A Distinct Role of the Temporal-parietal Junction in Predicting Socially Guided Decisions

R. McKell Carter[1,2,3], **Daniel L. Bowling**[1,2], **Crystal Reeck**[1,4], and **Scott A. Huettel**[1,3,4,*]

[1]Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

[2]Department of Neurobiology, Duke University, Durham, NC, USA

[3]Brain Imaging and Analysis Center, Duke University, Durham, NC, USA

[4]Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

Abstract

To make adaptive decisions in a social context, humans must identify relevant agents in the environment, infer their underlying strategies and motivations, and predict their upcoming actions. We used functional magnetic resonance imaging (fMRI), in conjunction with combinatorial multivariate pattern analysis (MVPA), to predict human participants' subsequent decisions in an incentive-compatible poker game. We found that signals from the temporal-parietal junction (TPJ) provided unique information about the nature of the upcoming decision, and that information was specific to decisions against agents who were both social and relevant for future behavior.

By observing the choices and behavioral cues of others, individuals can learn about, and rapidly adapt their behavior to, dynamic environments *(1, 2)*. Indeed, social information obtained by observing other agents' actions optimizes the learning of environmental contingencies *(3)*. Mental state attribution is the lynchpin in this process, enabling individuals to decipher the motivations underlying the behavior of other agents and predict their upcoming actions *(1, 4, 5)*. Different agents have different relevance for behavior, and failing to distinguish relevant from irrelevant agents may lead to following deceptive advice *(6)* or conforming with the attitudes or opinions of others *(7, 8)*. Accordingly, some of the most striking examples of human social cognition reflect the differentiation between more and less important social partners: more heavily weighting the preferences of key players when making group decisions *(9, 10)*, discriminating strategic actions from random behavior *(11)*, and comparing oneself to more similar others *(12)*.

A key challenge for neural models of social cognition comes from the very ubiquity of social signals. Social signals in the natural environment are frequent, highly salient, and good predictors of needed future behavior. Information derived from social agents, therefore, might be processed by generalized neural systems that are geared towards identifying *any* important events in the environment, rather than by specialized systems that carry information in specifically social contexts. This challenge, and the difficulty in distinguishing its predictions from that of social cognition models, lies at the heart of debates in the field *(13, 14)*.

To explore the interaction of social agency with behavioral relevance, we used a simplified poker game *(15)* in which participants (*n* = 18) played against human and computer opponents (Fig. 1). The opponents alternated across eight functional runs, and the same real

[*]To whom correspondence should be addressed. scott.huettel@duke.edu.

human opponent competed against all participants in an independent and incentive-compatible manner (see Supporting Methods). The human and computer opponents were matched on overall decision probabilities, so that differences in observed behavior and brain function could be attributed to perceived social agency.

In each trial, participants first viewed a picture of their opponent. Participants were then presented with either a high card that would win if they chose to bet and were called, or a low card, indicating that would lose if they chose to bet and were called. When the participant did bet, control of the game passed to the opponent who then decided to call or fold (i.e., bet by adding more money, or fold by ceding the pot to the participant). Participants won money in this game when opponents bet on trials when they held high cards or folded on trials where the participant bet while holding a low card. Thus, participants maximized earnings in this game through bluffs that prevented the opponent from accurately guessing the cards that were held. On average, participants bluffed 54% of the time (range 29-73%), consistent with the equilibrium solution to this game *(15)*. Participants' bluffing decisions were significantly influenced by their opponent's behavior on the previous trial [$t(17) = 8.8$, $P < 0.001$], with greater effects found for human compared to computer opponents [$t(17) = 2.3$, $P = 0.032$].

Our fMRI analyses used MVPA at the time of card presentation to predict the participant's subsequent decision, which occurred 6 seconds later in the trial. We conducted a combinatorial whole-brain analysis that allowed estimation of the unique information carried within local brain segments *(16, 17)*. For every segment, we calculated the average increase or decrease in predictive performance when paired with other brain segments (see Supporting Materials). We refer to this quantity as the Unique Combinatorial Performance (UCP). A high UCP means that a region carries information that can predict future behavior, beyond that obtained from other brain regions. A low UCP means that information within a region is largely redundant with that of other regions. Across both opponents, significant UCP was found in regions often associated with social cognition, including regions in both the dorsal and medial prefrontal cortices (Table S2).

A standard approach for establishing functional specificity in neuroscience research is subtraction; i.e., statistical comparison of two conditions that differ only in one process of interest. However, activations identified by subtraction may not be unique to the process of interest (Fig. 2C). In contexts like the current task, signals associated with selective attention and vigilance may be present in both social and non-social contexts, but are more engaged when a social stimulus is present *(18, 19)*. To remove these general cognitive effects, we regressed UCP for a human opponent against UCP for a computer opponent (Fig. 2B). Across regions, there was a robust correlation with a slope greater than one (slope = 1.4, $p = 0.016$). The general consistency of this response – including for brain regions often described within social networks – implies that the processes engaged during the performance of this task are greater for, but not specific to, social context.

To quantify whether any local brain regions carried distinctly social information, we defined a metric of *social bias* (Fig. 2D) using the normalized residual distance from the whole-brain regression line shown in Fig. 2B (see Supporting Online Methods). A region with a positive social bias has a greater UCP against the human opponent than against the computer opponent. One region, the TPJ, exhibited an extreme social bias that was five standard deviations higher than the mean of all other brain segments (Fig. 2E). The gap between the social bias found in the TPJ and that found in the second-most-biased region was greater than the range of variation throughout the remainder of the brain. This independent contribution of the TPJ was robust to multiple forms of correction for brain segment

characteristics (Fig. S5 and S6) and to Bonferroni correction for multiple comparisons (see Supporting Materials).

These findings identify a unique and independent role for the TPJ in representing information predictive of behavioral actions during social interactions. To rule out any potential biasing effects of this particular segmentation approach – which defined segments of the brain according to prior anatomical and functional divisions – we repeated all analyses using random, voxel-based segmentations of the brain (see Supporting Materials). Voxel-based UCP metrics were then calculated using the median UCP across the different segmentations (Fig. S7). The analysis replicated both previous findings: a non-specific overall increase in UCP against human compared to computer opponents (slope of 1.4, across all gray matter voxels), but strong social bias in the TPJ (Fig. S7D). Moreover, the specificity of TPJ was maintained when extending analyses to combinations of three regions (Figs. S9, S10).

We next evaluated how the subjective behavioral relevance of each opponent modulated information within the TPJ. After completion of the scanning session, participants indicated whether the human or computer was a better opponent. The social bias found in TPJ was present only for participants who judged the human opponent to be superior (Fig. 3A); again, its bias was more than five standard deviations from the mean normalized residual (Fig. 3B). Conversely, no social bias was observed in TPJ for those who judged the computer opponent as superior to the human [$t(6) = 1.3$, $P = 0.26$; difference between groups: $t(16) = 2.6$, $P = 0.02$].

For comparison we also plotted UCP for the medial prefrontal cortex (MPFC) and medial posterior cortex (MPC), a network of anatomical regions that, along with the TPJ, have been linked to aspects of social cognition (20-23). None of these regions exhibited a significant social bias. These results were robust to neural segmentation method selection, with analyses employing random voxel-based neural segmentation producing very similar results (Fig. S8). Our findings indicate the MPFC and MPC contribute similarly to predicting future behavior in both our social and non-social settings, suggesting that they support computations that are critical for social cognition but that are also applied in at least some non-social settings (24, 25).

Our results point to a specific role for TPJ within the social cognition network and demonstrate the unique sensitivity of this region to perceived behavioral relevance of other agents. The TPJ stands out as a contributor of unique and independent social information, even compared to other regions considered to be recruited by social cognition (Figs. 2,3). This finding indicates a potential reconciliation for the opposing perspectives that have been proffered for TPJ function: mental inference (26) and orienting to salient external stimuli (14, 27). Responses in the TPJ during social cognition tasks may reflect the nexus of these two operations – that is, interacting with an opponent whose internal states can be modeled (i.e., another human) and whose behavior is also relevant for guiding one's future actions. This interpretation is consistent with demonstrations that TPJ is activated when inferring an advisor's motives when the advice is relevant for an upcoming decision (6) or on potentially highly rewarding trials when implementing an advanced strategy during an interactive choice game (28). However, the present results demonstrate that the TPJ is not generally recruited by external social stimuli, but specifically contributes to choice deliberation when modeling the behavioral and cognitive tendencies of a social agent. Accordingly, when others are of lower status (29) or not considered relevant for future behavior – as in the case of dehumanization (30) – the TPJ may be disengaged. The sensitivity of TPJ to both social context and perceived relevance highlights a critical role for this region in coordinating

behavior in a dynamic, social environment and demonstrates the capacity of behavioral relevance to modify neural signals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and Notes

1. Lee D. Nature neuroscience. Apr.2008 11:404–9.

2. Danchin E, Giraldeau L-A, Valone TJ, Wagner RH. Science. Jul.2004 305:487–91. [PubMed: 15273386]

3. Rendell L, et al. Science (New York, N.Y.). Apr.2010 328:208–13.

4. Behrens TEJ, Hunt LT, Rushworth MFS. Science. 2009; 324:1160–4. [PubMed: 19478175]

5. Olsson A, Ochsner KN. Trends in cognitive sciences. Mar.2008 12:65–71. [PubMed: 18178513]

6. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS. Nature. Nov.2008 456:245–9. [PubMed: 19005555]

7. Asch SE. Scientific American. Nov.1955 193:31–35.

8. Raafat RM, Chater N, Frith C. Trends in cognitive sciences. Oct.2009 13:420–8. [PubMed: 19748818]

9. Kameda T, Ohtsubo Y, Takezawa M. Journal of Personality and Social Psychology. 1997; 73:296–309.

10. Bonaccio S, Dalal R. Organizational Behavior and Human Decision Processes. Nov.2006 101:127–151.

11. Yoshida W, Seymour B, Friston KJ, Dolan RJ. Journal of Neuroscience. 2010; 30:10744. [PubMed: 20702705]

12. Festinger L. Human Relations. May.1954 7:117–140.

13. Mitchell JP. Cerebral Cortex. 2008; 18:262. [PubMed: 17551089]

14. Corbetta M, Patel G, Shulman GL. Neuron. May.2008 58:306–24. [PubMed: 18466742]

15. Rapoport A, Erev I, Abraham EV, Olson DE. Organizational Behavior and Human Decision Processes. Jan.1997 69:31–49.

16. Hampton AN, O'Doherty JP. Proceedings of the National Academy of Sciences of the United States of America. Jan.2007 104:1377–82. [PubMed: 17227855]

17. Clithero JA, Carter RM, Huettel SA. NeuroImage. May.2009 45:1329–38. [PubMed: 19349244]

18. Goren CC, Sarty M, Wu PY. Pediatrics. Oct.1975 56:544–9. [PubMed: 1165958]

19. Ro T, Russell C, Lavie N. Psychological science. Jan.2001 12:94–9. [PubMed: 11294237]

20. Blakemore S-J. Cerebral Cortex. Aug.2003 13:837–844. [PubMed: 12853370]

21. Tankersley D, Stowe CJ, Huettel SA. Nature neuroscience. Feb.2007 10:150–1.

22. Fletcher P. Cognition. Nov.1995 57:109–128. [PubMed: 8556839]

23. Mitchell JP, Schirmer J, Ames DL, Gilbert DT. Journal of cognitive neuroscience. Apr.2011 23:857–66. [PubMed: 20350058]

24. Heider F, Simmel M. The American Journal of Psychology. 1944; 57:243–259.

25. Krach S, et al. PloS one. Jan.2008 3:e2597. [PubMed: 18612463]

26. Saxe R, Kanwisher N. NeuroImage. Aug.2003 19:1835–1842. [PubMed: 12948738]

27. Decety J, Lamm C. The Neuroscientist. 2007; 13:580. [PubMed: 17911216]

28. Bhatt MA, Lohrenz T, Camerer CF, Montague PR. Proceedings of the National Academy of Sciences of the United States of America. Nov.2010 107:19720–5. [PubMed: 21041646]

29. Shepherd SV, Deaner RO, Platt ML. Current biology: CB. Feb.2006 16:R119–20. [PubMed: 16488858]

30. Harris LT, Fiske ST. European Review of Social Psychology. 2009; 20:192–231.

One Sentence Summary

The brain's temporal-parietal junction provides unique information toward predicting future decisions in a social context.

Fig. 1. A single-card poker game
Each trial begins with a picture of the opponent (i.e., a photograph of the human opponent or of a computer). The participant's card is revealed and they then decide whether to bet (B) or fold (F). Buttons then appear – with left/right position randomized across trials – and the participant indicates their response. If the participant chooses to fold, the trial ends. If the participant chooses to bet on the card, control passes to the opponent who has 3 s to decide whether or not to match the participant's bet or fold. Predictions are made using fMRI data associated with the time period highlighted in green.
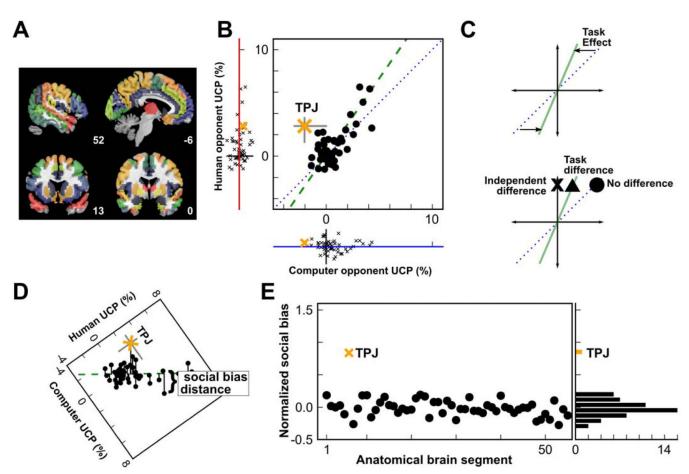
Fig. 2. TPJ plays a distinct role in predicting social decisions
(A) Whole-brain fMRI data was segmented into 55 bilateral anatomical regions, as described in Supplementary Material. (B) UCP for a human opponent was regressed against UCP for a computer opponent, across regions (green dashed line, slope = 1.4). For the range reported here, negative values of UCP correspond to a return to chance levels and not negatively predictive models. (C) The slope of the regression line in panel B is consistent with a general task effect (upper graph), such that social contexts tend to engage task-related regions across the brain more than similar non-social contexts. A region that was highly engaged by the task might exhibit activation in a standard fMRI contrast (triangle), even if that region carried no independent task-related information. A systematic shift away from the regression line ('x') would be necessary to show an independent effect in a region. (D) Social bias was defined by the residual difference (social minus non-social) in predictive power from the whole-brain regression line. (E) The magnitude of social bias in each region (normalized by the variance estimates for each segment's UCP), sorted by UCP against the human opponent. Only one region, the TPJ, contributed significantly more UCP against a human opponent than against a computer opponent, independent of the overall task effect.
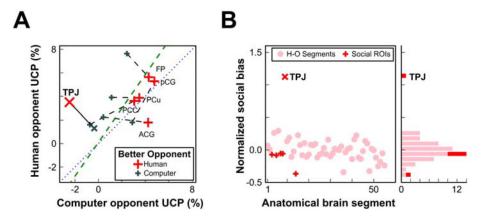
Fig. 3. Social bias in the TPJ depends on the participant's evaluation of the opponent
(A) The social bias observed in the TPJ was found only for those participants who considered the human to be a more informative opponent. Other putative social cognition regions do not show the same relationship with behavioral relevance (MPFC: FP-frontal pole, pCG-paracingulate gyrus, and ACG-anterior cingulate gyrus; MPC: PCC-posterior cingulate cortex and PCu-precuneus). (B) Social bias in all regions (normalized by the variance estimates for each segment's UCP), sorted by UCP against the human opponent.