{coop} > NJA_{coop}$, $NJA_{coop} > JA_{coop}$.

3. Results

3.1 Behavioral

3.1.1 Effect of gaze contingency and context on participants' decisions

Blocks containing invalid trials (naïve context: 6.67 %, cooperative context: 6.11 %) were excluded from the analysis. In the remaining blocks, results demonstrated a main effect of gaze contingency on participants' ratings, $F(2.45, 46.55) = 13.19$, $p < .001$, $\omega^2 = .23$, and a significant interaction between instruction and contingency, $F(3.13, 59.35) = 11.19$, $p < .001$, $\omega^2 = .08$. This interaction was scrutinized by separate analyses of ratings in the naïve and the cooperative context. In the naïve context (Figure 1C), there was a significant main effect of gaze contingency on participants' ratings, $F(2.76, 52.38) = 3.55$, $p = .023$, $\omega^2 = .03$. Planned polynomial contrasts revealed that this effect was characterized by a significant linear trend, $F(1, 19) = 7.84$, $p = .011$, $\omega^2 = .29$, thereby indicating that the proportion of blocks the proportion of 'human'-rated blocks were positively correlated with the number of joint attention trials per block. This means that 'human' ratings increased with increasing numbers of joint attention trials and thus argues for a particular importance of congruent reactions during unconstrained interactions. In the cooperative context

(Figure 1D), contingency also had a significant effect on participants' ratings, $F(2.79, 52.95) =$ 21.79, $p < .001$, $\omega^2 = .27$. This was characterized by a significant linear trend, $F(1, 19) = 20.21$, p $< .001$, $\omega^2 = .19$, and a quadratic trend, $F(1, 19) = 36.63$, $p < .001$, $\omega^2 = .39$, which indicates that 'human' ratings were more closely related to the general contingency of reactions. This suggests that any consistent behavior over an entire block was taken as indicative of a human counterpart and replicates the findings of a behavioral between-subject version of the present task (Pfeiffer et al., 2011).

3.1.2 Temporal integration of information

The results of the logistic regression over both interaction contexts are listed in the top half of Table 1. The final three trials showed up as main effects across both contexts, with the fourth trial weighing in the heaviest: if on this trial the agent engaged in joint attention, the chances of rating 'human' are over 1.59 times higher than when the agent looks the other way. Most importantly, however, the second trial loaded only in interaction with condition. The second and third logistic regression, for the naïve and cooperative contexts respectively, confirmed the persistence of the main effects for the final three trials in both contexts (Table 1, bottom half). In the naïve context, already the second trial had a significant influence on the final rating, with 'human' becoming 1.38 times more likely than 'computer' if the agent engaged in joint attention on that trial. Such an early component was absent in the cooperative context where the second trial did not load at all. In sum, there is an early influence of trial type in the naïve context, which is absent in the cooperative context, where the integration of information related to decision-making is approximately linear.

18

Table 1. Regression coefficients for the logistic regression models.

	Coeff	SE	Wald	Odds ratio	95 % CI
NAIVE + COOP: Main effect of trials and interaction of trials*condition					
2nd trial (NJA vs JA) * Condition (NAIVE vs COOP)	0.284	0.087	10.70**	1.33	[1.12 - 1.58]
3rd trial NJA vs JA	0.293	0.079	13.63***	1.34	[1.15 - 1.57]
4th trial NJA vs JA	0.464	0.08	34.08***	1.59	[1.36 - 1.86]
5th trial NJA vs JA	0.309	0.08	15.02***	1.36	[1.17 - 1.59]
NAIVE: Main effect of trials					
2nd trial NJA vs JA	0.32	0.113	8.07**	1.38	[1.10 - 1.72]
3rd trial NJA vs JA	0.255	0.114	4.98*	1.29	[1.03 - 1.62]
4th trial NJA vs JA	0.432	0.114	14.32***	1.54	[1.23 - 1.93]
5th trial NJA vs JA	0.264	0.115	5.31*	1.3	[1.04 - 1.63]
COOP: Main effect of trials					
3rd trial NJA vs JA	0.328	0.111	8.76**	1.39	[1.12 - 1.72]
4th trial NJA vs JA	0.489	0.112	19.09***	1.63	[1.31 - 2.03]
5th trial NJA vs JA	0.346	0.112	9.52**	1.41	[1.14 - 1.76]

Notes: *** $p<.001$; ** $p<.005$; * $p<.05$.

3.2 Imaging

3.2.1 Neural correlates of active engagement in social interaction

Initial analysis of functional imaging data was driven by participants' ratings. Irrespective of

interaction context, blocks rated as 'human' ($hum_{all\_block}>com_{all\_block}$) were accompanied by

enhanced activation of the ventral striatum (VS) and the medial orbitofrontal cortex (mOFC;

Figure 3A, Table 2A). In contrast, during blocks rated as 'computer' ($com_{all\_block}>hum_{all\_block}$)

there was increased activation of a fronto-parietal attention network (FPAN, Corbetta and

Shulman, 2002) including the inferior parietal cortex, precuneus, and the lateral prefrontal cortex

(Figure 3B, Table 2B). When considering only the naïve context, 'human'- versus 'computer'-rated blocks ($hum_{naïve\_block} > com_{naïve\_block}$) engaged the mesolimbic reward system (Berridge et al., 2009) including the ventral tegmental area (VTA) and the VS (Figure 4A, Table 2C), while there were no significant results for the reverse contrast ($com_{naïve\_block} > hum_{naïve\_block}$). Conversely, in the cooperative context, there was increased activation of the FPAN during 'computer'-rated blocks ($com_{coop\_block} > hum_{coop\_block}$, Table 2D), while no regions were more active during 'human'-rated blocks ($hum_{coop\_block} > com_{coop\_block}$).
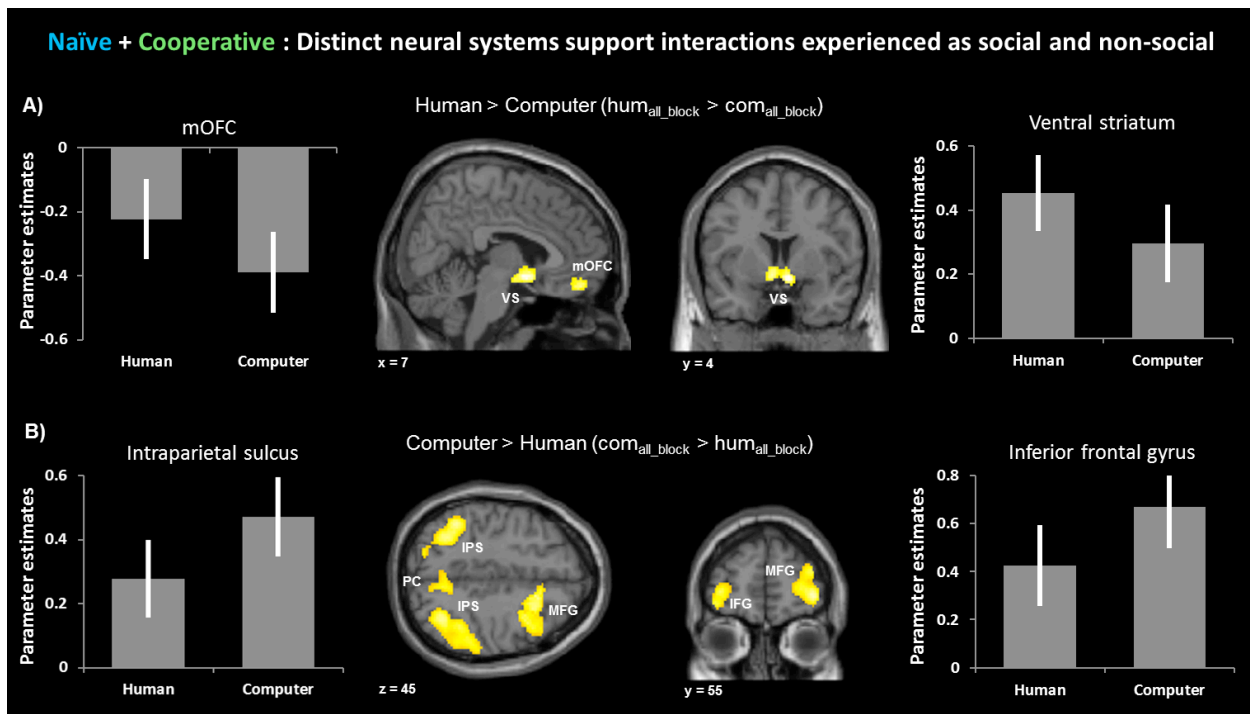

Figure 3: Neural activity during blocks rated as 'human' and 'computer'

(A) The experience of interaction with another human participant recruits the ventral striatum (VS) and medial orbitofrontal cortex (mOFC). (B) The experience of an interaction as computer-driven is associated with activity in a fronto-parietal network including the inferior parietal sulcus (IPS), precuneus (PC), premotor cortex (PMC), and lateral prefrontal cortex (for all fMRI graphs: statistical threshold is $p < .05$ cluster-level-corrected for multiple comparisons; error bars depict 90% confidence intervals).

Table 2. Analyses based on participants' subjective ratings of the nature of their interaction partner ('human' versus 'computer').

Region	Cluster		Side	MNI Coordinates			T
	Size	$p_{FWE\text{-}corr}$		x	y	z	
Whole block							
A) hum_{all_block} > com_{all_block}							
Nucleus accumbens (Ventral striatum)	468	.000	R	8	4	-10	4.96
Nucleus accumbens (Ventral striatum)			L	-6	10	-8	4.57
mOFC	326	.002	R	4	48	-16	4.61
mOFC			L	-8	42	-12	3.98
B) com_{all_block} > hum_{all_block}							
Supramarginal gyrus	2919	.000	R	52	-42	36	4.88
Intraparietal sulcus			R	34	-60	44	4.88
Precuneus			R	10	-60	40	4.67
Intraparietal sulcus	1513	.000	L	-30	-48	32	5.63
Inferior frontal gyrus	1455	.000	R	40	56	-2	5.04
Lateral orbital sulcus			R	46	48	-12	4.93
Middle frontal gyrus			R	52	38	22	4.39
Middle frontal gyrus	1207	.000	R	32	6	60	4.89
Superior frontal sulcus			R	22	14	42	4.82
Inferior frontal gyrus	332	.013	L	-36	56	2	4.39
C) hum_{naïve_block} > com_{naïve_block}							
Nucleus accumbens (Ventral striatum)	836	.000	R	10	4	-10	5.17
Medial orbitofrontal cortex			R	6	22	-10	4.40
Putamen			R	24	20	0	4.14
Anterior cingulate cortex	299	.020	L	-8	32	6	4.54
Anterior cingulate cortex			R	10	28	16	3.75
Nucleus accumbens	248	.039	L	-10	2	-2	4.19
Putamen			L	-8	16	2	4.14
Medial orbitofrontal cortex			L	-8	28	-12	3.98
Substantia nigra/Subthalamic nucleus	243	.043	L	-6	-6	-16	4.53
Ventral tegmental area			R	6	-24	-18	4.46
Ventral tegmental area			L	-6	-20	-18	4.07

D) com$_{coop\_block}$ > hum$_{coop\_block}$

Intraparietal sulcus	10451	.000	R	36	-44	34	6.23
Precuneus			R	8	-58	42	6.11
Supramarginal gyrus			R	50	-44	36	6.03
Intraparietal sulcus			L	-32	-54	40	5.87
Inferior frontal sulcus	6743	.000	R	28	52	4	6.56
Inferior frontal gyrus			R	40	56	-2	6.29
Middle frontal gyrus			R	32	8	60	6.21
Middle frontal gyrus	936	.000	L	-50	24	34	6.47
Lateral orbital gyrus	868	.000	L	-35	56	-8	5.17
Inferior frontal gyrus			L	-34	48	2	4.98

First two trials of block

E) hum$_{naïve\_early}$ > com$_{naïve\_early}$

Nucleus accumbens (Ventral striatum)	704	.012	L	-12	6	-6	4.35
Caudate nucleus (head)			L	-10	18	2	4.08
Putamen			L	-18	18	-8	3.92
Medial orbital gyrus			L	-22	14	16	3.93

Parametric increase over block

F) hum$_{coop\_param}$ > com$_{coop\_param}$

Nucleus accumbens (Ventral striatum)	719	.000	R	6	6	-4	4.49
Nucleus accumbens (Ventral striatum)			L	-6	10	-8	4.45
Putamen			R	20	14	-10	3.84

G) com$_{coop\_param}$ > hum$_{coop\_param}$

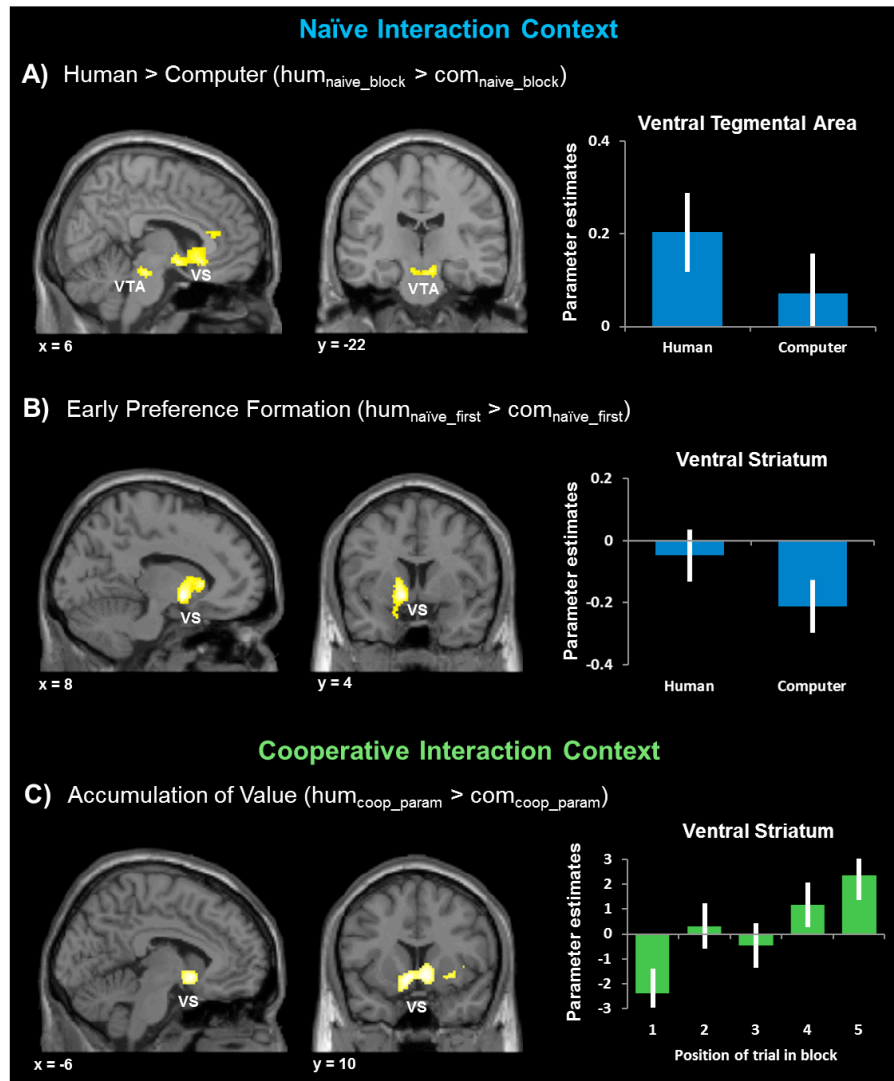|							
Angular gyrus	474	.001	R	42	-64	52	4.25
Angular gyrus			R	48	-64	34	4.07
Inferior frontal gyrus	201	.051	R	44	30	24	4.13

Figure 4: Modulation of reward processing by interaction context

(A) In the naïve context, blocks rated as 'human' recruit mesolimbic reward areas relative to 'computer'-rated blocks. This includes the ventral tegmental area (VTA) and the ventral striatum (VS). (B) The activity of the VS is predictive of participants' ratings already during the first two trials of a block. (C) In the cooperative context, activity in the ventral striatum (VS) unfolds over the time course of interaction blocks (i.e. with increasing trial progression) rated as 'human'.

3.2.2 Neural correlates of temporal integration of information

Further fMRI analyses (analyses 2 and 3) were driven by the behavioral finding of early versus linear integration of gaze reactions in the naïve and cooperative context, respectively. With respect to the early integration during the first two trials of blocks, VS activity during the first two trials of a block was predictive of participants' 'human' ratings exclusively in the naïve context ($hum_{naïve\_first} > com_{naïve\_first}$, Figure 4B, Table 2E). On the contrary, only in the cooperative context, linear parametric analyses including trial progression as a parametric regressor revealed an increase of VS activity with increasing trial progression over the full length of blocks rated as 'human' ($hum_{coop\_param} > com_{coop\_param}$, Figure 4C, Table 2F). Neural differentiation of social interaction therefore occurs early during naïve interactions while developing over time during cooperative interactions. The latter finding explains the lack of differential activity during the entire block for 'human'- versus 'computer'-rated blocks in the cooperative context, when activity during the entire block is compared ($hum_{coop\_block} > com_{coop\_block}$).

3.2.3 Neural processing of gaze in interaction

Finally, the agent's gaze behavior was used to analyze fMRI data with respect to gaze contingency (analysis 4) and at the level of single events (analysis 5). In the naïve context, single joint attention trials ($JA_{naïve} > NJA_{naïve}$) recruited regions associated with the 'social brain network' (Frith, 2007) involving the bilateral anterior cingulate cortex (ACC), and the left amygdala, medial prefrontal cortex (mPFC), temporal pole, and superior temporal sulcus (Figure 5, Table 3). However, the analysis focusing on effects of gaze contingency ($naïve_{increase\_JA}$) revealed increases of activity in the bilateral paracentral lobule, but not in social brain regions (Table 4A). In the cooperative context, there was no differential activity at single-trial level

(JA_{coop}>NJA_{coop}). Instead, results showed an increase of activity with increasing positive contingency ($coop_{increase\_JA}$) in the dorsal striatum, the thalamus, the ACC and the mPFC (Figure 6A, Table 4C). This indicates that the social brain network is recruited by behavioral consistency. Increasing negative contingency ($coop_{increase\_NJA}$) recruited the FPAN (Figure 6B, Table 4D).
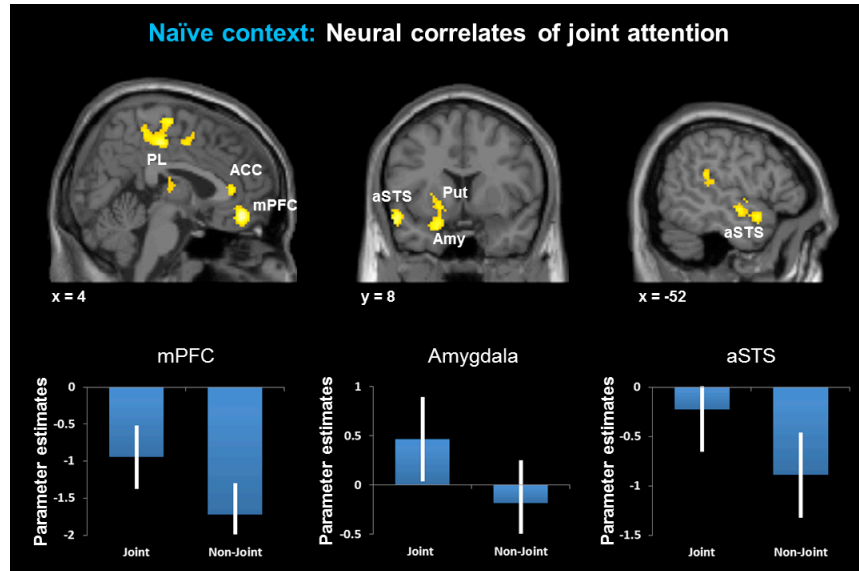

Figure 5: Event-related analysis of joint attention in the naïve context

Event-related analysis of single events of joint attention revealed activation in regions of the social brain network including the medial prefrontal cortex (mPFC), amygdala (Amy), and the anterior region of the superior temporal sulcus (aSTS). In addition, activity in the paracentral lobule (PL) was enhanced.

Table 3. Event-related analyses of gaze reactions compared JA with NJA trials.

Region	Cluster		Side	MNI Coordinates			T
	Size	*p_{FWE-corr}*		*x*	*y*	*z*	
JA_{naïve} > NJA_{naïve}							
Precentral gyrus	3207	.000	R	26	-22	60	5.86
Postcentral gyrus			R	32	-30	62	5.31
Postcentral gyrus			L	-20	-40	56	5.05
Middle cingulate cortex			R	12	-16	42	4.66
Middle cingulate cortex			L	-8	-2	34	4.60

Region			Side	x	y	z	T
Paracentral lobule			R	4	-24	48	4.57
Paracentral lobule			L	-8	-24	48	4.56
Amygdala	510	.001	L	-22	-6	-14	4.82
Medial orbitofrontal cortex			L	-22	8	-22	4.28
Putamen (Ventral striatum)			L	-20	8	-6	3.80
Thalamus	410	.003	L	-4	-18	14	4.66
Superior temporal sulcus (anterior region)	398	.004	L	-46	-6	-8	4.59
Temporal pole			L	-54	8	-14	3.93
Ventromedial prefrontal cortex	364	.006	R	4	44	-16	5.17
Anterior cingulate cortex			R	4	36	10	3.90
Anterior cingulate cortex			L	-4	38	2	3.61

Table 4. Analysis based on the contingency of the agent's gaze behavior.

The number of joint attention (JA) trials per interaction block was used as a parametric regressor. Both the contrasts referring to increasing numbers of JA as well as increasing numbers of non-joint attention (NJA) trials were reported to obtain information about the neural integration of positive and negative contingency of gaze reactions.

Region	Cluster		Side	MNI Coordinates			T
	Size	$p_{FWE-corr}$		x	y	z	
A) naïve$_{increase\_JA}$							
Paracentral lobule	217	.002	R	4	-32	54	3.89
Paracentral lobule			L	-2	-22	58	3.63
B) naïve$_{increase\_NJA}$							
Superior occipital gyrus	277	.009	R	30	-78	20	4.09
Middle occipital gyrus			R	30	-72	32	3.97
C) coop$_{increase\_JA}$							
Caudate nucleus	3048	.000	L	-16	14	-6	5.38
Anterior cingulate cortex			R	2	20	22	5.15
Caudate nucleus			R	22	20	4	5.02
Putamen			L	-18	14	0	4.98
Thalamus	495	.000	R	18	-16	12	4.20
Thalamus			R	0	22	6	3.98

D) coop$_{increase\_NJA}$

Precuneus	2493	.000	R	10	-58	48	6.20
Intraparietal sulcus			R	40	-46	44	5.53
Superior parietal lobule			R	36	-60	62	5.09
Supramarginal gyrus			R	46	-36	40	4.96
Middle frontal gyrus	932	.000	R	50	24	34	5.91
Middle frontal gyrus			R	36	12	60	4.49
Intraparietal sulcus	878	.000	L	-36	-56	40	4.77
Inferior frontal gyrus	869	.000	R	32	60	8	5.10
Lateral orbital gyrus			R	44	48	-14	4.68
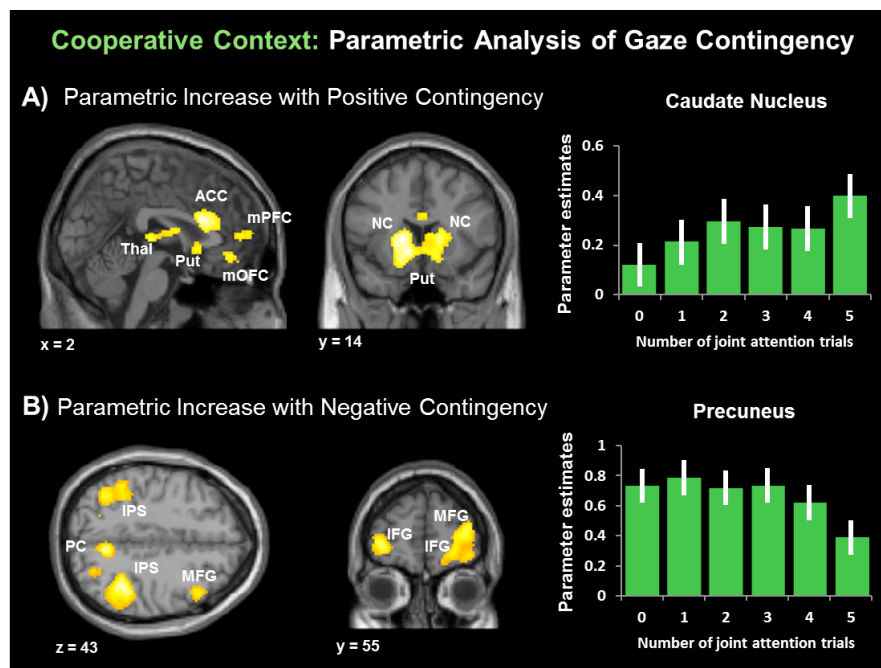
Figure 6: Temporal integration of gaze reactions in the cooperative context

(A) Parametric increases of activity with increasing numbers of joint attention trials per block in the caudate nucleus (NC), the putamen (Put), the thalamus (Thal), the anterior cingulate cortex (ACC), and the medial prefrontal cortex (mPFC). (B) Parametric increases of activity with increasing numbers of non-joint attention trials per block in the FPAN including the precuneus (PC), intraparietal sulcus (IPS), and regions of the IFG and MFG corresponding to the dorso- and ventrolateral prefrontal cortex. Parameter estimates are only shown for the maxima of the largest clusters.

4. Discussion

The present study aimed at unraveling the neural substrates of the subjective experience of engagement in social interaction with another person in real-time. We made use of an interactive eye-tracking paradigm in which participants indicated whether they experienced an interaction as human- or computer-mediated based on the gaze behavior of a virtual agent (Pfeiffer et al., 2011). Behavioral judgments of humanness and the accompanying neural activations were influenced substantially by the interaction context. When participants interacted with a naïve – interaction partner, the congruency of gaze reactions provided the major cue to humanness of the partner. When the interaction partner was explicitly cooperative, general contingency as compared to mere congruency of gaze reactions was interpreted as indicative of being in interaction with another human being. These results constitute a within-subject replication of an extensive between-subject pilot study (Pfeiffer et al., 2011), thereby implicating a universal difference in mindset between unconstrained and cooperative interactions. In-depth regression analyses of the influence of trial type with increasing trial progression revealed that there was an early influence of trial type (i.e. whether a trial was a JA or an NJA trial) in the naïve context, whereas there was an increasing influence of trial type in the cooperative context. This strongly suggests fundamental differences in the integration of information as a function of interaction context. This is complemented by neuroimaging results showing that the subjective experience of being engaged in social interaction is predicted by early ventral striatal activation in the naïve context. In contrast, during cooperative interactions, activity in this region increased differentially with increasing trial position in those interaction blocks rated as human. Taken together, these findings provide first-time evidence that the mere subjective experience of social

interaction with another human is sufficient to recruit the mesolimbic reward system, including the VTA, the VS, and the mOFC (Alcaro et al., 2007). Furthermore, they argue for different functions of the reward system during unconstrained and cooperative social interactions.

4.1 The rewarding nature of social interactions

The present results provide a fundamental extension of previous studies on social rewards. In addition to the more general observation that inert social stimuli recruit the VS similar to monetary rewards (Izuma et al., 2008; Spreckelmeyer et al., 2009), neuroeconomic studies have demonstrated an involvement of both mOFC and ventral striatum in social interactions. For instance, Rilling and colleagues scanned participants playing a Prisoner's Dilemma Game with another person or a computer (Rilling et al., 2002). In each round, players could choose to defect or to cooperate, with cooperation representing the riskier choice in terms of monetary outcome. Mutual cooperation generally led to increased activity in the anterior cingulate cortex, the mOFC, and the VS. However, the VS was not activated during cooperation with a computer, thus suggesting that activity in this area is specifically related to positive reinforcement by mutual cooperation with a human conspecific. Using multi-round trust games, it has accordingly been demonstrated that both the intention to trust someone (King-Casas et al., 2005) and another person's reputation for positive reciprocity are encoded in the striatum (Phan et al., 2010).

While these findings suggest a role of reward-related processes during human social encounters, they rely on static social stimuli or complex economic interactions and do not answer the question whether the experience of human social interaction *per se* recruits the reward system. Furthermore, in other studies employing a human-computer distinction subjects were informed a priori whether they would be interacting with another person or a computer program

(e.g. Decety et al., 2004; McCabe et al., 2001; Sanfey et al., 2003). This prevents an assessment of participants' phenomenological experience of an interaction as social (De Jaegher and Di Paolo, 2007). In contrast, the present study required participants to determine the nature of their counterpart via the interaction itself and hence focuses on the subjective experience.

The results reported here raise the question whether the engagement in social interaction and the processing of rewards share common anatomical substrates. Intriguingly, a recent study revealed a significant correlation between social reward dependence – i.e. a measure of an individual's propensity to engage in social interaction – and gray matter density in the VS and the mOFC (Lebreton et al., 2009a). Another morphometric study found a positive correlation between mOFC volume and participants' capacity to infer intentions from stories describing social interactions (Powell et al., 2010). These studies suggest an overlap between structural predispositions for the engagement and performance in social interactions and brain regions involved in the processing of rewards and thereby lend support to the major finding of this study.

While social affiliation is among our most basic needs (Baumeister and Leary, 1995), caution must be exercised because not *any* kind of social interaction is necessarily related to reward and positive affect. For example, studies using designs involving competition with a human counterpart have not found reward-related neural activations in situations of explicit competition rather than cooperation (Decety et al., 2004; Polosan et al., 2011). Furthermore, it has long been known that intergroup relations constitute an important determinant of the emotional valence of an interaction (Cikara and Bavel, 2014) – for example, interacting with a member of an out-group might foster conflict and negative emotions such as fear and disgust, while the interaction with in-group members is related to positive emotions (e.g. Rilling et al., 2008).

4.2 The 'wanting' and 'liking' of social interactions

The processing of rewards has been divided into 'wanting' and 'liking' components (Berridge et al., 2009). In the naïve context, striatal activity during the first two trials reliably predicted that an interaction will be rated as 'human'. Concordantly, post-hoc ratings revealed that participants relied on their intuition in this context (Figure 6B), and preferred thinking about the behavior of a human conspecific rather than a computer when making the decision (Figure 6D). There are two possible interpretations of this finding. The first possibility is that the VS encodes a prediction signal. It has been repeatedly shown that the VS is involved in the prediction of rewards (e.g. Schultz and Dickinson, 2000; Bromberg-Martin et al., 2010). Specifically, dopaminergic neurons of the nucleus accumbens encode a reward prediction error which relates to the difference between a predicted reward and the reward which actually occurs (Schultz et al., 1997). It is hence possible that the early activation of the VS in human-rated blocks within the naïve context represents a reward prediction signal. It has further been argued that neural activity related to reward prediction should be involved in encoding the contingency between a stimulus potentially predicting reward and the actual reward (Schultz, 2006). In the case of our study, however, there is no measurable contingency between the second trial of a block and the continuing trial progression because trial order is completely randomized. This implicates that the VS should not be able to calculate a reliable reward prediction error.

In other words, the agent's behavior during the first two trials is actually inconclusive regarding the nature of the counterpart. Therefore, it is possible that the early striatal activation in human-rated blocks in the naïve interaction context might relate to the 'wanting' component associated with the human need to interact (Baumeister and Leary, 1995). Albeit speculative, this interpretation would be consistent with the previous observation that the VS conveys automatic

incentive signals to the mOFC during initial stages of impression formation (Kim et al., 2007). If such an association of the VS with the motivational drive for interaction could be confirmed in future studies, these results might contribute to a larger picture of the VS as an automatic valuation system which encodes preferences irrespective of stimulus modality and task demands (Lebreton et al., 2009b).

Although our results do not provide a direct proof of this idea, one might speculate that if a need for interaction represents the 'wanting' component, its fulfillment by the actual experience of engagement in interaction might correspond to reward 'liking'. Behaviorally, this idea gains support by participants' post-experience ratings which indicate that social interactions are experienced as more pleasant than non-social interactions (Figure 2C). The subjective hedonic experience of rewards has been attributed to the mOFC rather than the VS (for recent reviews see Diekhof et al., 2012; Peters and Büchel, 2010). Interestingly, the mOFC was active during blocks rated as 'human' irrespective of interaction context, which might possibly reflect a general 'liking' of being engaged in interaction. Although task structure does not allow any direct proof of this interpretation, post-experiment questionnaires did not provide any hint to a difference in the perceived pleasantness of naïve and cooperative human interactions. This might be interpreted as evidence that neural activity related to the subjective experience of reward should not be affected by the context of the interaction. Indeed, our results show that the interaction context exclusively modulates VS function and thus implicate that activation of the mOFC – and not the VS – might be related to the pleasantness of being in interaction with another human. At present, however, explanations which address such detailed functional segregations must remain speculative. Future studies involving carefully devised online

interaction paradigms specifically aimed at differentiating the desire for human interaction and the pleasure of being engaged in interaction are needed to confirm these hypotheses.

In a more general framework, the concept that social interaction represents a reward in itself is supported by previous studies indicating that social exclusion – which can be regarded as an externally forced disengagement from social interaction – is correlated with activity in the pain network (Eisenberger et al., 2003). Specifically, novel social encounters create instant expectations regarding another person's behavior whose violation is correlated with activation of the anterior cingulate cortex, an important component of this network (Somerville et al., 2006). Moreover, the fulfillment of such expectations might recruit the reward system and thereby pave the way to the establishment of a prolonged relationship (Baumeister and Leary, 1995) which is consequently assigned with high reward-value in future interactions (Fareri et al., 2012).

4.3 Accumulation of value in cooperative interactions

Importantly, there are ultimate as well as proximate definitions of cooperation. The former relate to the survival value of a behavior, whereas the latter relate to the underlying mechanisms (West et al., 2007). In its ultimate definition, cooperation is a behavior selected to create mutual benefit for an actor and a recipient (e.g. Fehr and Rockenbach, 2004). It often requires the actor to sacrifice an immediate reward (i.e. reward discounting) in order to create a mutual, but delayed benefit for both actor and recipient (Axelrod, 1984). However, cooperation has also been described as a form of behavioral coordination which is a proximate mechanism required for obtaining mutual benefit (e.g. Noë, 2006; Taborsky, 2007). Obviously, the introduction of a cooperative interaction partner in the present study is not directly related to evolutionary fitness but rather to behavioral coordination. Accordingly, behavioral results (Figure 1D) showed that –

irrespective of overall congruency – coordinated behavior is judged as indicative of a human interaction partner. This replicates results of a behavioral pilot study which assessed naïve and cooperative interactions in a between-subject design (Pfeiffer et al., 2011).

Neuroimaging data complemented these behavioral observations. As expected, the cooperative context was also associated with reward processing. Notably, however, there was no early activation of the VS as in the naïve context, but a gradual increase of striatal activity with increasing trial progression. Considering that the human interaction partner allegedly facilitated participants' decisions, consistent behavior had to be detected by an accumulation of information over time rather than by trusting initial intuitions (Pfeiffer et al., 2011). The question arising here is whether the differential increase of VS activity in blocks rated as human relates to a general accumulation of evidence for a choice (e.g. Heekeren et al., 2008) or to an accumulation of value inherent to that choice (e.g. Rangel et al., 2008). In our opinion, the differential increase of VS activity over cooperative interaction blocks rated as human is reflective of value rather than evidence accumulation. The latter has mainly been investigated in studies on perceptual decision-making and is considered to be an effortful and noisy process due to fuzzy category boundaries (Ratcliff and McKoon, 2008). In contrast, the decision criterion for social interaction during cooperation is comparatively straightforward because only behavior that is highly contingent is taken as evidence for the agent being controlled by another human.

A recent study directly investigated whether VS activation during decision-making is related to the accumulation of evidence in general or the accumulation of value (Gluth et al., 2012). In a buying task, a number of positive and negative ratings of a product were sequentially disclosed to participants. The general disclosure of the ratings correlated with activity in the pre-SMA and anterior insula. As this activity was neither influenced by the valence of the ratings nor

by the buying decision, it was argued to reflect the accumulation of evidence. In contrast, the VS specifically updated the representation of value when positive ratings were disclosed. This can be directly compared to our study: In each trial, novel information regarding the nature of the interaction partner is revealed in the form of the agent's gaze reactions. Each trial of an interaction block rated as 'human' during cooperative interactions therefore *must* have been interpreted as *positive evidence* that the agent is controlled by a human interaction partner. Furthermore, the specificity of striatal activation for 'human'-rated blocks argues against general evidence accumulation. The differential linear increase of VS activity with these trials thus reflects the accumulation of value rather than evidence per se (Figure 3C), which is consistent with the previously described role of this region in encoding benefit signals during decision-making (Basten et al., 2010).

4.4 Contextual modulation of gaze processing in social interaction

The contextual modulation of the neural processing of the agent's gaze reactions allows a deeper understanding of the integration of information underlying reward-based decisions. First insights come from event-related analyses of JA and NJA. In the naïve context, activity in the mPFC and aSTS was confined to single joint attention trials. The involvement of these regions in the inference of mental states suggests that the social salience of gaze behavior is processed in a trial-by-trial fashion (Schilbach et al., 2010; Redcay et al., 2012). Furthermore, activation of the reward system is confined to single events of joint, but not non-joint attention (Gordon et al., 2013; Schilbach et al., 2010), thereby indicating that making someone follow your gaze represents a reward in itself. In contrast, in cooperative interactions, single events are only indicative of a human counterpart when part of contingent behavior. Accordingly, positive

contingency was processed by the dorsal striatum, while negative contingency recruited the FPAN. Possibly, this interplay of reward and attention networks reflects a distribution of cognitive resources required to ensure that joint attention is *only* considered as a social cue in cases of high contingency. The dorsal striatum is also recruited in iterated trust-games (King-Casas et al., 2005), which require participants to monitor a cooperators' actions across multiple rounds and demand similar temporal binding of contingent information as the present task (McCabe et al., 2001). Furthermore, it plays a role whenever participants experience contingency between their actions and a reward (Tricomi et al., 2004). This is consistent with the recent observation that successful initiation of joint attention results in an increased experience of agency (Pfeiffer et al., 2012), and hence suggests that the reward value of cooperative interactions relies on the experience of contextually meaningful contingencies between one's own actions and another individual's behavior.

4.5 Limitations of the present study

Despite the novelty of our findings, there are several limitations with respect to study design and the interpretation of the results.

First of all, it must be emphasized that *no* interaction in our study is an actual interaction with a real human interaction partner. It would therefore be misleading to claim that the paper discusses the neural substrates of engagement in human social interaction. Instead, task design elicited participants' *subjective experience* of being engaged in interaction with a human counterpart. Although it is conceivable that the underlying neural mechanisms are similar, the distinction between subjective experience and actual engagement must be kept in mind when interpreting the results.

Directly related to the first limitation, a more general limitation is imposed by the interaction task itself. Although ecological validity and behavioral realism of the interactive eye-tracking paradigm are high (e.g. Fox et al., 2009; Schroeder, 2002), the gaze-based interactions are still comparably inflexible in contrast to everyday social interactions which are rich in dynamics and sensory detail. However, while some EEG studies have investigated more dynamic interactions (e.g. Dumas et al., 2010; Lachat et al., 2012; Tognoli et al., 2007), possibilities to do so are limited in MRI scanners which are required to detect activity in subcortical structures related to motivational and reward processes (for a methodological discussion see Pfeiffer et al., 2013). For this reason, neuroimaging studies of real-time social interactions are currently still limited by a trade-off between social realism and technical feasibility. To move towards even more naturalistic interactions, future gaze-based tasks could involve important modulators of our experience of social interaction such as emotional expressions (Adams and Kleck, 2005) or culture (Krämer et al., 2013).

Third, it must be noted here that the present study was not explicitly designed to disentangle the neural mechanisms underlying the motivational desire for social interaction on the one hand and the pleasantness of being engaged in interaction on the other. To some extent, the corresponding discussion is thus speculative. More work is clearly needed to uncover the precise roles of the VS and mOFC in encoding the motivational aspects of 'online' social interactions (Schilbach et al., 2013). Such studies would also need to include refined behavioral measures of the subjective pleasantness of social interactions which could then be included in parametric analyses of functional imaging data. In addition, it would be helpful to obtain indicators of participants' propensity to engage in social interaction, for example by assessing

their social reward dependence using Cloninger's Temperament and Character Inventory (Cloninger et al., 1993).

Besides these conceptual issues, there are also some methodological caveats. The first concerns the lack of randomization of the order of interaction contexts due to the importance of keeping participants naïve with respect to the task in the naïve context. Despite the fact that behavioral results replicate those of a between-subject study (Pfeiffer et al., 2011) and that participants were given an extended practice session to prevent novelty and habituation effects, we cannot fully exclude the possibility of sequence effects. Future studies should therefore use designs allowing a more direct comparison of unconstrained and cooperative interactions.

The final methodological aspect relates to the limited number of blocks per experimental condition (i.e. six), which decreases statistical power and thereby prevents in-depth assessment of mechanisms relating to reward prediction and accumulation of evidence in general. As reward prediction plays an important role in virtually all decision-making tasks (Berridge et al., 2009; Daniel and Pollmann, 2014), future studies would have to be designed in such a way that the motivational and learning-related aspects of VS function can clearly be disentangled.

Taken together, we believe that in light of the scarcity of studies investigating social encounters in real-time and despite the limitations discussed above, the present study provides important insights into the motivational mechanisms underlying the subjective experience of engagement in social interaction (Becchio et al., 2010; De Jaegher et al., 2010; Schilbach et al., 2013).

5. Conclusion

In sum, this study demonstrates how the neural systems possibly sustaining the active engagement in social interaction can be examined using a novel, interactive paradigm. It provides first evidence that the social nature of human primates rests upon an urge to interact and upon the rewarding nature of the active participation in social interactions. The present results hence further endorse the proposal of a human predisposition for cooperation (Tomasello, 2009) by suggesting that a fundamental motive for cooperation could be the sustainment of an interaction with another person. A final note concerns the potential of the present study to foster our understanding of autism spectrum disorders (ASD). It has recently been claimed that autism is primarily an impairment of social motivation, with disturbances of both 'wanting' and 'liking' of social rewards (Chevallier et al., 2012). Although first studies support this idea by demonstrating hypoactivation of the nucleus accumbens during the anticipation of social rewards (Delmonte et al., 2012; Richey et al., 2014), there are no studies examining the brains of persons with ASD while they are actively engaged in social interaction. It is hence conceivable that the present paradigm could provide a tool to test the social motivation hypothesis under ecologically valid but controlled conditions.

6. References

Adams RB Jr, Kleck RE (2005) Effects of direct and averted gaze on the perception of facially communicated emotion. Emotion 5:3–11.

Alcaro A, Huber R, Panksepp J (2007) Behavioral functions of the mesolimbic dopaminergic system: an affective neuroethological perspective. Brain Res Rev 56:283–321.

Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. Hum Brain Mapp 7:254–266.

Astolfi L, Toppi J, De Vico Fallani F, Vecchiato G, Salinari S, Mattia D, Cincotti F, Babiloni F (2010) Neuroelectrical hyperscanning measures simultaneous brain activity in humans. Brain Topogr 23:243–256.

Axelrod R (1984) The Evolution of Cooperation. New York: Basic Books.

Babiloni F, Cincotti F, Mattia D, De Vico Fallani F, Tocci A, Bianchi L, Salinari S, Marciani M, Colosimo A, Astolfi L (2007) High resolution EEG hyperscanning during a card game. Conf Proc IEEE Eng Med Biol Soc 2007:4957–4960.

Báez-Mendoza R, Schultz W (2013) The role of the striatum in social behavior. Front Neurosci 7:233.

Basten U, Biele G, Heekeren HR, Fiebach CJ (2010) How the brain integrates costs and benefits during decision making. Proc Natl Acad Sci USA 107:21767–21772.

Baumeister RF, Leary MR (1995) The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 117:497–529.

Becchio C, Sartori L, Castiello U (2010) Toward You: The Social Side of Actions. Curr Dir Psychol Sci 19:183–188.

Berridge KC, Robinson TE, Aldridge JW (2009) Dissecting components of reward: "liking", "wanting", and learning. Curr Opin Pharmacol 9:65–73.

Bohil CJ, Alicea B, Biocca FA (2011) Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 12:752–762.

Bromberg-Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68:815–834.

Chevallier C, Kohls G, Troiani V, Brodkin ES, Schultz RT (2012) The social motivation theory of autism. Trends Cogn Sci 16:231–239.

Cikara M, Bavel JJV (2014) The Neuroscience of Intergroup Relations - An Integrative Review. Perspect Psychol Sci 9:245–274.

Cloninger CR, Svrakic DM, Przybeck TR (1993) A psychobiological model of temperament and character. Arch Gen Psychiatry 50:975–990.

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192–205.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215.

Daniel R, Pollmann S (2014) A universal role of the ventral striatum in reward-based learning: Evidence from human studies. Neurobiol Learn Mem.

De Jaegher H, Di Paolo E, Gallagher S (2010) Can social interaction constitute social cognition? Trends Cogn Sci 14:441–447.

De Jaegher HD, Di Paolo ED (2007) Participatory sense-making. Phenom Cogn Sci 6:485–507.

Decety J, Jackson PL, Sommerville JA, Chaminade T, Meltzoff AN (2004) The neural bases of cooperation and competition: an fMRI investigation. Neuroimage 23:744–751.

Delmonte S, Balsters JH, McGrath J, Fitzgerald J, Brennan S, Fagan AJ, Gallagher L (2012) Social and monetary reward processing in autism spectrum disorders. Mol Autism 3:7.

Diekhof EK, Kaps L, Falkai P, Gruber O (2012) The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude − An activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. Neuropsychologia 50:1252–1266.

Dumas G, Nadel J, Soussignan R, Martinerie J, Garnero L (2010) Inter-brain synchronization during social interaction. PLoS ONE 5:e12166.

Duvernoy HM (1999) The human brain: Surface, three-dimensional sectional anatomy with MRI, and blood supply, 2nd ed. Heidelberg: Springer.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25:1325–1335.

Eisenberger NI, Lieberman MD, Williams KD (2003) Does Rejection Hurt? An fMRI Study of Social Exclusion. Science 302:290–292.

Fareri DS, Niznikiewicz MA, Lee VK, Delgado MR (2012) Social network modulation of reward-related signals. J Neurosci 32:9045–9052.

Fehr E, Camerer CF (2007) Social neuroeconomics: The neural circuitry of social preferences. Trends Cogn Sci 11:419–427.

Fehr E, Rockenbach B (2004) Human altruism: economic, neural, and evolutionary perspectives. Curr Opin Neurobiol 14:784–790.

Fox J, Arena D, Bailenson JN (2009) Virtual reality. J Media Psychol 21:95–113.

Frith CD (2007) The social brain? Phil Trans R Soc B 362:671–678.

Gallagher HL, Jack AI, Roepstorff A, Frith CD (2002) Imaging the intentional stance in a competitive game. Neuroimage 16:814–821.

Gluth S, Rieskamp J, Büchel C (2012) Deciding when to decide: Time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. J Neurosci 32:10686–10698.

Gordon I, Eilbott JA, Feldman R, Pelphrey KA, Vander Wyk BC (2013) Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses. Soc Neurosci 8:544–554.

Hari R, Kujala MV (2009) Brain basis of human social interaction: From concepts to brain imaging. Physiol Rev 89:453–479.

Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. Nat Rev Neurosci 9:467–479.

Izuma K, Saito DN, Sadato N (2008) Processing of social and monetary rewards in the human striatum. Neuron 58:284–294.

Kim H, Adolphs R, O'Doherty JP, Shimojo S (2007) Temporal isolation of neural processes underlying face preference decisions. Proc Natl Acad Sci USA 104:18253–18258.

King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: Reputation and trust in a two-person economic exchange. Science 308:78–83.

Krach S, Paulus FM, Bodden M, Kircher T (2010) The rewarding nature of social interactions. Front Neurosci 4:22.

Krämer K, Bente G, Luo S, Pfeiffer UJ, Han S, Vogeley K (2013) Influence of ethnic group-membership and gaze direction on the perception of emotions. A cross-cultural study between Germany and China. PLoS ONE 8:e66335.

Lachat F, Hugueville L, Lemaréchal J-D, Conty L, George N (2012) Oscillatory Brain Correlates of Live Joint Attention: A Dual-EEG Study. Front Neurosci 6:156.

Lebreton M, Barnes A, Miettunen J, Peltonen L, Ridler K, Veijola J, Tanskanen P, Suckling J, Jarvelin M-R, Jones PB, Isohanni M, Bullmore ET, Murray GK (2009a) The brain structural disposition to social interaction. Eur J Neurosci 29:2247–2252.

Lebreton M, Jorge S, Michel V, Thirion B, Pessiglione M (2009b) An automatic valuation system in the human brain: evidence from functional neuroimaging. Neuron 64:431–439.

Macey PM, Macey KE, Kumar R, Harper RM (2004) A method for removal of global effects from fMRI time series. Neuroimage 22:360–366.

Maslow AH (1943) A theory of human motivation. Psych Rev 50:370–396.

Materna S, Dicke PW, Thier P (2008) Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. J Cogn Neurosci 20:108–119.

McCabe K, Houser D, Ryan L, Smith V, Trouard T (2001) A functional imaging study of cooperation in two-person reciprocal exchange. Proc Natl Acad Sci USA 98:11832–11835.

McDonald JH (2009) Handbook of biological statistics, 2nd ed. Baltimore, MD: Sparky House Publishing.

Mundy P, Newell L (2007) Attention, joint attention, and social cognition. Curr Dir Psychol Sci 16:269–274.

Noë R (2006) Cooperation experiments: coordination through communication versus acting apart together. Anim Behav 71:1–18.

Noonan MP, Kolling N, Walton ME, Rushworth MFS (2012) Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. Eur J Neurosci 35:997–1010.

Peters J, Büchel C (2010) Neural representations of subjective reward value. Behav Brain Res 213:135–141.

Pfeiffer U, Schilbach L, Timmermans B, Jording M, 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. Front Psychol 3:537.

Pfeiffer UJ, Timmermans B, Bente G, Vogeley K, Schilbach L (2011) A non-verbal Turing test: Differentiating mind from machine in gaze-based social interaction. PLoS ONE 6:e27591.

Pfeiffer UJ, Vogeley K, Schilbach L (2013) From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. Neurosci Biobehav Rev 37:2516–2528.

Phan KL, Sripada CS, Angstadt M, McCabe K (2010) Reputation for reciprocity engages the brain reward center. Proc Natl Acad Sci USA 107:13099–13104.

Polosan M, Baciu M, Cousin E, Perrone M, Pichat C, Bougerol T (2011) An fMRI study of the social competition in healthy subjects. Brain Cogn 77:401–411.

Powell JL, Lewis PA, Dunbar RIM, García-Fiñana M, Roberts N (2010) Orbital prefrontal cortex volume correlates with social cognitive competence. Neuropsychologia 48:3554–3562.

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545–556.

Ratcliff R, McKoon G (2008) The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. Neural Comput 20:873–922.

Redcay E, Dodell-Feder D, Pearrow MJ, Mavros PL, Kleiner M, Gabrieli JDE, Saxe R (2010) Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. Neuroimage 50:1639–1647.

Redcay E, Kleiner M, Saxe R (2012) Look at this: The neural correlates of initiating and responding to bids for joint attention. Front Neurosci 6:169.

Reddy V, Morris P (2004) Participants Don't Need Theories Knowing Minds in Engagement. Theory Psychology 14:647–665.

Richey JA, Rittenberg A, Hughes L, Damiano CR, Sabatino A, Miller S, Hanna E, Bodfish JW, Dichter GS (2014) Common and distinct neural features of social and non-social reward processing in autism and social anxiety disorder. Soc Cogn Affect Neurosci 9:367–377.

Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002) A neural basis for social cooperation. Neuron 35:395–405.

Rilling JK, Dagenais JE, Goldsmith DR, Glenn AL, Pagnoni G (2008) Social cognitive neural networks during in-group and out-group interactions. Neuroimage 41:1447–1461.

Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. Annu Rev Psychol 62:23–48.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. Science 300:1755–1758.

Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, Vogeley K (2013) Toward a second-person neuroscience. Behav Brain Sci 36:393–414.

Schilbach L, Wilms M, Eickhoff SB, Romanzetti S, Tepest R, Bente G, Shah NJ, Fink GR, Vogeley K (2010) Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. J Cogn Neurosci 22:2702–2715.

Schroeder R (2002) The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments. London: Springer.

Schultz W (2006) Behavioral theories and the neurophysiology of reward. Annu Rev Psychol 57:87–115.

Schultz W, Dayan P, Montague PR (1997) A Neural Substrate of Prediction and Reward. Science 275:1593–1599.

Schultz W, Dickinson A (2000) Neuronal Coding of Prediction Errors. Annu Rev Neurosci 23:473–500.

Somerville LH, Heatherton TF, Kelley WM (2006) Anterior cingulate cortex responds differentially to expectancy violation and social rejection. Nat Neurosci 9:1007–1008.

Spreckelmeyer KN, Krach S, Kohls G, Rademacher L, Irmak A, Konrad K, Kircher T, Gründer G (2009) Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. Soc Cogn Affect Neurosci 4:158–165.

Taborsky M (2007) Cooperation built the Tower of Babel. Behav Processes 76:95–99.

Tognoli E, Lagarde J, DeGuzman GC, Kelso JAS (2007) The phi complex as a neuromarker of human social coordination. Proc Natl Acad Sci USA 104:8190–8195.

Tomasello M (2009) Why we cooperate. Cambridge, MA: The MIT Press.

Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of caudate activity by action contingency. Neuron 41:281–292.

West SA, Griffin AS, Gardner A (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J Evol Biol 20:415–432.

Wilms M, Schilbach L, Pfeiffer U, Bente G, Fink GR, Vogeley K (2010) It's in your eyes - Using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. Soc Cogn Affect Neurosci 5:98–107.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996) A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58–73.