Neural mechanisms underlying the computation of socially inferred rewards

Natalia Vélez (nvelez@stanford.edu) and Hyowon Gweon (gweon@stanford.edu)
Department of Psychology, Stanford University, 450 Serra Mall
Stanford, CA 94305 USA

Abstract

No one knows everything. Therefore, it is often not enough to rely solely on one's own knowledge, nor to indiscriminately follow advice from others. The current work examines the neural systems that support the human ability to capitalize on imperfect social information to support decision-making. Participants completed an fMRI task where they could choose to stay with an option of known value or switch to a hidden option, while receiving advice from an advisor who had access to both options, no options, or only the option that was hidden from participants. First, we find that value-guided regions (including dorsal striatum, dMPFC) preferentially track the expected value of the hidden option when it is the only option the advisor can access. Second, the advisor's knowledge state is represented in regions that support social reasoning (precuneus, vMPFC). Our results suggest that neural systems that support social cognition and value-based decision-making support computations that enable humans to harness social information to vicariously explore the value of latent options.

Keywords: fMRI; social learning; decision making

Introduction

Social learning—learning from others—is essential for adaptive behavior. By learning from others, humans can learn more about the world than what they can directly experience (Boyd et al., 2011). However, other people's knowledge is often as limited as our own, and their advice may not be perfectly helpful. Thus, it is often not enough to accept social information indiscriminately nor to ignore it entirely; to arrive at optimal decisions, one must integrate one's own knowledge with information from others (Bahrami et al., 2010).

Recent computational work has examined how human learners make utility-maximizing decisions by "putting two heads together" (Vélez & Gweon, 2018). In this study, participants played a card game where they chose to "stay" with a card of known value or "switch" to an unknown card, given an advisor's advice to stay or switch. Participants used advice strategically based on which cards the advisor could see and how the advisor selected advice, and their responses were consistent with a Bayesian Theory of Mind model that leverages the advice to infer the value of the unknown card. These results support the idea that human learners do not simply accept social information at face value. Instead, people are able to *infer* the value of options that they have not directly experienced by harnessing an intuitive understanding of how others' knowledge and goals gives rise to their observable options.

Cognitive models offer computational- and algorithmic-level descriptions of how mental processes generate overt decisions and behaviors (Marr, 1982), but often without explaining how these inferences are implemented in the brain. One stunning exception is reinforcement learning (RL) models, which have successfully bridged precise computational descriptions of the cognitive mechanisms underlying decision-making to their neural implementation (Dayan & Niv, 2008). However, past work on RL approaches to social decision-making has largely focused on identifying neural correlates of *observable* aspects of social information, such as others' accuracy and trustworthiness (e.g., Boorman et al., 2013; King-Casas et al., 2005). Less is known about the mechanisms that enable humans to use social information to discover the value of *latent* options.

The current work uses model-based fMRI to investigate how neural systems that support social cognition and value-based decision making track socially inferred rewards. Building on the computational model and the behavioral task developed in prior work (Vélez & Gweon, 2018), the current work has two major goals. The first is to identify neural signals that track the expected value of options that are inferred through social information (here, the value of the hidden card) and to test whether regions that support value-guided decision-making preferentially track this value when the advisor only has access to the hidden card, compared to conditions where the advice is fully informative or totally uninformative. The second goal is to examine whether brain regions implicated in social cognition (i.e., the Theory of Mind network Dodell-Feder et al., 2011) represent the advisor's epistemic state.

Methods

Participants 20 participants (10F, ages 19–40, M(SD) age = 24.9(6.7)) were recruited for an fMRI study. Participants were right handed, had normal or corrected-to-normal vision, and provided informed consent in accordance with the requirements of the IRB. Participants were paid $40 for their time and a bonus of up to $10 proportional to their final score.

fMRI task Participants played a simple card game in the scanner (Figure 1A; for more details on the experiment procedure and results, see Vélez & Gweon, 2018). On each trial, two cards with different point values between +1 and +6 were drawn. One card was visible to the participant (*visible card*), while the other was hidden from them (*hidden card*). Participants chose to *stay* and keep the points in the visible card, or *switch* to the points in the hidden card.

In every trial, participants received advice from an "advisor" who saw a subset of the cards and recommended whether participants should stay or switch. Before the fMRI task, par-
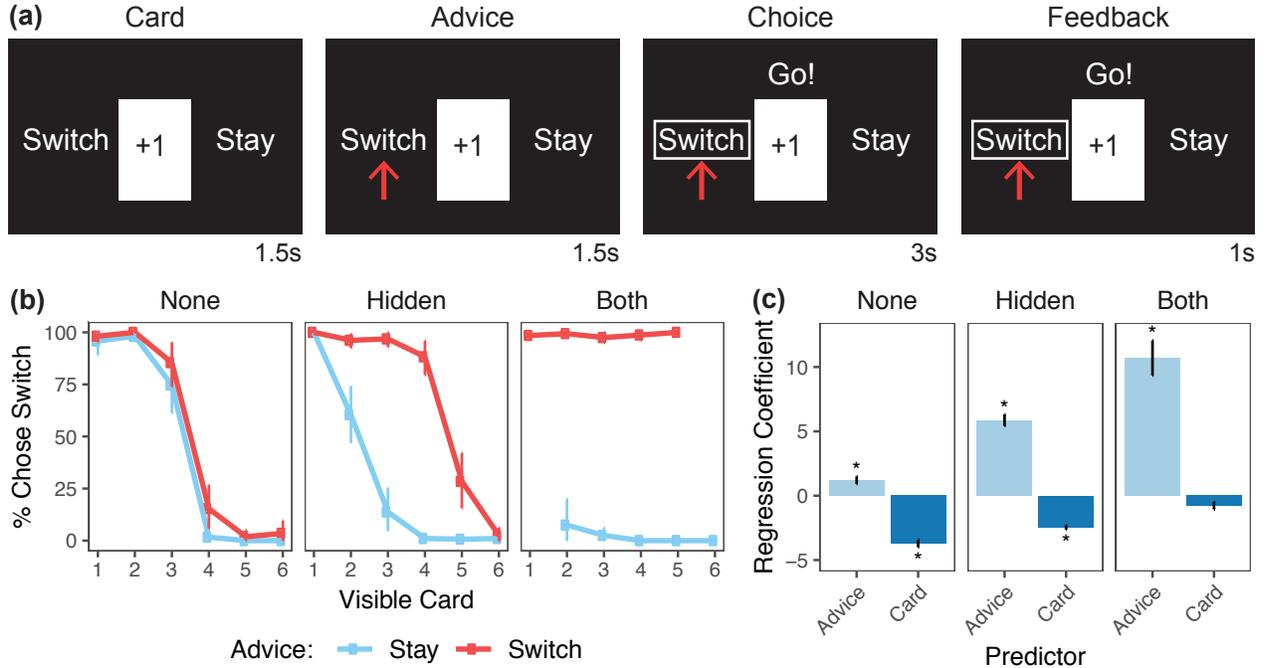
Figure 1: fMRI task design and behavioral results. (a) Trial structure. (b) % of trials in which participants switched. (c) Mixed-effects logistic regression coefficients relating the advisor's advice and the value of the visible card to participants' choices to stay or switch. $*$ denote deviations from 0 ($p < .05$). Error bars denote (b) bootstrapped 95% CI and (c) model standard errors.

ticipants were introduced to a human confederate playing the role of the advisor and were led to believe that they would play the game with the confederate; in reality, in the scanner task, the advice was generated by a simulated agent. The advisor's access to information varied across three within-subjects conditions. In the *Both* condition, the advisor saw both cards and deterministically recommended the best option (i.e., advising to stay if the visible card has the larger value, and switch if the hidden card has the larger value). In the *None* condition, the advisor saw no cards and provided advice at random. In our critical condition, the *Hidden* condition, the advisor saw only the card that was hidden from the participant. In this condition, the advisor sampled advice (A) using a softmax function:

$$P(A = \text{Switch}|H) = \frac{1}{1 + e^{-(\beta_A \times (H - H_{med}))}} \quad (1)$$

where H is the value of the hidden card; $H_{med} = 3.5$, the median card value, centers the advisor's choice function on the range of possible card values; and $\beta_A = 1.5$ is a free temperature parameter that was estimated from average human responses in an earlier task (Vélez & Gweon, 2018). Thus, the advisor's advice was generally informative; the advisor was more likely to recommend switching as the value of the hidden card increased. However, because the advisor did not have full access to information about the cards, following the advisor's advice does not guarantee the best outcome.

Participants completed 5 runs of the task, each comprising 6 blocks of 6 trials (2 blocks/condition, palindromic order).

fMRI task Based on prior work (Vélez & Gweon, 2018), we modeled participants' responses using a model that uses advice to infer the value of the hidden card. Our model infers the value of the hidden card on each trial (C_H) given the advice (A) and the value of the visible card (C_V):

$$P(C_H|A, C_V) \propto P(C_H|C_V)P(A|C_H) \quad (2)$$

where $P(C_H|C_V)$, a discrete distribution that is 0 for $C_H = C_V$ and uniform everywhere else, represents the learner's prior belief about the value of the hidden card, and $P(A|C_H)$ (Eq. 1) represents the likelihood of the observed advice based on the advisor's belief about the hidden card value.

Based on its beliefs about the value of the hidden card ($P(C_H|A, C_V)$), the model then computes an expected value ($\mathbf{E}[C_H|A, C_V]$). Finally, the model selects an action (L) by comparing the value of the visible card to the expected value of the hidden card using a softmax choice rule:

$$P(L = \text{Switch}|A, C_V) = \frac{1}{1 + e^{-\beta_C \times (\mathbf{E}[C_H|A, C_V] - C_V)}} \quad (3)$$

where β_C is a free parameter that modulates how much the learner's choices are influenced by expected rewards.

fMRI Processing & Analysis

Data acquisition MRI data were collected using a 3T MRI scanner (GE Discovery MR750). Functional images were acquired in interleaved order using an EPI sequence (45 transverse slices, TR = 2s, TE = 30ms, voxel size = 2.9mm isometric). Spiral fieldmap images were collected every 20 minutes between functional runs. Anatomical images were acquired at the end of the session (T1 MPRAGE, voxel size = 0.9mm isometric).

Preprocessing Data were preprocessed with fmriprep, using fieldmap images for distortion correction (v1.0.8 Esteban et al., 2018). Functional runs were later smoothed with FSL SUSAN (FWHM = 5mm).

ROI definition Functional ROIs were defined in each participant using an independent functional localizer (Dodell-Feder et al., 2011). We masked thresholded images ($p < .001$ uncorrected) using ROI hypothesis spaces (Dufour et al., 2013) and selected the cluster within each hypothesis space containing the peak voxel.

GLM We defined a GLM to identify neural correlates of the expected value of the hidden card. For each condition (Both, Hidden, None) we defined 6 regressors of interest (18 regressors total): a boxcar regressor spanning all trials in a condition (from the card phase to the end of the choice phase), two parametric regressors, each marking the value of the visible card and the expected value of the hidden card, two boxcar regressors spanning trials where the advisor suggested to stay or switch, respectively, and a boxcar regressor spanning the length of the feedback phase. We also included nuisance regressors estimated using fmriprep (framewise displacement, 6 motion regressors, 6 noise components). We estimated the models using FSL FILM and generated cluster-corrected p-values using FSL Randomise.

MVPA To test whether Theory of Mind ROIs represent the advisor's access to information, we estimated trial-by-trial responses in each voxel using a beta series regression. We then trained a linear, multiclass classifier to label trials by condition based on the pattern of responses within each functional ROI, and we measured cross-validation accuracy in the left-out run.

Results

Behavioral results

Participants believed that they played the task with a human partner. In a post-test, 18/20 participants indicated that they slightly or strongly believed that the advice they received was provided by a human (4-point Likert scale, "strongly disbelieved" to "strongly believed"; median: 4, mode: 4).

Participants flexibly combined the advisor's advice with the value of the visible card to make decisions (Figure 1b–c). In the Both condition, participants' responses were primarily influenced by the advisor's advice (main effect of advice: $\beta = 10.7$, $z = 7.9$, $p < .0001$; visible card: $\beta = -0.8$, $z = -2.7$, $p = .007$). In the None condition, participants' re-

sponses were strongly influenced by the value of the visible card ($\beta = -3.7$, $z = -13.0$, $p < .0001$) and slightly biased by advice ($\beta = 1.2$, $z = 4.1$, $p < .0001$). In the Hidden condition, participants' responses were influenced by both the the value of the visible card ($\beta = -2.5$, $z = -14.6$, $p < .0001$) and advice ($\beta = 5.9$, $z = 13.0$, $p < .0001$).

Expected value of the hidden card

We tested whether regions that support value-based decision-making track the expected value of the hidden card. One region, intraparietal sulcus, tracked the expected value of the hidden card in all conditions (Figure 2a). Our key question was whether there are regions that selectively track this value in the hidden condition, where participants had to *infer* the value of the cards based on both the advisor's advice and their own knowledge of the value of the visible card. We identified several regions that are more strongly modulated by the expected value of the hidden card in the Hidden condition than in the Both and None conditions: bilateral dorsal striatum, middle frontal gyrus, precuneus, and dorsomedial prefrontal cortex (Figure 2b, circled, left-to-right).

Advisor's access to information

Of the functionally-defined regions within the Theory of Mind network, precuneus (PC, 36.1%; accuracy, 95% CI: $[34.2, 38.1]$, $t(18) = 2.98, p < 0.008$) and ventromedial prefrontal cortex (vMPFC, 36.1%; 95% CI: $[34.1, 37.7]$; $t(15) = 3.00, p = .008$) discriminated between conditions (Figure 3A). These regions represent the conditions in qualitatively different ways: While PC contains distinct representations of each condition (Figure 3b), vMPFC discriminates conditions where social information is relevant to choice (Both, Hidden) from conditions where it is not (None; Figure 3c).

Discussion

The current work leveraged a Bayesian Theory of Mind model to examine how the computations involved in rich mental state reasoning support value-based decision making. First, we identified regions that track the inferred value of the hidden
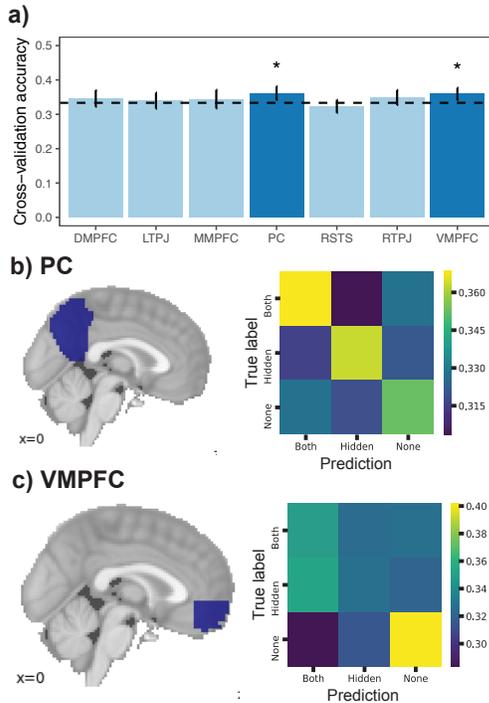
Figure 3: MVPA results. a) 3-way cross-validation accuracy in each region within ToM network. denotes deviations from chance (one-sample t-test). b–c) Confusion matrices in PC and VMPFC. y-axes denote true condition; x-axes denote model predictions; fill denotes average co-occurrence.

card. While IPS, a region implicated in numerical cognition (Nieder & Dehaene, 2009), reflected this value across all conditions, striatum, precuneus, frontal gyrus, and dMPFC reflected this value selectively in the Hidden condition, where neither participants' knowledge nor the advisor's advice alone were sufficient to make optimal decisions. Second, precuneus and vMPFC represented the advisor's access to information. These results are consistent with prior work, which finds that regions that support Theory of Mind represent abstract features of mental states, such as emotional valence and the strength of evidence supporting a belief (Koster-Hale et al., 2017).

Put together, our work identifies how neural systems that support Theory of Mind and value-guided decision-making each contribute to decision-making based on imperfect social information. However, it does not speak to how these networks may interact, and how interactions between these networks might shift based on how strongly one relies on social information make decisions. Ongoing work is currently addressing this question.

Our results complement reinforcement learning approaches to social cognition, which have provided foundational insights into how regions that support value-based choice represent and make use of social information (Behrens

et al., 2009). Going beyond this work, we find that these regions track not only observable properties of social information—such as their accuracy—but also the value of latent options that are only visible to others. Our work provides an example of how Bayesian models of social cognition can directly inform hypotheses about neural computations. We hope that this work is a first step towards a more pluralistic and complete computational account of the neural mechanisms underlying human social cognition.

References

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081–1085.

Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behavior. *Science*, *324*(5931), 1160–1164.

Boorman, E. D., ODoherty, J. P., Adolphs, R., & Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron*, *80*(6), 1558–1571.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*, 10918–10925.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology*, *18*(2), 185–196.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fmri item analysis in a theory of mind task. *Neuroimage*, *55*(2), 705–712.

Dufour, N., Redcay, E., Young, L., Mavros, P. L., Moran, J. M., Triantafyllou, C., ... Saxe, R. (2013). Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PloS one*, *8*(9), e75468.

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Aliaga, A. E., ... others (2018). Fmriprep: a robust preprocessing pipeline for functional mri. *bioRxiv*, 306951.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, *308*(5718), 78–83.

Koster-Hale, J., Richardson, H., Velez, N., Asaba, M., Young, L., & Saxe, R. (2017). Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *NeuroImage*, *161*, 9–18.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: WH Freeman.

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual review of neuroscience*, *32*, 185–208.

Vélez, N., & Gweon, H. (2018). Integrating incomplete information with imperfect advice. *Topics in Cognitive Science*.