Supporting Online Material for

A Key Role for Similarity in Vicarious Reward

Dean Mobbs,* Rongjun Yu, Marcel Meyer, Luca Passamonti, Ben Seymour, Andrew J. Calder, Susanne Schweizer, Chris D. Frith, Tim Dalgleish

*To whom correspondence should be addressed. E-mail: dean.mobbs@mrc-cbu.cam.ac.uk

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 to S5
Table S1
References

Supporting Online Materials (SOM)

A Key Role for Similarity in Vicarious Reward

Dean Mobbs[*], Rongjun Yu, Marcel Meyer, Luca Passamonti, Ben Seymour, Andrew Calder, Susanne Schweizer, Chris D. Frith, and Tim Dalgleish

*To whom correspondence should be addressed:

Email: dean.mobbs@mrc-cbu.cam.ac.uk

Volunteers

We scanned 14 healthy volunteers (8 females; mean age and SD 25.8 ± 3.7). All participants were students of Cambridge University, right-handed, fluent speakers of English and screened for psychiatric or neurological problems. This study was authorized by the National Health Service Local Research Ethics Committee for Cambridge.

Videos

To create socially desirable (SD), socially undesirable (SU) and neutral (e.g. neither extremely SD or SU) contestants, several videos were filmed of actors answering a number of personal, social and ethical questions in front of the camera. The questions were similar to those found on contestant forms for a game show (i.e. Who Wants to be a Millionaire) and, as part of the manipulation, the answers were predetermined. To illustrate, in the answer to the question: "What would the happiest day of your life look like?" the SU contestant said "Win the lottery, be married to Angelina Jolie and have a nice car.", whereas the SD contestant replied "I think it would involve having all my friends and family around, all at the same time, and we are getting on. And also, something else, which I think would be nice: if there was some treaty or something that said 'none of the countries are going to argue anymore and everyone has to try and be nice to one another'". An example of the neutral contestant is "Not sure it would be a day but a night and would be running off into the night with a wonderful lady". Based on pilot testing (N = 8) we selected two videos with characters that were perceived as neutral (one male, one female), and one each of where a contestant was perceived as SD (female) and SU (male), respectively.

Trait Word Task

We devised a trait word task as a measure of the likeability of the contestants. For this purpose we selected 80 trait words (40 positive and 40 negative) from a database of trait words that were considered relevant to the characters (*S1*). In the trait word task a neutral picture of the relevant contestant was displayed in the centre of the screen (taken as a screenshot from the character's video). The trait word was presented just below the picture and a five-point rating scale (1 = *not at all* and 5 = *extremely*). We used E-prime stimulus presentation and response recordings for the trait word task as well as the experimental paradigm (described below). A standard keyboard was utilized to record participants' responses using the number buttons (1-5). In a pilot study (N = 8) participants rated the desirability of each of the 80 trait words on a 10-point scale (e.g. 1= *highly undesirable* to 10 = *highly desirable*). Paired samples t-tests indicated that the desirability between the two categories of trait words was statistically significantly different (positive words [mean = 7.2; SD = .49] versus negative words [mean = 2.2; SD = .56], t^{7} = 15.6, P < .05]). A paired t-test showed that participants attributed higher positive traits word ratings to the SD contestant (t^{13} = 7.3; P < .00005) and higher negative traits word ratings to the SU contestant (see Figure 1 A).

Experimental Paradigm

At the beginning of the experiment, participants were asked to complete several questionnaires and a contestant form. They were then filmed reciting their answers. The contestant form consisted of the same questions as those that were used for the videos of the contestants. For this form, as well as all other self-report measures collected in this study, it was emphasized that questions should be answered according to the participant's personal view and that there were no right or wrong answers. Participants were further informed that their video would be edited and then presented to three future participants of this study. To make this statement more credible, prior to the filming participants were asked to provide written a consent specifically to have their videos shown to other participants. After this, participants viewed three videos and in order to reduce suspicion regarding the socially desirability manipulation, we interleaved the confederates with several videos of other potential contestants encouraging the participants to believe that we had randomly chosen these contestants for their fMRI study. The order of presentation was counterbalanced, but with the restriction that the video of the neutral characters was always shown in second place. Participants were told that these videos were recordings from participants who had done the same study in the previous testing week and that one of the characters behaved slightly provocatively despite having received the same instructions as every other participant in this study. We felt that it was important to prepare participants for seeing the socially undesirable character in the latter manner, as he might otherwise not have been perceived as a genuine participant, thus evoking demand characteristics.
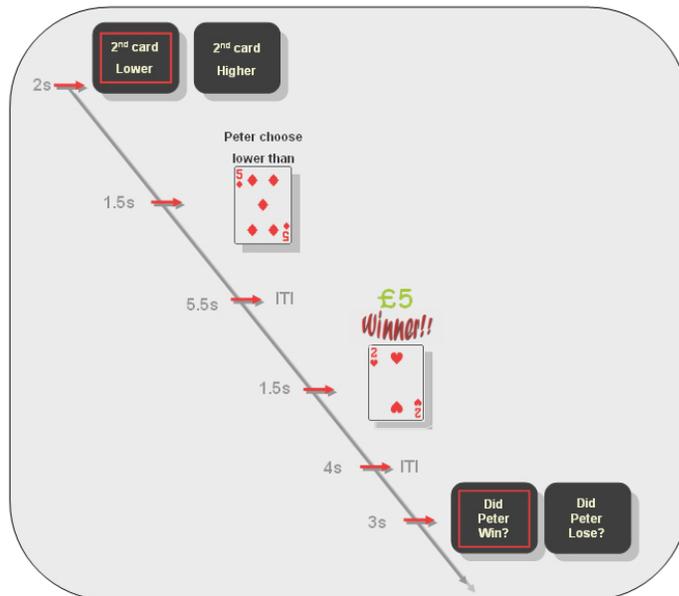
Figure S1. The experimental paradigm.

Before the scanning, participants were asked to practice the gambling task both for the self-play (i.e. where they played themselves) and the observer (i.e. where the SD or the SU character played) blocks. Participants played the practice tasks until the experimenter felt that they had fully mastered them. The practice tasks were identical to the experimental tasks, except for the characters who were playing.

Participants then observed two contestants (i.e. SD and SU) and one self-play block with 80 trials per block. At the beginning of each block participants were informed about who would be playing. Before the self-play block, participants watched the two contestants. The order of appearance for the observer blocks was counterbalanced. Participants were informed that the two videos had been randomly selected from the three videos that they had viewed previously and that the participant's own game would also be recorded for this purpose. Unbeknown to participants, the character selection and the rounds had been predetermined. Each block consisted of 80 trials and the order of the contestants was counterbalanced across subjects.

On each trial, two playing cards were randomly drawn from a deck of five cards numbered between 1 to 5. Before seeing either card, participants were first required to guess whether the number of the second card would be higher or lower than that of the first card by pressing the left or right button (with the first and second fingers of the right hand). After that, participants were presented with the first card for 1.5s, followed by the second card for 1.5s.

In each block, each number ("1", "2", "3", "4", "5") appeared on the first card for 16 trials. The probability of winning varied as a function of the number on card 1 (see, Table 1). For example, if the player's guess was "second card higher" and the number on card 1 was a "2", then the player would win with a 75 % probability. That is, three numbers on card 2 can be higher ("3", "4", "5"), but only one can be lower ("1") than the number on card 1 ("2"). For the observer blocks, these choices were predetermined so that both outcomes occurred with equal frequency. Participants were told that the contestants from the observer blocks had played for the same amount of money as they would play for in the self-play block. In the self-play block, participants could win £5 for each correct guess and lose £5 for each incorrect guess. At the end of the experiment, each participant was paid the same amount of £40 for their participation

Our task aimed to intensely engage participants in a realistic-like gambling game. To further increase participants' interest, a bell sound, a shuffling sound, a cash register, and a buzzer were played when the first, the second card, win outcome, and loss outcome were shown, respectively. To check participants' attention to the task, they were required to press either of the two buttons on the presentation of the first card and they had to press the corresponding yes/no button to confirm the outcome (e.g. Did Peter win?). The recording of yes/no buttons and the win/loss outcome questions were counterbalanced.

Any confounds of interpersonal attraction (e.g. anticipation of seeing an attractive contestant) were reduced by only showing the name of the contestant and not a picture. This also controlled for any effects of emotional contagion, for example, observing a joyful face. Strangers were used so that there was no direct consequence of seeing the contestants winning (e.g. if they win, will they give me a portion of the money). We aimed to reduce any influence of complex emotions such as jealously at seeing others win by making participants play the game after seeing the contestants play. No significant differences were found between how jealous the participants felt when observing the SD and SU contestants win ($t_{13} = .618$; $P = 0.547$). Any competition confounds were reduced by making participant play the game independent of the observed contestants and telling the participant that the contestant had played the game within the previous week.

Table 1. Probabilities of winning a trial dependent on decision and card 1.

Card 1	Higher*	Lower*
1	1.00	0.00
2	0.75	0.25
3	0.50	0.50
4	0.25	0.75
5	0.00	1.00

*Probabilities of winning after card 1 is shown if the decision is:

Questionnaire Measures

Priori to starting the experiment, participants were further asked to complete the Interpersonal Reactivity Index (IRI; S2, S3). We were interested in the sub-scales of empathy and perspective taking on the IRI. The empathy score is an index of "other-oriented" feelings of sympathy and concern for unfortunate others. The perspective taking scale assesses the tendency to adopt another's point of view. Following the MRI scan, participants answered a number of questions concerning the gambling task (post-scan questionnaire). The rating scale ranged from 1 and 2 (Very slightly or not at all) over 3 and 4 (A little), 5 and 6 (Moderately), 7 and 8 (Quite a bit) to 9 and 10 (Extremely). Examples of the questions include: "How rewarding was it when contestant 1 won?" and "How similar are you to contestant 2?"

Image acquisition
MRI scanning was conducted at the Medical Research Council Cognition and Brain Sciences Unit on a 3-Tesla Trio Tim Magnetic Resonance Imaging scanner (Siemens, Germany) by using a head coil gradient set. Whole-brain data were acquired with echoplanar T2*-weighted imaging (EPI), sensitive to BOLD signal contrast (48 sagittal slices, 3 mm-thickness; TR = 2400 ms; TE = 30 ms; flip angle = 78°; FOV 192 mm; voxel size: 3x3x3 mm). To provide for equilibration effects the first 5 volumes were discarded. T1 weighted structural images were acquired at a resolution of 1x1x1 mm.

Image preprocessing
SPM5 software (www.fil.ion.ucl.ac.uk/spm/) was used for data analysis. The EPI images were sinc interpolated in time for correction of slice timing differences and realignment to the first scan by rigid body transformations to correct for head movements. Field maps were estimated from the phase difference between the images acquired at the short and long TE and unwrapped, employing the FieldMap toolbox. Field map and EPI imaging parameters were used to establish voxel displacements in the EPI image. Application of the inverse displacement to the EPI images served the correction of distortions. For each participant the mean EPI was calculated and examined to guarantee that none exhibited excessive signal dropout in insula and ventral striatum. Utilising linear and non-linear transformations, and smoothing with a Gaussian kernel of full-width-half-maximum (FWHM) 8-mm, EPI and structural images were coregistered and normalized to the T1 standard template in MNI space (Montreal Neurological Institute (MNI) – International Consortium for Brain Mapping). Moreover, global changes were removed by proportional scaling and high-pass temporal filtering with a cut-off of 128 s was used to remove low-frequency drifts in signal.

Statistical analysis
After preprocessing statistical analysis was performed using the general linear model. Analysis was carried out to establish each participant's voxel-wise activation during the observe contestants and self-play blocks. Activated voxels in each experimental context were identified using an event-related statistical model representing each of the four experimental contexts, convolved with a canonical hemodynamic response function and mean-corrected. Six head-motion parameters defined by the realignment were added to the model as regressors of no interest. Multiple linear regression was then run to generate parameter estimates for each regressor at every voxel. For group statistics random effects analysis was utilized. A statistical threshold of $P < 0.05$ corrected for multiple spatial comparisons across the whole-brain was used, except for a priori hypothesized regions which were thresholded at $P = 0.001$ uncorrected (only clusters involving $k > 10$ or more contiguous voxels are reported). Small volume correction was used on several a priori regions of interest including the amygdala, medial orbital frontal cortex, ventral striatum, insula and dorsal and ventral medial PFC including the ventral ACC.

Connectivity analyses
Psycho-Physiological Interaction (PPIs)
We sought to identify 'target' regions which had differential connectivity with the ventral striatum ('source region') as a function of viewing a SU contestant winning compared to a SD one. This was achieved using a moderator variable, derived from the product of the source activation and the psychological context. For each participant, we computed the above contrasts to determine the local maximum that was the nearest voxel to the activation peak in the ventral striatum defined by the whole group cluster. Analysis employed an 8-mm sphere across all participants for ventral striatum: seed location: x = 18, y = 14, z = -8, which was the maximal voxel. Next, the time-series for each participant was computed by using the first eigenvariate from all voxels' time series in this common ventral striatum ROI. The BOLD time series for each participant was deconvolved to estimate a 'neuronal time series' for this region (S4). The psycho-physiological interaction term (PPI regressor) was calculated as the element-by-element product of the ventral striatum neuronal time

series and a vector coding for: [1 = SD win vs -1 = SU win]. This product was reconvolved by the canonical hemodynamic response function (hrf). The model also included the main effects of task convolved by the hrf, and the movement regressors as effects of no interest. Participant specific PPI models were run, and contrast images generated for positive and negative PPIs. The identified regions have greater or lesser connectivity with the source region according the context of the contrast above.

Supplementary results
Subjects were asked to respond when they saw the first card appear, and respond after the second card by pressing either the left or right button to indicate if the contestant won or lost. Responses were limited to the first card and no significant differences were found for the amount of time subjects pressed to indicate the presence of the first card ($t_{13} = 1.076$; $P = 0.3$).

Self-ratings
Following the scanning session, we asked participants to use a 10-point Likert scale to indicate their subjective responses in relation to the task. Compared to the SU contestant, participants rated themselves to be significantly more similar to the SD contestant ($t_{13} = -4.41$; $P = 0.001$) and agreed more with the SD contestant ($t_{13} = -6.996$; $P < 0.0005$). As mentioned above, no significant differences were found between SD and SU contestants in terms of how jealous the participants reported feeling when seeing them win ($t_{13} = .618$; $P = 0.547$).

Test for sex differences
Given that we used one female and one male as the confederate contestants, we tested whether attitudinal similarity was influenced by contestant's sex. We found no sex differences between attitudinal similarity to the SD ($t_{12} = .207$; $P = 0.840$) and SU contestant ($t_{12} = -1.112$; $P = 0.288$). Moreover, no significant sex differences were found for how rewarding participants found it to see the SD ($t_{12} = .-628$; $P = 0.543$) and SU contestant win ($t_{12} = 0.00$; $P = 0.100$). Lastly, we found no sex differences in the propensity to select higher positive traits for the SD ($t_{12} = -.843$; $P = 0.416$) and SU contestant ($t_{12} = 1.803$; $P = 0.104$).

Self-report correlations
To ensure that participants were using attitudes as a measure of similarity, we examined the correlation between participants' perceived similarity to the SD contestant and how much they agreed with the SD contestant (Pearson's Correlation: $r = .685$ $P = 0.003$). No such significant correlation was found with the SU contestant ($r = .095$ $P = 0.374$). Similarity to the SD contestant also correlated with how rewarding it was to see the SD contestant win ($r = .547$; $P = 0.021$), empathy scores ($r = .492$; $P = 0.037$), and perspective taking scores ($r = .580$; $P = 0.015$). How rewarding it was to see the SD contestant win, empathy and perspective taking scores did not correlate with how similar participants felt they were to the SU contestant. However, similarity to the SU contestant did correlate with how jealous the participants felt seeing the likeable contestant win ($r = .744$; $P = 0.002$), decreased empathy scores ($r = -.726$; $P = 0.002$) and how frustrated they felt when they lost on the game themselves ($r = .748$; $P = 0.001$).

Similarity correlations
As mentioned above (also see Fig 1B), participant's found themselves to be more similar to the SD than the SU contestant. Similarity correlated with left vACC (-12 34 0; $Z = 3.19$, $P = 0.031$ small volume corrected), right vmPFC (16, 48, 2; $Z = 3.21$, $P = 0.039$ small volume corrected), left anterior vmPFC (-16, 60, -22; $Z = 3.02$, $P < 0.0005$ uncorrected), and right hippocampus activation (28, -16, -22; $Z = 3.89$, $P < 0.0005$ uncorrected; Fig. S2).
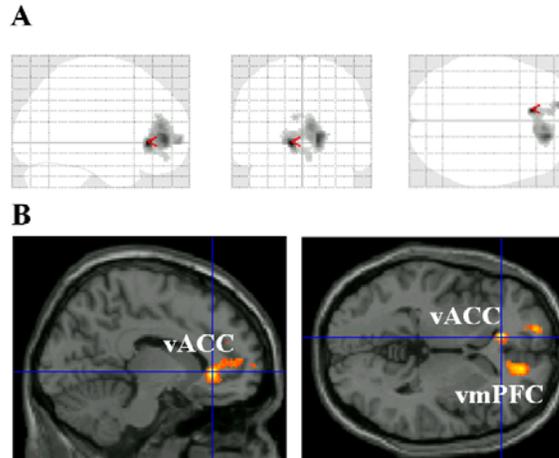
Figure S2. Activity overlaid on the **(A)** glass brain and **(B)** canonical brain for the regression of similarity to SD minus SU contestant.

Reward when observing the contestants win
Participants found it more rewarding to watch the SD contestant win (See Fig. 1C). A correlational analysis involving how rewarding the participants found it when observing the SD > SU contestant winning revealed significant increases in the right hippocampus (32, -16, -26; Z = 3.58, P < 0.0005 uncorrected), right ventral striatum (10, 16, -6, Z = 3.05, P = 0.044 small volume corrected), bilateral frontal pole (Left: -16, 60, -4; Z = 3.06, P < 0.0005 uncorrected; Right: 20, 60, -8; Z = 3.04, P < 0.0005 uncorrected), and left dorsomedial PFC (-4, 46, 48; Z = 3.15, P = 0.0005 uncorrected). No vACC was found for this correlation.

Likeability correlations
Because we found a strong correlation between likeability and similarity, we next examined the neural systems that correlate with likeability. As with similarity, the vACC was active for likeability. (-12, 36, -2; Z = 3.05; P = 0.001 uncorrected. PPI anaylsis also showed the vACC to be connected to the ventral striatal activity, albeit at a liberal threshold (2, 40, -2. P = 0.004 uncorrected). To separate likeability from similarity, we did a direct subtraction of likeability from similarity scores (i.e. we subtracted similarity to SU from similarity to SD. We then subtracted positive traits for the socially undesirable contestant from those for the socially desirable contestant).

Similarity minus likeability correlations
Although previous research suggests that similarity and likeability strongly correlate, we attempted to dissociate these two variables by subtracting likeability scores from similarity scores. We found that these residual similarity scores also significantly correlated with increased activity in the left vACC (-10, 32, 0; Z = 3.70, P < 0.0005 uncorrected), right vmPFC (16, 48, 2; Z = 3.17, P < 0.0005 uncorrected) and right hippocampus (28, -20, -20; Z = 3.04, P < 0.0005 uncorrected).

PPI results
We explored the connectivity between the vACC and the VS, and we found that the coupling between these regions was changed as a function of both viewing a SD contestant winning and as a function of how much participants felt similar to that contestant (-6, 38, -8; P = 0.043 small volume corrected). An additional figure shows the exact location of the activity (Figure S3).

A **B**
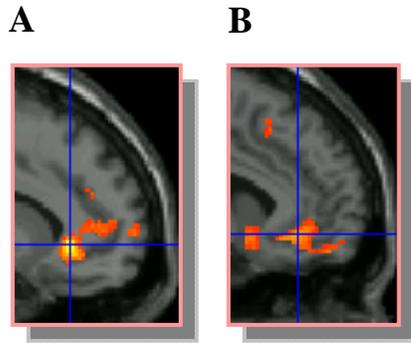
Figure S3. (A) vACC region correlated with similarity and (B) vACC connectivity to the VS seed.

Self-play gambling task.
In addition to viewing the contestant play participants were also asked to play the game. The fMRI results were similar to those of a near identical study (*1*) showing that winning minus loss resulted in activity in the ventral striatum (18, 14, -8; Z = 3.88; P < 0.0005whole-brain corrected; Figure S4). We used this MNI coordinate as a seed for the PPI (see PPI analysis above).
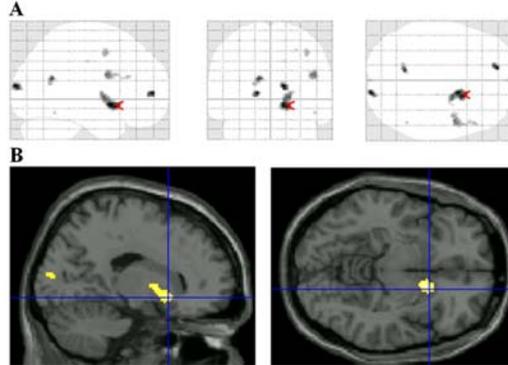
Figure S4. Activity overlaid on the **(A)** glass brain and **(B)** canonical brain for the self winning.

Comparison between self-win and SD>SU win.
A follow-up analysis directly compared experiencing win > loss with observing SD > SU contestant win. To do so, we treated two experiments as two sessions in a single design matrix. We found that the ventral striatum (14, 12, -4; Z = 3.41 P < 0.0005 uncorrected) was more activated for experiencing win > loss, suggesting that experiencing monetary winning is more rewarding than merely observing others winning (Figure S5 A). The bilateral superior temporal sulcus (STS; -54 -32, 10; Z = 3.04; P = 0.001 uncorrected; Right 64 -18 8; Z = 3.53, P < 0.0005 uncorrected) was more engaged when observing SD > SU contestant win minus self-win (Figure. S5 B), suggesting that observing others win recruits brain areas associated with social attention and theory of mind (S5, S6).
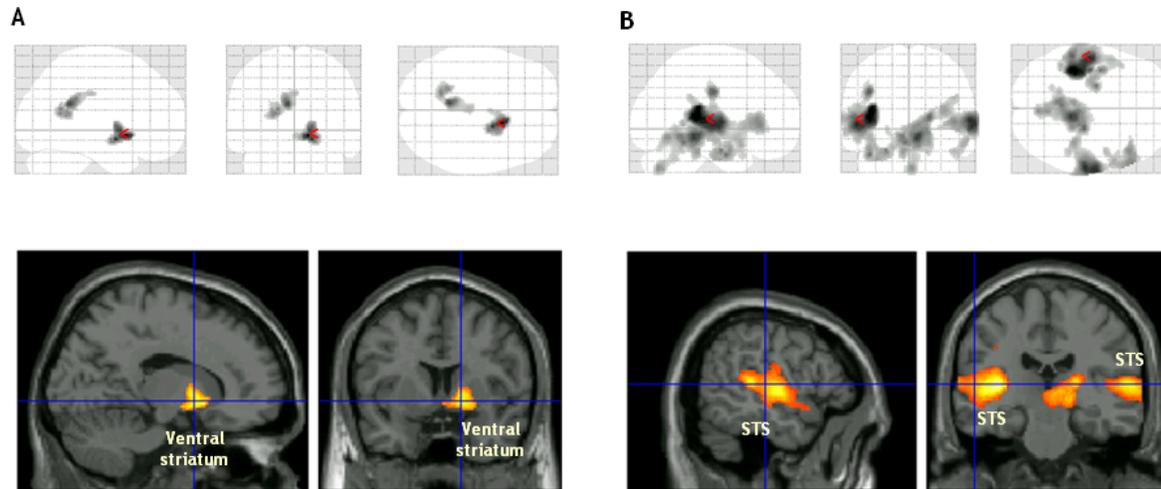
Figure S5. Activity overlaid on a glass and canonical brain for the direct comparison between (A) Self-win minus SD>SU other winning; and (B) SD > SU minus self-win winning. STS = superior temporal sulcus.

References

S1. Anderson, N.H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 9, 272-279.

S2. Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology,* 10, 85.

S3. Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology,* 44, 113-126.

S4. Gitelman, D.R., Penny, W.D., Ashburner, J., Friston, K.J., 2003. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage* 19, 200–207.

S5. Nummenmaa L., **Calder, A.J.** (In Press) *Neural mechanisms of social attention*, Trends in Cognitive Science.

S6. Frith,C., Frith,U. (1999). Interacting minds - A biological basis. *Science* 286 (5445), 1692-1695.