A neural link between affective understanding and interpersonal attraction

Silke Anders[a,1], Roos de Jong[a], Christian Beck[a], John-Dylan Haynes[b,c,d], and Thomas Ethofer[e,f]

[a]Social and Affective Neuroscience, Department of Neurology, Universität zu Lübeck, 23562 Luebeck, Germany; [b]Bernstein Center for Computational Neuroscience, Charité Universitätsmedizin Berlin, 10115 Berlin, Germany; [c]Berlin Center for Advanced Neuroimaging, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany; [d]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10099 Berlin, Germany; [e]Department of Biomedical Magnetic Resonance, University of Tübingen, 72076 Tuebingen, Germany; and [f]Clinic for Psychiatry and Psychotherapy, University of Tübingen, 72076 Tuebingen, Germany

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved February 11, 2016 (received for review August 20, 2015)

Being able to comprehend another person's intentions and emotions is essential for successful social interaction. However, it is currently unknown whether the human brain possesses a neural mechanism that attracts people to others whose mental states they can easily understand. Here we show that the degree to which a person feels attracted to another person can change while they observe the other's affective behavior, and that these changes depend on the observer's confidence in having correctly understood the other's affective state. At the neural level, changes in interpersonal attraction were predicted by activity in the reward system of the observer's brain. Importantly, these effects were specific to individual observer–target pairs and could not be explained by a target's general attractiveness or expressivity. Furthermore, using multivoxel pattern analysis (MVPA), we found that neural activity in the reward system of the observer's brain varied as a function of how well the target's affective behavior matched the observer's neural representation of the underlying affective state: The greater the match, the larger the brain's intrinsic reward signal. Taken together, these findings provide evidence that reward-related neural activity during social encounters signals how well an individual's "neural vocabulary" is suited to infer another person's affective state, and that this intrinsic reward might be a source of changes in interpersonal attraction.

affective communication | confidence | intrinsic reward | multivoxel pattern analysis | human social relations

Finding the "right" cooperation partner is an important task for individuals living in complex environments that require social interaction and cooperation. To accomplish a common goal, interaction partners must understand and continuously update information about their partner's current intentions, motivation, and affect, anticipate the other's behavior, and adapt their own behavior accordingly. From a sociobiological point of view, one thus might expect that evolution has favored a neural mechanism that permits individuals to select other individuals as their cooperation partners whose behavior and communication signals they can easily decode. However, the neural mechanisms that control human interpersonal attraction and the selection of cooperation partners are not well-understood.

Several influential theories in social psychology have stressed the role of reward in interpersonal attraction (1, 2). The idea is that if a social encounter with another person is rewarding, then the reward will become associated with the other person, resulting in interpersonal attraction (2–4). Until recently, neuroscientific research into interpersonal attraction has focused mainly on determining the neural mechanism underlying the evaluation of others based on the physical attractiveness of their faces (e.g., 5–11). These studies consistently show that neural activity in the ventral striatum and medial orbitofrontal cortex (mOFC), core regions of the brain's reward system that also respond to food and money (12, 13), increases in response to faces that are perceived as attractive. Other studies show that these brain regions also respond to another person's prosocial behavior (14–20). Although this research documents the role of the brain's reward system in interpersonal attraction, it does not explain why social encounters often result in relational effects in interpersonal attraction (21) such that one individual is particularly attracted to one person whereas another individual is more attracted to another person. Here we focus on the role of nonverbal understanding in interpersonal attraction. Specifically, we ask whether the human brain possesses a neural mechanism that permits individuals to select and approach other individuals as interaction partners whose affective behavior they can easily understand.

Recent work on perceptual learning provides a first hint that the brain's reward system might play an important role not only in signaling facial attractiveness but also in the individual adjustment of interpersonal attraction during social interaction. This work suggests that whenever the brain evaluates sensory information, it generates a neural signal in the ventral striatum that reflects the amount of evidence available for stimulus evaluation (22) and that, at the experiential level, is associated with subjective confidence (23). Importantly, it has been proposed that such intrinsic confidence signals can act as positive reinforcement signals (24). We hypothesized that a similar confidence signal, reflecting the amount of evidence available to decode another person's nonverbal behavior, might serve as an intrinsic reward that becomes associated with the interaction partner and thereby increases or decreases the perceiver's interpersonal attraction toward the interaction partner during social encounters.

Considering further evidence from social neuroscience, we reasoned that an individual's confidence in their understanding of the other's behavior might reflect how well the individual's "neural

Significance

Humans interacting with other humans must be able to understand their interaction partner's affect and motivations, often without words. We examined whether people are attracted to others whose affective behavior they can easily understand. For this, we asked participants to watch different persons experiencing different emotions. We found the better a participant thought they could understand another person's emotion the more they felt attracted toward that person. Importantly, these individual changes in interpersonal attraction were predicted by activity in the participant's reward circuit, which in turn signaled how well the participant's "neural vocabulary" was suited to decode the other's behavior. This research elucidates neurobiological processes that might play an important role in the formation and success of human social relations.

Author contributions: S.A. and R.d.J. designed research; S.A. and R.d.J. performed research; S.A. analyzed data; C.B. provided analysis tools; and S.A., J.-D.H., and T.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: silke.anders@neuro.uni-luebeck.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516191113/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1516191113

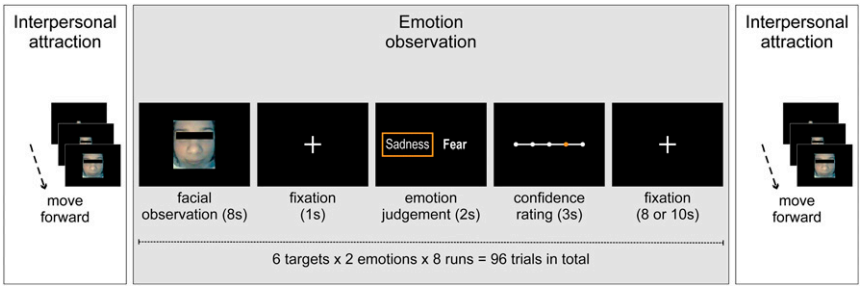

Fig. 1. Experimental design. To measure changes in interpersonal attraction during emotion observation, the participants' attraction toward each target was assessed before (*Left*) and after (*Right*) emotion observation. (*Middle*) A sample emotion observation trial is shown (time intervals and screen shots are taken from experiment II). A trial consisted of a facial observation period during which a short video clip of the target experiencing fear or sadness was shown, followed by a fixation cross, an emotion judgment period, and a confidence rating period. Responses were given with a button box and fed back to the observer (orange frame around the selected emotion and orange dot on the confidence scale).

vocabulary" is suited to decode and interpret the target's behavior. Neurobiological accounts of social cognition suggest that when humans evaluate the inner state of another person, they implicitly use neural representations of their own states as reference (e.g., 25–28), and empirical studies have provided evidence that is consistent with this idea (e.g., 29–39). Thus, we predicted that an individual's intrinsic confidence that they correctly understood another person's affective state would reflect how well the target's behavior matches the observer's neural representation of the underlying state.

To examine the role of mutual understanding in interpersonal attraction, we conducted two experiments, a behavioral experiment (experiment I) and a combined behavioral–fMRI (functional magnetic resonance imaging) experiment (experiment II) with two independent samples of volunteers. Participants in both experiments were shown short video clips of six different female targets who experienced and facially expressed two different emotions, fear or sadness. After each video, participants were asked to judge the target's affective state (fear or sadness) and to indicate how confident they were that they had correctly understood the target's affective state (Fig. 1). Before and after emotion observation, we assessed the participants' interpersonal attraction toward each target at two levels. First, we asked participants to enlarge a small picture of each target on a computer screen by repeatedly pressing a button until the picture had a size that corresponded to a subjectively pleasant conversational distance. The number of button presses executed to enlarge the picture of a given target was taken as measure of the participant's approach behavior toward that target (modified after ref. 6). Next, participants were given three statements about each target and asked to indicate how much they agreed with each of these statements (Table 1). This way, we derived a motivational–behavioral measure (approach behavior) and a self-report measure of each participant's interpersonal attraction toward each target before and after emotion observation. This experimental design allowed us to link the participants' confidence that they correctly understood the targets' affective state to individual changes in interpersonal attraction. In experiment II, we additionally measured the participants' brain activity during emotion observation. This enabled us to identify neural activity in the brain's reward system that predicted the participants' self-reported confidence in their emotion judgments and to link this neural activity to changes in interindividual attraction. Finally, participants in experiment II completed an emotion experience task immediately after the emotion observation task in which they were asked to experience and express the two emotions (fear and sadness) themselves, using instructions similar to those that were used when recording the videos of the targets (37). This permitted us to compare the patterns of neural activity elicited during emotion observation to those associated with the observer's own emotional experience. We refer to the level of correspondence between these patterns as neural observation–experience matching (NOE matching).

Experiment I tested whether observing another person's affective behavior can lead to individual changes in interpersonal attraction, and whether these changes are predicted by the observer's subjective confidence that they correctly understood the other's affective state. Experiment II validated the findings of experiment I and additionally investigated the neural mechanisms that might mediate between affective understanding and interpersonal attraction. To this end, we first examined whether the participants' subjective confidence in their emotion judgments was predicted by neural activity in the reward system of their brains. Second, we used multivoxel pattern analysis (MVPA) (40) to examine the relation between subjective confidence, neural confidence signals, and NOE matching. Both analyses were performed in a cross-validated hierarchical approach, using whole-brain analyses to identify relevant brain regions, followed by region-of-interest (ROI) analyses to examine the relation between neural signals within these regions and individual changes in interpersonal attraction (Fig. S1).

Table 1. Statements used to assess interpersonal attraction

Original statement (German)	Translation
Willingness to meet	
Ich würde Sie gerne im echten Leben treffen	I would like to meet her in real life
Expectation of intimate communication	
Ich habe das Gefühl, dass sie mich verstehen würde	I feel that she would understand me
Ich glaube, dass ich mit ihr über persönliche Probleme reden könnte	I think I could discuss personal problems with her

Participants were asked to indicate how much they agreed with each statement on a Likert-type 7-point scale ranging from 1 (not at all) to 7 (definitely). The first statement estimated the participant's overt willingness to meet the target; the last two questions were averaged to estimate the participant's expectation that they could have an intimate communication with the target.

Results

Experiment I.

Emotion judgments and self-reported confidence. In the behavioral experiment, observers (21 women, 19 men) correctly labeled the target's emotional state in the majority of trials (hit rate 74 ± 3.6% [mean ± SEM], $T[39] = 21$, $P < 0.001$), and the observers' average self-reported confidence for a given target closely reflected the actual correctness of their emotion judgments for this target (mean $r = 0.71 \pm 0.07$ [back-transformed mean of Fisher-transformed correlation coefficients], $T[39] = 9.8$, $P < 0.001$). This indicates that observers had a valid internal model that allowed them to infer the targets' affective state and to accurately estimate the correctness of their understanding. A two-way ANOVA with between-subject factor observer (40 levels), within-subject factor target (6 levels), and the observers' self-reported confidence as dependent variable ($n = 40 \times 6 \times 16 = 3{,}840$ trials) revealed, in addition to a significant main effect of target ($F[5,195] = 95$, $P < 0.001$, $\text{eta}^2 = 0.71$), a significant observer-by-target interaction ($F[195,3600] = 2.4$, $P < 0.001$, $\text{eta}^2 = 0.12$). This indicates that an observer's subjective confidence that they correctly understood a target's affective state did not only depend on the observer's general ability to recognize facial emotions, or the target's general ability to express their emotion, but also on how well a particular observer could "tune in" to a particular target's affect.

Changes in interpersonal attraction. Overall, observers conducted more button presses to "approach" the observed targets after than before emotion observation ($T[39] = 3.7$, $P < 0.001$; please see Table S1 for results for self-reported interpersonal attraction). This is in line with previous research that shows that familiarity increases interpersonal attraction (e.g., 41). However, the critical question in the current study was whether emotion observation can lead to changes in interpersonal attraction that differ between observers, such that one observer feels more attracted to a particular target after emotion observation whereas another observer feels less attracted to the same target, even though both observers saw exactly the same behavior. Intriguingly, this was the case. Between-subject variability of the number of button presses observers conducted to approach a given target (measured as the width of the 66% interval of button presses for each target) was significantly larger after emotion observation (mean width of the 66% interval for each target, 7.4 ± 0.6 button presses) than before emotion observation (mean width of the 66% interval for each target, 6.2 ± 0.4 button presses) ($T[5] = 2.7$, $P = 0.040$; Fig. 2; please see Table S1 for results for self-reported interpersonal attraction).

Self-reported confidence and individual changes in interpersonal attraction. Next, we asked whether the observed changes in interpersonal attraction were predicted by the observers' confidence that they correctly understood the targets' affective state. To test this, we computed partial correlations between each observer's confidence ratings and postobservation attraction scores for each target. Importantly, to ensure that this correlation was not driven by the observer's initial attraction toward the targets, any variance that could be explained by preobservation attraction (Table S2) was removed from confidence ratings and postobservation attraction scores. This revealed significant positive partial correlations between the observer's confidence and postobservation attraction both at the behavioral–motivational level (approach behavior) and at the level of self-report (Table 2). To further control for potential differences between targets in physical attractiveness and facial expressivity, we removed average confidence ratings and average attraction scores for each target ("general target effects," ref. 21) from each individual dataset and performed the same partial correlation analyses as above. As predicted, partial correlations between self-reported confidence and self-reported interpersonal attraction remained significant after general target effects had been removed (Table 2). This indicates that the link between confidence and changes in interpersonal attraction cannot be fully explained by a target's general attractiveness or expressivity.

Experiment II.

Behavioral data: Emotion judgments, self-reported confidence, and individual changes in interpersonal attraction. Behavioral data of experiment II largely replicated those of experiment I. Observers (28 women, 24 men) correctly labeled the target's emotional state in the majority of trials (mean hit rate 75 ± 2.5%, $T[51] = 30$, $P < 0.001$), and their self-reported confidence closely reflected the actual correctness of their emotion judgments for a given target (mean back-transformed $r = 0.42 \pm 0.01$, $T[51] = 4.1$, $P < 0.001$). Furthermore, there was a significant observer-by-target interaction in self-reported confidence similar to that observed in experiment I (two-way ANOVA with between-subject factor observer [52 levels], within-subject factor target [6 levels], and the observers' confidence ratings as dependent variable [$n = 52 \times 6 \times 16 = 4{,}992$ trials]; main effect target, $F[5,255] = 41$, $P < 0.001$, $\text{eta}^2 = 45$; observer-by-target interaction, $F[255,4624] = 2.9$, $P < 0.001$, $\text{eta}^2 = 0.14$). As in experiment I, observers conducted more button presses to approach the observed targets after than before emotion observation ($T[51] = 2.7$, $P = 0.005$), and between-subject variability of the number of button presses observers conducted to approach a given target increased significantly from preobservation (mean width of the 66% interval for each target 7.8 ± 0.5 button presses) to postobservation (mean width of the 66% interval for each target 9.1 ± 0.5 button presses)

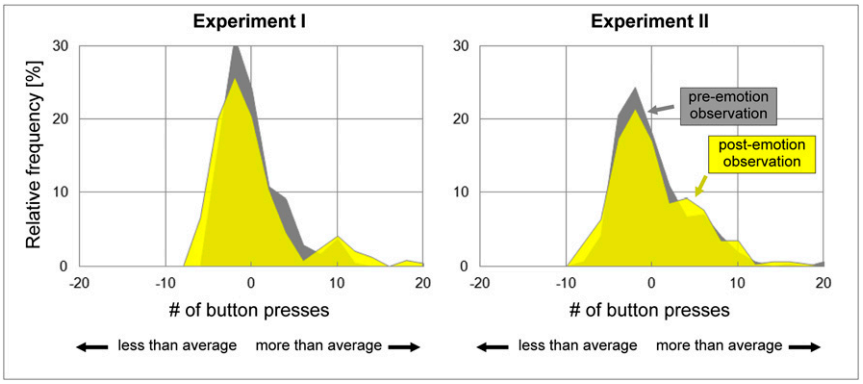

Fig. 2. Interindividual variability in the observers' approach behavior toward the targets before and after emotion observation. Data were first centered, separately for each target and pre- and postobservation runs (i.e., the mean number of button presses executed for each target in each run was set to 0), and then averaged across all targets, separately for pre- and postobservation runs.

Table 2. Partial correlations between self-reported confidence, neural confidence signals in the brain's reward system, neural observation–experience matching in the anterior insula, and interpersonal attraction after emotion observation

	Postobservation attraction											
	Partial correlations						Residual partial correlations (general target effects removed)					
	Approach behavior		Willingness to meet		Expected intimacy of communication		Approach behavior		Willingness to meet		Expected intimacy of communication	
Measure	*r*	*T*	*r*	*T*	*r*	*T*	*r*	*T*	*r*	*T*	*r*	*T*
Experiment I ($n = 40$)												
Self-reported confidence	0.45*	(4.5)	0.58*	(5.4)	0.61*	(7.5)	0.03	(0.3)	0.23*	(1.7)	0.24*	(1.7)
Experiment II ($n = 52$)												
Self-reported confidence	0.44*	(4.5)	0.45*	(5.2)	0.59*	(6.9)	0.17	(1.6)	0.19*	(1.7)	0.30*	(2.8)
Confidence signals in VS	0.20*,†	(2.2)	0.06	(0.1)	0.12	(1.3)	0.21*	(2.0)				
Confidence signals in mOFC	0.26*,†	(3.1)	−0.03	(−0.3)	0.01	(0.1)	0.22*,†	(2.8)				
NOE matching (cluster 1)	0.10	(1.2)	0.02	(0.2)	0.04	(0.4)						
NOE matching (cluster 2)	0.11	(1.2)	0.00	(0.0)	−0.02	(−0.2)						

Variance that can be explained by the observer's initial interpersonal attraction toward the targets is removed from both variables in all analyses. Residual correlations (i.e., correlations after general target effects are removed from both variables) are only reported for significant main correlations. *r*, back-transformed average partial correlation coefficients; *T*, *t* values at random-effects group level; VS, ventral striatum. Please see Fig. 3 for the location of the two clusters in the anterior insula.

*Significant correlations ($P < 0.05$, one-tailed).

†Correlations that remain significant in the split-half analysis ($P < 0.05$, one-tailed) (see *SI Materials and Methods* for details).

($T[5] = 4.2$, $P < 0.001$; Fig. 2; please see Table S1 for results for self-reported interpersonal attraction). Finally, the pattern of correlations between self-reported confidence and interpersonal attraction closely reflected that of experiment I, with partial correlations between self-reported confidence and self-reported interpersonal attraction remaining significant after general target effects had been removed from each individual dataset (Table 2).

Neural confidence signals in the brain's reward system. In the first step of the fMRI data analysis, we examined whether the observers' subjective confidence that they correctly understood the target's affective state was associated with neural activity in reward-related brain regions. For this, we used whole-brain univariate correlation analyses. First, we computed the correlation between each observer's trial-by-trial confidence ratings and trial-by-trial neural activity during the facial observation period. This revealed a significant positive correlation between self-reported confidence and neural activity in the right ventral striatum ($x = 18$, $y = 6$, $z = -15$; $T[51] = 5.7$, $P = 0.022$, familywise error [FWE]-corrected at voxel level; Fig. 3 *A* and *B*). Second, we computed the correlation between each observer's trial-by-trial confidence ratings and trial-by-trial neural activity during the emotion judgment period. This revealed a significant positive correlation between self-reported confidence and neural activity in the mOFC ($x = -3$, $y = 39$, $z = -18$; $T[51] = 4.9$, $k = 503$, $P < 0.001$, FWE-corrected at cluster level; Fig. 3 *E* and *F*) (please see Table S3 for brain regions outside the reward system where neural activity increased significantly with increasing subjective confidence).

ROI analysis: Neural confidence signals in the brain's reward system and individual changes in interpersonal attraction. Having shown that the observer's self-reported confidence reflected neural activity in the brain's reward system, we next asked whether these confidence-related neural signals would predict changes in interpersonal attraction. For this, we averaged the neural activity within the two clusters in the ventral striatum and mOFC, separately for each observer and target, and performed a partial correlation analysis between this neural activity and the observer's postobservation attraction scores for each target (with variance explained by preobservation attraction removed from both variables). This revealed significant positive partial correlations between confidence-related neural activity and the observer's approach behavior in both clusters (Fig. 3 *C* and *G* and Table 2). Again, these partial correlations remained significant when general target effects (i.e., average levels of confidence-related neural activity and average attraction scores for each target across all observers) were removed from each individual dataset (Fig. 3 *D* and *H* and Table 2). The only brain region outside the reward system that showed a similar pattern of partial correlations was the lingual gyrus (Table S4).

To ensure that the observed partial correlations between confidence-related neural activity and postobservation attraction were not due to the fact that we used the observer's confidence ratings (which we already knew predicted changes in interpersonal attraction) to identify clusters in the reward system that showed confidence-related neural activity, we performed a split-half cross-validation analysis (42, 43) (see *SI Materials and Methods* for details). This analysis replicated the significant partial correlations between confidence-related neural activity and postobservation attraction in the ventral striatum and mOFC (ventral striatum, mean back-transformed $r = 0.17$, $T[51] = 1.9$, $P = 0.033$; mOFC, mean back-transformed $r = 0.23$, $T[50] = 2.9$, $P = 0.003$) and the significant partial correlation between confidence-related neural activity and postobservation attraction after general target effects had been removed in the mOFC (mean back-transformed $r = 0.11$, $T[50] = 1.7$, $P = 0.047$). This indicates that the observed correlations between neural confidence signals in the reward system and interpersonal attraction cannot be explained by nonindependencies.

NOE matching. In the next step of our fMRI data analysis, we asked whether the observer's self-reported confidence and neural confidence signals are linked to the degree to which the patterns of neural activity elicited during emotion observation matched those associated with the observer's own emotional experience. For this, we used searchlight-based MVPA (44, 45), a technique that allows estimation of the level of correspondence between local patterns of neural activity within spherical neighborhoods (the "searchlights") across the entire brain volume (we refer to this level of correspondence as neural observation–experience matching; please see *Materials and Methods* for details). NOE-matching maps were computed for each trial and observer. These maps were then subjected to whole-brain correlation analyses with (*i*) the observer's trial-by-trial confidence ratings

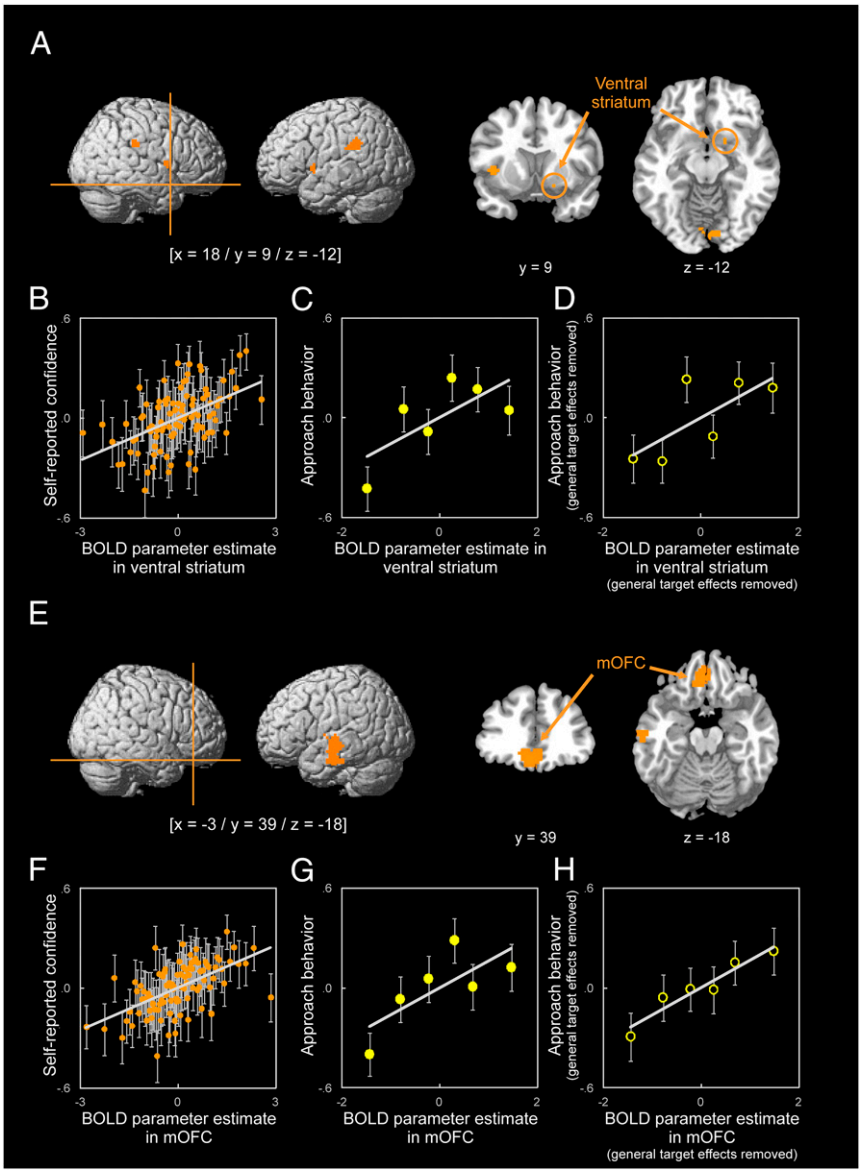

Fig. 3. Confidence-related neural activity in the brain's reward system and individual changes in interpersonal attraction. (*A*) Brain regions where neural activity during the facial observation period covaried with self-reported confidence. (*B*) Scatter plot illustrating the correlation between neural activity in the right ventral striatum and self-reported confidence. (*C*) Scatter plot illustrating the partial correlation between confidence-related neural activity in the right ventral striatum and the observer's postobservation approach behavior (variance that can be explained by preobservation approach behavior is removed). (*D*) Scatter plot illustrating the partial correlation between confidence-related neural activity in the right ventral striatum and the observer's postobservation approach behavior (variance that can be explained by preobservation approach behavior and general target effects are removed). (*E*) Brain regions where neural activity during the emotion judgment period covaried with self-reported confidence. (*F*) Scatter plot illustrating the correlation between neural activity in the mOFC and self-reported confidence. (*G*) Scatter plot illustrating the partial correlation between confidence-related neural activity in the mOFC and the observer's postobservation approach behavior (variance that can be explained by preobservation approach behavior is removed). (*H*) Scatter plot illustrating the partial correlation between confidence-related neural activity in the mOFC and the observer's postobservation approach behavior (variance that can be explained by preobservation approach behavior and general target effects are removed). Note: SPMs (height threshold $T[51] = 5.5$, $P = 0.05$, FWE-corrected at voxel level in *A*; height threshold $T[51] = 3.2$, extent threshold $k = 100$ voxels, $P = 0.001$, FWE-corrected at cluster level in *E*) are superimposed onto a rendered surface and coronal/axial sections of a T1-weighted map of a standard brain (MNI space). For the scatter plots in *B* and *F*, trialwise data of each observer (i.e., 96 data points) were *z*-standardized and rank-ordered according to BOLD parameter estimates and then averaged across observers, separately for each rank. For the scatter plots in *C*, *D*, *G*, and *H*, targetwise data of each observer (i.e., 6 data points) were *z*-standardized and rank-ordered according to BOLD parameter estimates and then averaged across observers, separately for each rank. Error bars represent SEMs.

and (*ii*) the observer's trial-by-trial confidence-related neural activity in the reward system (ventral striatum and mOFC, respectively). This revealed (*i*) a significant positive correlation between the observer's self-reported confidence and NOE matching in a cluster in the anterior insula ($x = -33$, $y = 21$, $z = -9$; $T[51] = 3.7$, $P = 0.001$, FWE-corrected at cluster level) and (*ii*) a significant positive correlation between neural confidence signals in the mOFC and NOE matching in a second, adjacent, cluster in the anterior insula ($x = -27$, $y = 30$, $z = -12$; $T[51] = 4.2$, $P = 0.001$, FWE-corrected at cluster level) (Fig. 4; please note that the actual overlap of the two clusters in the anterior insula is much larger than shown in Fig. 4 because each voxel represents a 9-mm spherical searchlight). No significant correlation was observed between confidence-related activity in the

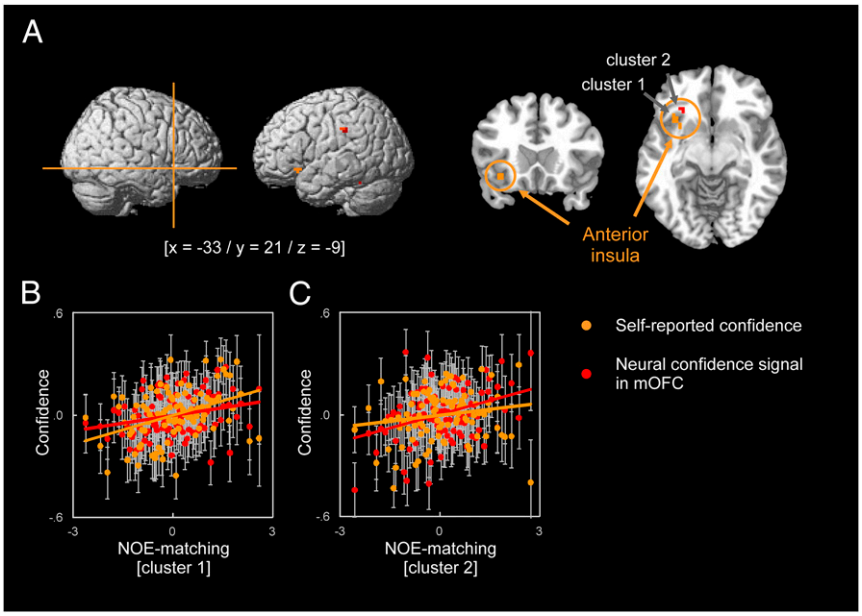

Fig. 4. Neural observation–experience matching (NOE matching), self-reported confidence, and neural confidence signals in the mOFC. (*A*) Brain regions where NOE matching covaried with self-reported confidence (orange, cluster 1) and neural confidence signals in the mOFC (red, cluster 2), respectively. (*B* and *C*) Scatter plots illustrating the correlation between NOE matching in each cluster and self-reported confidence (orange)/neural confidence signals in the mOFC (red). Note: SPMs (height threshold $T[51] = 3.2$, extent threshold $k = 10$ voxels, $P = 0.05$, FWE-corrected at cluster level) are superimposed onto a rendered surface and coronal/axial sections of a T1-weighted map of a standard brain (MNI space). For the scatter plots in *B* and *C*, trialwise data of each observer (i.e., 96 data points) were *z*-standardized and rank-ordered according to NOE matching and then averaged across observers, separately for each rank. Error bars represent SEMs.

ventral striatum and NOE matching (Table S5). As above, a split-half cross-validation analysis, performed to ensure that the correlation between NOE matching and confidence-related neural activity in the mOFC was not due to nonindependencies, replicated this effect ($x = -27$, $y = 30$, $z = -12$; $T[51] = 3.5$, $P = 0.040$, FWE-corrected at cluster level).

ROI analysis: NOE matching and individual changes in interpersonal attraction. Finally, we tested whether NOE matching in the anterior insula also predicted changes in interpersonal attraction directly. Interestingly, this was not the case (Table 2). This is in line with our hypothesis that NOE matching is not directly associated with changes in interpersonal attraction but that these changes are mediated by neural confidence signals in the brain's reward system.

Discussion

The goal of this study was to examine whether the human brain possesses a neural mechanism that attracts individuals to other individuals whose nonverbal signals they can easily understand. To pursue this goal, we conducted two experiments. In line with our first prediction, data of experiment I and behavioral data of experiment II show that an individual's interpersonal attraction toward another person can change after a few minutes of emotion observation, depending on the individual's subjective confidence that they correctly understood the other's affective state. fMRI data from experiment II provide an initial understanding of the neural processes that mediate between subjective understanding and interpersonal attraction. First, we found that individual changes in interpersonal attraction were predicted by confidence-related neural signals in the ventral striatum and the mOFC, core regions of the brain's reward system (12, 13). Second, we found that both the observer's subjective confidence and neural confidence signals in the mOFC covaried with the degree of similarity between patterns of neural activity elicited during emotion observation and those associated with the observer's own emotional experience (NOE matching). This suggests that an individual's confidence in their interpersonal judgments of affect, and ensuing changes in interpersonal attraction, is partly determined by how well the other person's affective behavior matches the observer's neural representation of the underlying state.

Confidence Signals in the Brain's Reward System. The first important finding of the current study is that confidence-related neural activity in the brain's reward system can act as an intrinsic reward signal, predicting individual changes in interpersonal attraction. Previous studies have shown that neural activity in the ventral striatum signals internal confidence when subjects make perceptual judgments about physical stimuli such as circles, lines, and moving dots (23, 24). The current study shows that neural activity in the ventral striatum/mOFC covaries with subjective confidence when participants try to infer another person's current affective state from their facial expression. This underlines a modality-independent role of the ventral striatum/mOFC in signaling confidence. Furthermore, confidence-related neural activity in the ventral striatum/mOFC predicted changes in the observer's interpersonal attraction toward the target, providing behavioral evidence that confidence-related activity in the brain's reward system can act as a positive reinforcement signal (24).

Three details of these findings are worth further discussion. First, we observed a temporal dissociation of neural confidence signals in the ventral striatum and in the mOFC: Confidence-related neural activity in the ventral striatum occurred during the facial observation period of each trial, whereas confidence-related neural activity in the mOFC occurred later, during the emotion judgment period. This is in line with previous work that has shown a similar temporal dissociation of neural activity in the ventral striatum/mOFC during face evaluation, which has led to the suggestion that the mOFC has a particular role in holding the outcome of stimulus evaluations online for further processing (9).

Second, neural confidence signals in both the ventral striatum and the mOFC were more closely associated with subsequent changes in the observer's approach behavior toward the target

than with changes in self-reported interpersonal attraction, whereas the observer's subjective confidence that they correctly understood the target's affective state was more closely associated with changes in self-reported interpersonal attraction. This is in line with previous studies that have stressed the role of the ventral striatum in motivated behavior (6) and suggests a partial dissociation between neural processes underlying motivational–behavioral and cognitive components of interpersonal attraction.

Third, confidence-related activity in the mOFC, but not in the ventral striatum, reflected the degree of similarity between patterns of neural activity elicited during emotion observation and those associated with the observer's own emotional states (NOE matching). Again, this supports a particular role of the mOFC in holding the outcome of stimulus evaluations online for further processing (9).

"Common Coding" and Success of Affective Communication. The second important finding of the current study is that both the observer's subjective confidence and neural confidence signals in the mOFC reflected the level of correspondence between patterns of neural activity elicited in the anterior insula during emotion observation and those associated with the observer's own emotional experience (NOE matching). This extends previous studies that observed overlapping activity in the anterior insula when participants experienced and observed pain (46, 47), disgust (30), or joy (32) and more recent studies that used MVPA to examine whether one's own pain and emotional experience and another person's pain and emotional experience are encoded in similar patterns of neural activity (37, 38).

Importantly, the results of the current study not only provide evidence that confidence can signal correspondence, they also indicate that confidence and correspondence covary across individual observer–target pairs. This provides empirical evidence for theoretical models of social interaction and communication that propose that the more similar the observer's and the target's internal model of a given behavior, the easier it should be for the observer to understand the target's inner state and to react accordingly (25, 48).

A similar link between correspondence of neural activity and success of communication has been observed in the medial prefrontal cortex (mPFC). A study using pseudohyperscanning (a technique where a "sender" and a "perceiver" are scanned one after the other in the same scanner but are connected by audio or video recordings such that their brain activity can be temporally aligned after scanning) showed that neural activity in the mPFC is time-locked between speakers and listeners involved in verbal communication. Strikingly, the listener's semantic understanding of the story told by the speaker varied as a function of the degree to which neural activity in the mPFC of the listener's brain at time point t_1 predicted the speaker's neural activity in that region at time point t_2 (49). However, in that study, all stories were told by a single speaker, so it remains unclear whether there was a specific listener-by-speaker interaction, similar to the observer-by-target interaction observed in the current study.

Success of Communication and Interpersonal Attraction. Until recently, neuroscientific research into interpersonal attraction has been guided by the view that an individual's primary goal when evaluating other individuals must be to identify potential mating partners who possess high genetic fitness and fertility (e.g., 11, 50). This research builds on a large literature that links physical attractiveness to genetic fitness (for a review, see, e.g., ref. 51). However, for species that live in complex environments that require social interaction and cooperation to maximize reproductive success, being able to identify the right cooperation partners might be equally important. The current study provides evidence that potential cooperation partners qualify as right not only by their willingness and competence to cooperate (14–17) (for a theoretical account see, e.g., ref. 52) but also by the degree to which their communication signals can be reliably decoded by the other individual. Importantly, unlike interpersonal attraction due to a target's fitness-signaling physical features, which seems to be fairly consistent across perceivers (53, 54), the confidence-dependent adjustment of interpersonal attraction found in the current study seems to be specific for specific interaction partners. Indeed, observers showed more disagreement about which target they felt attracted to after than before emotion observation. In social psychology it has long been recognized (21) that attraction between individuals is not only determined by general target effects (e.g., a target's physical attractiveness) but also by specific perceiver-by-target effects (relational effects) such as the match between a target's affective behavior and a perceiver's neural vocabulary we describe here. Interestingly, it has been suggested that such interaction partner-specific effects could underlie the forming of social cliques within larger groups (1, 55). The current study provides evidence that the brain's reward system, signaling how well one's neural vocabulary is suited to decode another person's behavior, might play an important role in these social processes.

Conclusion. In sum, we have shown that subjective understanding during social interaction can modulate interpersonal attraction. Interestingly, the findings of the current study suggest that the neural mechanisms underlying individual adjustments of interpersonal attraction during social encounters might act through internal reward signals that are partly independent of external feedback, which makes them perhaps less prone to cheating by potential cooperation partners. To investigate the interaction between intrinsic confidence signals and other—honest or manipulative—signals sent back and forth between communication partners and to examine the neural determinants of the dynamics of human social relations in larger groups ("social connectomes") remain challenging tasks for future studies. The current study suggests that mutual understanding is an important factor in interpersonal attraction, and that further research into the role of a common neural vocabulary in interpersonal attraction will lead to a better understanding of the neurobiological factors that define human social relations.

Materials and Methods

Participants. Forty volunteers (21 women, 19 men, all Caucasian, mean age 22.3 y, range 18–30 y) completed experiment I, and 54 volunteers completed experiment II. In experiment II, data of two participants were discarded because of estimated head movements >3 mm within a functional imaging run. The final sample in experiment II comprised 52 participants (28 women, 24 men, all right-handed and Caucasian, mean age 25.3 y, range 18–35 y). Participants reported no history of neurological or psychiatric disorders and had normal or corrected-to-normal vision. All participants gave written consent before participation and both studies were approved by the local ethics committee (Universität zu Lübeck).

Stimuli. Videos of women experiencing fear or sadness were recorded in a previous fMRI study in which participants were asked to imagine and submerge themselves into a cued emotional situation and to facially express their feeling to their romantic partner (36). Using prerecorded videos of women who experienced and facially expressed fear and sadness toward their romantic partner ensured that all participants saw exactly the same behavior, and allowed us to exclude the possibility that individual changes in the participants' interpersonal attraction toward the targets were due to differences in the target's behavior toward different participants. Women were chosen as targets because women have been shown to express their emotions more accurately than men (56, 57). For the current study, videos of fear and sadness of six different women (all Caucasian, age 20–25 y) were selected. Videos were cut into short clips, each covering the first 8 s of a 20-s emotional period. The final set consisted of 48 different video clips (4 videos for each target and emotion). Each of the 48 emotion video clips was shown twice, resulting in a total of 96 emotion observation trials per participant. For preexperiment familiarization with each target and the assessment of

interpersonal attraction (see below), a still picture of each target was cut from the original recordings, showing the target's face during a 20-s rest period.

Cover Story and Preexperiment Familiarization. Experimental procedures were similar in both studies (Fig. 1). Upon arrival in the laboratory, participants were told that the aim of the study was to investigate the relation between response times and the neural processing of faces and emotional expressions. Participants were then seated in front of a computer screen using head phones and a chin rest to avoid distraction and to ensure that they viewed all facial stimuli at the same distance. To support the cover story and to familiarize participants with the targets, assessment of interpersonal attraction was preceded by a motor task in which participants were required to press one of two response buttons in response to a visual cue (an arrow pointing to the left or to the right, or a negative or positive word) as quickly as possible. Response time trials were intermixed with a total of 96 short presentations (200 ms) of still pictures of each target (16 presentations per target) and targets were fully balanced over arrows and words, so that after completion of the motor task, participants were well-familiarized with each target. Participants in experiment I completed all parts of the study on the same computer screen, and participants in experiment II completed all parts except emotion observation and emotion experience (which were performed during fMRI) on the same computer screen.

Assessment of Interpersonal Attraction. First, to assess the participants' interpersonal attraction toward each target at the motivational–behavioral level, participants were asked to imagine that they would approach targets, one after the other, for a casual conversation. At the beginning of each trial, a small picture of the target appeared on the computer screen (about 40% of the original picture size) and participants were asked to increase the size of the picture by repeatedly pressing a button (increase about 4% per button press, no decrease button) until a pleasant conversational distance was reached. This task is a modified version of a task originally introduced by Aharon and colleagues (6) to measure the reward value of faces, except that the task used in the current study additionally mimics approach behavior by increasing the size of the target with each button press.

Second, to assess the participants' self-reported attraction toward each target, they were shown a still picture of each target (1 s) followed by three statements about their interpersonal attraction toward the target. The statements were adapted from the "social attraction" items in McCroskey and McCain (58) and related to the participant's subjective motivation to meet the target in real life (willingness to meet) and the participant's expectation that they could have an intimate communication with the target (expectation of intimate communication) (Table 1). Participants were asked to indicate how much they agreed with each statement on a Likert-type 7-point visual scale ranging from 1 (not at all) to 7 (definitely) by pressing the corresponding key on a keyboard. In both tasks, targets were presented in different random orders before and after emotion observation.

Emotion Observation. In experiment I, emotion observation was divided into four runs. During each run, 24 video clips, balanced across the six targets and two emotions, were presented in randomized order. Each video clip was followed by an emotion judgment question ["Hat sie Furcht oder Trauer empfunden?" ("Did she feel fearful or sad?")]. After the participant had entered their emotion judgment by pressing the corresponding button on the keyboard, a confidence question ["Wie sicher bist Du, dass sie Furcht/Trauer empfunden hat?" ("How confident are you that she felt fearful/sad?")] and a 5-point visual scale ranging from 1 (I am guessing) to 5 (I am absolutely sure) appeared on the screen. Responses were entered by pressing the corresponding number on the keyboard. The orientation of the scale (increasing confidence values from left to right or right to left) was balanced across participants. Each trial terminated with an intertrial interval of 1 s, during which a fixation cross was shown.

In experiment II, emotion observation was divided into eight runs. During each run, 12 video clips, balanced across the six targets and two emotions, were presented in randomized order. Each video clip was followed by a fixation cross (1 s), an emotion judgment screen (2 s), and a confidence screen (3 s). The emotion judgment screen showed the words "Trauer" (sadness) and "Furcht" (fear) side by side at the center of the screen, indicating that the participant should convey their emotion judgment by the response button in their left or right hand, respectively. The order of emotion words (left or right) was balanced across targets and emotions within participants. After the participant had entered their response, an orange frame appeared around the chosen emotion word, providing the participant with feedback about their response. The confidence screen showed a five-dot visual scale ranging from 1 (I am guessing) to 5 (I am absolutely sure). An orange dot at the central position of the scale indicated the starting position of the cursor. Participants were asked to move the orange dot to the left or to the right by pressing the button in their left or right hand, respectively, to indicate their confidence about their emotion judgment. The orientation of the scale (increasing confidence from left to right or right to left) was balanced across participants. Each trial terminated with an intertrial interval of 8 or 10 s, during which a fixation cross was shown (Fig. 1).

Stimulus presentation and response logging were controlled with Presentation software (Neurobehavioral Systems).

Emotion Experience. After completion of the emotion observation part, participants in experiment II participated in four additional fMRI runs during which they were asked to experience and express fear and sadness themselves. Participants were informed that the experimental setup during this part of the experiment would be very similar to that for the women they had just observed, except that their facial expression would not be recorded, and that they would be asked to submerge themselves into frightening or sad situations and to feel and express their feelings as soon as they saw the corresponding word [Furcht (fear) or Trauer (sadness)] on the screen (please see *SI Materials and Methods* for details). Each run (two runs per emotion) comprised four emotional periods (20 s) and five interspersed periods (20 s), during which participants were asked to relax. The order of emotions was balanced across participants.

MRI Data Acquisition. MRI data were acquired on a 3-T scanner (Siemens MAGNETOM Trio). A T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) image [MPRAGE, 176 sagittal slices, resolution $1 \times 1 \times 1$ mm^3, field of view (FOV) 256×256 mm^2, flip angle 8°, inversion time 1,100 ms], used for spatial normalization of individual data, and a T2-gradient echo image [39 axial slices per volume, slice thickness 3 mm + 1-mm gap, interleaved order, in-plane resolution 3×3 mm^2, FOV 192×192 mm^2, echo time (TE) 1 5.19 ms, TE2 7.65 ms, repetition time (TR) 425 ms], used to compute individual field maps for correction of image distortions, were obtained from each participant before functional imaging. One hundred forty-five T2*weighted echoplanar images (EPIs) covering the whole brain were acquired during each emotion observation run, and 96 EPIs were acquired during each emotion experience run (35 axial slices per volume, slice thickness 4 mm + 0.4-mm gap, interleaved order, in-plane resolution 3×3 mm^2, FOV 192×192 mm^2, TE 30 ms, TR 2,000 ms, generalized autocalibrating partially parallel acquisition, factor 2). Functional runs were preceded by five functional images not included in the analysis to allow for T1 saturation.

Data Analysis. MRI data were preprocessed with SPM8 (Wellcome Department of Imaging Neuroscience, University College London; www.fil.ion.ucl.ac.uk/spm/software/spm8/). Preprocessing followed standard procedures and included concurrent spatial realignment and correction of image distortions and normalization into standard Montreal Neurological Institute (MNI) space (59) at a spatial resolution of $3 \times 3 \times 3$ mm^3 using DARTEL (60). An additional receiver coil sensitivity bias correction [using the New Segment tool of SPM8 with very light regularization (0.0001) and 60-mm smoothness] was conducted after realignment and unwarping that corrected for differences in the scanner's bias correction between functional runs that occurred due to a technical problem.

For the analysis of confidence-related neural activity, individual maps of parameter estimates were computed for each participant based on a standard generalized linear model (GLM) that accounted for first-order autocorrelations and low-frequency drifts (high-pass cutoff period 128 s). BOLD (blood oxygen level-dependent) activity was modeled separately for each emotion observation trial ($n = 96$) using box car functions (three per trial), convolved with a standard hemodynamic response function (hrf) that modeled (*i*) video onset and duration (8 s), (*ii*) emotion judgment onset and duration (3 s), and (*iii*) confidence rating onset and duration (3 s) (the latter were included as regressors of no interest). For the analysis of neural observation–experience matching (see below), a second set of maps of parameter estimates ($n = 16$, 20-s box car functions convolved with the hrf) was obtained for the emotion experience runs of each participant.

For the whole-brain analysis of confidence-related neural activity, trial-by-trial correlation maps (BOLD parameter estimates–self-reported confidence) were computed for each participant, Fisher-transformed, spatially smoothed (8-mm isotropic Gaussian kernel), and tested at random-effects group level (using *T* statistics).

For the analysis of NOE matching, we used a linear support vector machine (SVM) as implemented in LIBSVM (https://www.csie.ntu.edu.tw/~cjlin/libsvm) with a linear kernel and a hard margin. The searchlight radius was set to

9 mm (123 voxels), and the searchlight was moved in steps of one voxel through the entire brain volume. The classifier was trained on patterns of neural activity associated with the participant's own emotional experience (n = 8 samples per class) and tested on patterns of neural activity elicited during video observation (n = 96). To ensure that classification was based on multivoxel patterns of neural activity and not on the average level of activity within a sphere, the spatial mean of each local pattern was set to zero. Because we reasoned that the classifier's decision confidence (the distance between test sample and decision border) would provide a more accurate estimate of the level of correspondence between a test pattern and the reference patterns than the classifier's decision accuracy alone (which is binary variable), we computed a measure that reflected both the classifier's decision accuracy and the classifier's confidence for this decision, the weighted decision confidence. Mathematically, the weighted decision confidence is the product of the classifier's decision accuracy [which was set to {1} for correct decisions (classifier's decision matched the participant's judgment) and to {−1} for incorrect decisions (classifier's decision did not match the participant's judgment)] and the classifier's decision confidence (which is defined for a linear SVM as the distance between the test sample and the hyperplane that separates the two classes).

For the whole-brain analyses of neural observation–experience matching, three trial-by-trial correlation maps (NOE-matching–self-reported confidence, NOE-matching–neural confidence signals in the ventral striatum, NOE-matching–neural confidence signals in the mOFC) were computed for each participant. As above, these maps were Fisher-transformed, spatially smoothed (4-mm isotropic Gaussian kernel; the small kernel size accounted for the fact that a searchlight-based SVM already introduces some smoothness into the data), and tested at random-effects group level (using T statistics) separately for each comparison.

Statistical significance of all random-effects statistical parametric maps (SPMs) was assessed allowing for a probability of false positives of P = 0.05, corrected for multiple tests (familywise errors) across the whole volume according to random-field theory (61). FWE correction was performed at voxel level for the ventral striatum and at cluster level (using a height threshold of $T[51]$ = 3.2 corresponding to P = 0.001) for all other regions. This accounted for the fact that clusters of activity were expected to be more distributed in cortical regions than in the ventral striatum.

For the ROI analyses (i.e., neural confidence signals in the ventral striatum–interpersonal attraction, neural confidence signals in the mOFC–interpersonal attraction, NOE matching in the anterior insula–interpersonal attraction), trialwise BOLD parameter estimates/trialwise weighted decision confidences were extracted from the corresponding cluster (identified in the whole-brain analyses) and averaged separately for each target and participant. For the cluster in the mOFC, BOLD parameter estimates were extracted from a 3-mm sphere centered at the peak voxel ([−3 39 −18]) because the mOFC cluster was a large cluster that extended into the dorsomedial OFC.

All ROI-based correlation analyses were checked for outliers, defined as values that deviated more than three times the interquartile range from the first or third quartile. No outliers were detected in the main analysis, and two outliers were detected in the split-half cross-validation (indicated by degrees of freedom less than n − 1). For all ROI-based analyses, a probability of false positives of P = 0.05 (one-tailed) was accepted unless indicated otherwise, and exact P values are reported for $P < 0.200$ and $P > 0.001$.

Observer-by-target interactions in self-reported confidence were computed with SPSS (version 22.0.0.1; IBM).

ACKNOWLEDGMENTS. The authors thank N. Dewies, E. Charyasz, and M. Erb for help with data acquisition, and N. Weiskopf for technical expertise. This work was funded by Bundesministerium für Bildung und Forschung (German Federal Ministry of Education and Research) Grant 01GQ1105 (to S.A.).

1. Newcomb TM (1956) The prediction of interpersonal attraction. *Am Psychol* 11(11): 575–586.
2. Byrne D (1997) An overview (and underview) of research and theory within the attraction paradigm. *J Soc Pers Relat* 14(3):417–431.
3. Byrne D, Rhamey R (1965) Magnitude of positive and negative reinforcements as a determinant of attraction. *J Pers Soc Psychol* 2(6):884–889.
4. Gouaux C, Summers K (1973) Interpersonal attraction as a function of affective state and affective change. *J Res Pers* 7(3):254–260.
5. Nakamura K, et al. (1999) Activation of the right inferior frontal cortex during assessment of facial emotion. *J Neurophysiol* 82(3):1610–1614.
6. Aharon I, et al. (2001) Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* 32(3):537–551.
7. O'Doherty J, et al. (2003) Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41(2):147–155.
8. Kampe KKW, Frith CD, Dolan RJ, Frith U (2001) Reward value of attractiveness and gaze. *Nature* 413(6856):589.
9. 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(46): 18253–18258.
10. Cloutier J, Heatherton TF, Whalen PJ, Kelley WM (2008) Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *J Cogn Neurosci* 20(6):941–951.
11. Bzdok D, et al. (2011) ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Struct Funct* 215(3-4):209–223.
12. Haber SN, Knutson B (2010) The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology* 35(1):4–26.
13. Bartra O, McGuire JT, Kable JW (2013) The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76:412–427.
14. Rilling J, et al. (2002) A neural basis for social cooperation. *Neuron* 35(2):395–405.
15. Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport* 15(16):2539–2543.
16. Tabibnia G, Satpute AB, Lieberman MD (2008) The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol Sci* 19(4):339–347.
17. Phan KL, Sripada CS, Angstadt M, McCabe K (2010) Reputation for reciprocity engages the brain reward center. *Proc Natl Acad Sci USA* 107(29):13099–13104.
18. Jones RM, et al. (2011) Behavioral and neural properties of social reinforcement learning. *J Neurosci* 31(37):13039–13045.
19. Korn CW, Prehn K, Park SQ, Walter H, Heekeren HR (2012) Positively biased processing of self-relevant social feedback. *J Neurosci* 32(47):16832–16844.
20. Bhanji JP, Delgado MR (2014) The social brain and reward: Social information processing in the human striatum. *Wiley Interdiscip Rev Cogn Sci* 5(1):61–73.
21. Kenny DA, West TV, Malloy TE, Albright L (2006) Componential analysis of interpersonal perception data. *Pers Soc Psychol Rev* 10(4):282–294.
22. Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455(7210):227–231.
23. Hebart MN, Schriever Y, Donner TH, Haynes J-D (2016) The relationship between perceptual decision variables and confidence in the human brain. *Cereb Cortex* 26(1): 118–130.
24. Daniel R, Pollmann S (2012) Striatal activations signal prediction errors on confidence in the absence of external feedback. *Neuroimage* 59(4):3457–3467.
25. Gallese V (2003) The manifold nature of interpersonal relations: The quest for a common mechanism. *Philos Trans R Soc Lond B Biol Sci* 358(1431):517–528.
26. Decety J, Jackson PL (2004) The functional architecture of human empathy. *Behav Cogn Neurosci Rev* 3(2):71–100.
27. Bastiaansen JACJ, Thioux M, Keysers C (2009) Evidence for mirror systems in emotions. *Philos Trans R Soc Lond B Biol Sci* 364(1528):2391–2404.
28. Iacoboni M (2009) Imitation, empathy, and mirror neurons. *Annu Rev Psychol* 60: 653–670.
29. Adolphs R, Damasio H, Tranel D, Cooper G, Damasio AR (2000) A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J Neurosci* 20(7):2683–2690.
30. Carr L, Iacoboni M, Dubeau M-C, Mazziotta JC, Lenzi GL (2003) Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proc Natl Acad Sci USA* 100(9):5497–5502.
31. Wicker B, et al. (2003) Both of us disgusted in *My* insula: The common neural basis of seeing and feeling disgust. *Neuron* 40(3):655–664.
32. Pourtois G, et al. (2004) Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *Eur J Neurosci* 20(12):3507–3515.
33. Hennenlotter A, et al. (2005) A common neural basis for receptive and expressive communication of pleasant facial affect. *Neuroimage* 26(2):581–591.
34. Mitchell JP, Banaji MR, Macrae CN (2005) The link between social cognition and self-referential thought in the medial prefrontal cortex. *J Cogn Neurosci* 17(8):1306–1315.
35. Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50(4):655–663.
36. Buchanan TW, Bibas D, Adolphs R (2010) Associations between feeling and judging the emotions of happiness and fear: Findings from a large-scale field experiment. *PLoS One* 5(5):e10640.
37. Anders S, Heinzle J, Weiskopf N, Ethofer T, Haynes J-D (2011) Flow of affective information between communicating brains. *Neuroimage* 54(1):439–446.
38. Corradi-Dell'Acqua C, Hofstetter C, Vuilleumier P (2011) Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *J Neurosci* 31(49):17996–18006.
39. Lorey B, et al. (2012) Confidence in emotion perception in point-light displays varies with the ability to perceive own emotions. *PLoS One* 7(8):e42169.
40. Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293(5539):2425–2430.
41. Moreland RL, Zajonc RB (1982) Exposure effects in person perception: Familiarity, similarity, and attraction. *J Exp Soc Psychol* 18(5):395–415.
42. Vul E, Harris C, Winkielman P, Pashler H (2009) Reply to comments on "Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition." *Perspect Psychol Sci* 4(3):319–324.
43. Poldrack RA, Mumford JA (2009) Independence in ROI analysis: Where is the voodoo? *Soc Cogn Affect Neurosci* 4(2):208–213.

44. Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7(7):523–534.
45. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103(10):3863–3868.
46. Singer T, et al. (2004) Empathy for pain involves the affective but not sensory components of pain. *Science* 303(5661):1157–1162.
47. Ochsner KN, et al. (2008) Your pain or mine? Common and distinct neural systems supporting the perception of pain in self and other. *Soc Cogn Affect Neurosci* 3(2): 144–160.
48. Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci* 358(1431): 593–602.
49. Stephens GJ, Silbert LJ, Hasson U (2010) Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci USA* 107(32):14425–14430.
50. Funayama R, et al. (2012) Neural bases of human mate choice: Multiple value dimensions, sex difference, and self-assessment system. *Soc Neurosci* 7(1):59–73.
51. Senior C (2003) Beauty in the brain of the beholder. *Neuron* 38(4):525–528.
52. Fiske ST (2012) Journey to the edges: Social structures and neural maps of inter-group processes. *Br J Soc Psychol* 51(1):1–12.
53. Said CP, Haxby JV, Todorov A (2011) Brain systems for assessing the affective value of faces. *Philos Trans R Soc Lond B Biol Sci* 366(1571):1660–1670.
54. Freeman JB, Stolier RM, Ingbretsen ZA, Hehman EA (2014) Amygdala responsivity to high-level social information from unseen faces. *J Neurosci* 34(32):10573–10581.
55. Hogan R, Mankin D (1970) Determinants of interpersonal attraction. A clarification. *Psychol Rep* 26(1):235–238.
56. Buck RW, Savin VJ, Miller RE, Caul WF (1972) Communication of affect through facial expressions in humans. *J Pers Soc Psychol* 23(3):362–371.
57. Sabatelli RM, Buck R, Dreyer A (1980) Communication via facial cues in intimate dyads. *Pers Soc Psychol Bull* 6(2):242–247.
58. McCroskey JC, McCain TA (1974) The measurement of interpersonal attraction. *Speech Monogr* 41(3):261–266.
59. Worsley KJ, et al. (1996) A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4(1):58–73.
60. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38(1):95–113.
61. 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(2):192–205.
62. Tzourio-Mazoyer N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1):273–289.