HHS Public Access
Author manuscript
Neuroimage. Author manuscript; available in PMC 2016 February 11.

Published in final edited form as:
Neuroimage. 2015 January 1; 104: 326–335. doi:10.1016/j.neuroimage.2014.09.011.

With you or against you: Social orientation dependent learning signals guide actions made for others

George I. Christopoulos[1,2] and **Brooks King-Casas**[1,3,4,5,6,*]

[1]Virginia Tech Carilion Research Institute, 2, Riverside Circle, Roanoke, VA, 24016, USA

[3]Department of Psychology, Virginia Tech, Blacksburg, VA, USA

[4]Department of Psychiatry, Virginia Tech Carilion School of Medicine, Roanoke, VA, USA

[5]Virginia Tech - Wake Forest University School of Biomedical Engineering and Sciences, Blacksburg, VA, USA

[6]Research Service Line, Salem VA Medical Center, Salem, VA, USA

Abstract

In social environments, it is crucial that decision-makers take account of the impact of their actions not only for oneself, but also on other social agents. Previous work has identified neural signals in the striatum encoding value-based prediction errors for outcomes to oneself; also, recent work suggests neural activity in prefrontal cortex may similarly encode value-based prediction errors related to outcomes to others. However, prior work also indicates that social valuations are not isomorphic, with social value orientations of decision-makers ranging on a cooperative to competitive continuum; this variation has not been examined within social learning environments. Here, we combine a computational model of learning with functional neuroimaging to examine how individual differences in orientation impact neural mechanisms underlying ‘other-value’ learning. Across four experimental conditions, reinforcement learning signals for other-value were identified in medial prefrontal cortex, and were distinct from self-value learning signals identified in striatum. Critically, the magnitude and direction of the other-value learning signal depended strongly on an individual’s cooperative or competitive orientation towards others. These data indicate that social decisions are guided by a social orientation-dependent learning system that is computationally similar but anatomically distinct from self-value learning. The sensitivity of the medial prefrontal learning signal to social preferences suggests a mechanism linking such preferences to biases in social actions and highlights the importance of incorporating heterogeneous social predispositions in neurocomputational models of social behavior.

[*]To whom correspondence should be addressed: T: +1-540-526-2009 bkcasas@vt.edu.
[2]Current affiliation: Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore, USA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could a3ect the content, and all legal disclaimers that apply to the journal pertain.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Keywords

reward learning; prediction error; social neuroscience; fMRI; social value orientation; social orientation

1. Introduction[1]

Navigating one's environment, whether it be foraging for food or interacting with social partners, requires evaluating available options and taking actions that are likely to benefit oneself. The application of formal learning models to the analysis of decision-related neural activity has begun to reveal the neural basis of computations underlying value-guided decision-making in humans (D'Ardenne et al., 2008; Daniel and Pollmann, 2014; Jocham et al., 2011). These data have shown that individuals learn the value associated with an action through experience by serially comparing expectations with outcomes (Krugel et al., 2009; Seymour et al., 2004; Sutton and Barto, 1988). Through this general process, humans dynamically learn how to value their actions and their environment, and dopaminergic signaling is believed to underlie these learning signals (Bayer and Glimcher, 2005; Delgado et al., 2008; den Ouden et al., 2010; Montague et al., 2006; Schultz et al., 1997).

This process becomes more complicated when making decisions that also impact others, whether friend, partner, adversary, or stranger. To successfully navigate such social transactions, it is crucial that decision-makers be able to assess (i) the value of the decision for oneself and (ii) the value of the decision to others, based upon one's own motivations toward oneself and the social partner. Previous studies have identified brain signals associated with outcomes delivered to oneself (Delgado et al., 2008; Galvan et al., 2005; Pessiglione et al., 2006; Ramnani et al., 2004) and outcomes delivered to others (Apps et al., 2013; Nicolle et al., 2012; O'Connell et al., 2013; Suzuki et al., 2012).

However, it is relatively less well understood how outcomes delivered to others are implemented in reinforcement learning environments. Behavioral research indicates that valuation of social outcomes, i.e. outcomes that involve other agents, depend on social preferences or motivations that can vary across individuals. For example, competitive types seek outcomes benefiting oneself at the expense of the social partner, while cooperative types seek outcomes benefiting both self and other (Fehr and Krajbich, 2014; Lurie, 1987; McClintock and Liebrand, 1988; Murphy and Ackermann, 2014). Studies of inequality aversion and guilt aversion have identified neural correlates of preferences over divisions of resources (Chang et al., 2011; Crockett et al., 2010; Fliessbach et al., 2007; Haruno and Frith, 2010; Tricomi et al., 2010), while measures of social value orientation have identified individual differences in neural correlates of these preferences (Haruno and Frith, 2010). However, the role of social preferences has not been taken into account in tasks that require learning the consequence of one's own action for social partners.

[1] Abbreviations. SVO: Social Value Orientation; mPFC: medial prefrontal cortex; PES: prediction error for outcomes (including negative rewards eg – $70) received for oneself (i.e. the decision maker); PEO: prediction error for outcomes received by another person. Vs: value of outcome delivered to Self; Vo: value of rewards punishments delivered to Other

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Here we examine the process by which decision-makers learn how actions map onto outcomes for others. In doing so, we first identify learning signals underlying value-based decision-making for others and differentiate these signals from value-based learning signals for oneself, replicating and extending previous research efforts; subsequently we show how these signals vary parametrically as a function of social value orientation. That is, in a large cohort of participants, we show that the direction and magnitude of learning signals based on the value of an outcome for a social partner vary with the cooperative or competitive orientation of the participant.

2. Materials and Methods

2.1. Overview of procedures

Prior to scanning, the social value orientation (competitive, individualistic or cooperative) of participants was assessed through a parametric estimation by a sequential testing procedure (PEST). In this assessment, participants chose between allocations of an endowment between the participant and an anonymous social partner. Participants were then instructed that they would make a series of choices while in the MRI scanner (Fig. 1). Additionally, they were told that their payment and the payment of the anonymous social partner would be based on a random subset of their choices. Seventy-two participants underwent 3T fMRI as they performed six manipulations of an instrumental learning task. In each condition, participants chose between two square fractals that were probabilistically (80:20) related to gains or losses for the decision-maker and another, unknown to them, participant (for instance $70 for the participant and −$70 for the other participant). The manipulations varied in the magnitude and valence of value assigned to oneself and the value assigned to the social partner. The order in which blocks were presented was pseudorandomized across participants.

2.2. Participants

Ninety participants (mean age 27.37 years; 28 female) were recruited from a college and community sample. Ten participants were excluded following the social value orientation assessment described below, as the parametric estimation by sequential testing (Luce, 2000) procedure did not produce reliable estimates across repeated measures. Seven additional subjects were excluded from neuroimaging analysis based on excessive movement during scanning. One subject was excluded as behavioral responses were not recorded for 40% of his/her trials. One condition [self −70 / other −70 ; self +70 / other +70] of a second subject was excluded for missing behavioral responses as well.

2.3. Social value orientation assessment

(see Fig. SM1). We employed a psychophysics-inspired assessment (Parameter Estimation by Sequential Testing – PEST) a non-learning choice task designed to assess social value orientation (SVO; (McClintock and Liebrand, 1988); (Kelly and Stahelski, 1970; Kuhlman and Marshello, 1975; Sattler and Kerr, 1991; Van Lange and Kuhlman, 1994). During the PEST procedure, participants serially made preference choices between two allocations. Each allocation included a number of points for the participant and a number of points to another anonymous participant. The two allocations were represented by a pair of numbers

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

placed to the left and right of a fixation cross (see SM Fig. 1a). One row indicated the number of points for the participant whereas the other row indicated the number of points for the second person. To increase the attention to both types of outcomes, the position of the ‘self’ and ‘other’ amounts (top or bottom row) was randomly determined on each trial. The participant pressed one of two keys to indicate the preferred choice. Following the key press, the choice was highlighted in red for 1 second and then a new trial started. The participant was informed that: (i) s/he would never meet the ‘other’ person or know each other’s identity; (ii) the other person would be paid according to one, randomly determined, choice outcome; and, (iii) the other person would not make similar or any kind of decisions influencing the participant’s payment.

Initial allocations—The allocation pairs presented to the participant for each decision were based on a two-dimensional geometric representation of value, where the x-axis represents outcomes to ‘self’ whereas the y-axis represents outcomes to ‘other’ (SM Fig. 1b). Allocations were positioned on a circle with a center at (0,0) and radius of 100. Initially, the allocations were rounded so that they were in multiples of 5. For example, an allocation with an angle of 45° between self and other axes and radius of 100 equals x-value of 70.7 (rounded to 70) and y-value of 70.7 (rounded to 70). The algorithm started by offering two different allocations. To ensure that the social value orientation derived from this procedure was reliable, we repeated the procedure three times during the same experimental session with differing initial allocations: in the first case, initial allocations were [SELF<0, OTHER<0] and [SELF<0, OTHER>0]; in a second case, initial allocations were [SELF<0, OTHER >0] and [SELF>0, OTHER<0]; and in a final case, initial allocations were [SELF<0, OTHER <0] and [SELF>0, OTHER>0].

Subsequent allocations—The values of the subsequent choice pairs were determined in part, by allocations of the previous choice pairs: the algorithm retained the unchosen allocation as one option whereas the chosen allocation was moved towards the unchosen option with a step angle derived from a uniform distribution with a mean of (2*pi/40) radians (9 degrees). The resulting values were again rounded to a multiple of 5. Based on this algorithm the two vectors gradually approximated the preferred vector for each participant. On each step the algorithm would check whether the step was larger than the difference between the two options, in order to prevent one option ‘crossing’ the other one (i.e. the originally more cooperative alternative becoming more competitive). If the step was larger then it could take one of the following actions:

1. Reduce the step to half. If the difference between the two alternatives was smaller than half of the step then it would move to algorithms (2) or (3), as described below

2. Change the values of the unchosen alternative by adding (or subtracting, depending on the direction of the change) 5 points to self- and other-values. If this also resulted in allocations ‘crossing’ each other then the algorithm (3) was implemented.

3. Change the values of the unchosen alternative by adding (or subtracting, depending on the direction of the change) 2 points to self- and other-values.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Notice that the algorithm could choose to directly implement (2) without implementing (1); also (3) could be directly implemented, without the need of implementing (1) and (2).

Final allocation—On each trial, the algorithm checked whether it should stop. If both the difference between the two ‘self’ values and the difference between the two ‘other’ values were smaller than 6 points, then the algorithm stopped. The average of the two angles associated with the final two allocations for each assessment were used as the measure of SVO.

This sequential testing algorithm yielded social value orientations measurements with strong test-retest reliability. Participants made choices in three separate assessments, and each version had different initial allocations (see description above). Estimated angles across the assessments showed strong test-retest reliability (mean R^2 = .54). To exclude the minority of participants with unreliable measures of SVO, any participant with a difference between SVO measurements greater than 15° was excluded. Of the 90 participants assessed, ten were excluded based on this criterion. The test-retest reliability of the remaining participants was high (mean R^2 = .92). The boundary criteria for the three groups (cooperative, individualists and competitive) were defined as the mean SVO of the sample plus (cooperative threshold) or minus (competitive threshold) a half standard deviation of the sample (yielding 10.16 and −9.57 as boundaries). The average SVO measures of the included participants are depicted in SM Fig. 2, and individual SVO estimates for each individual are reported in SM Table 2.

3. Social Value Learning task (fMRI)

Prior to scanning, all participants received instructions about the mechanics of the tasks, and it was explained to each participant that they would be making choices that would form the basis of their own payment and the payment of a second person. The participant was informed that: (i) s/he would never meet the other person or know each other’s identity; (ii) the other person would be paid according to the outcomes of a randomly chosen subset of decisions; and, (iii) the other person would not be making similar decisions for the participant.

The scanning session was separated into 6 blocks. Each block consisted of 30 trials in which participants chose between two squares depicting fractals. Six different fractals were used, so players learned values associated with each fractal a single time, and fractals were randomly positioned to the left or right of a fixation cross. The outcomes associated with each fractal were randomly determined for each participant. Within each block, one fractal was associated with one allocation with 80% probability and a second allocation with 20% probability. The other fractal was associated with the first allocation with 20% probability and the second outcome with 80% probability. The probabilities on each trial were pseudorandom, so that every 10 trials included 2 less probable outcomes.

The allocations for each block were as follows, where ε is a uniform discrete distribution with mean 0 and range of 10:

- [Self: −70 + ε, Other: −70 + ε] vs. [Self: −70 + ε, Other: +70 + ε]
- [Self: −70 + ε, Other: −70 + ε] vs. [Self: +70 + ε, Other: −70 + ε]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- [Self: −70 + ε, Other: −70 + ε] vs. [Self: +70 + ε, Other: +70 + ε]
- [Self: −70 + ε, Other: +70 + ε] vs. [Self:+70 + ε, Other: −70 + ε]
- [Self: −70 + ε, Other: +70 + ε] vs. [Self: +70 + ε, Other: +70 + ε]
- [Self: +70 + ε, Other: −70 + ε] vs. [Self: +70 + ε, Other: +70 + ε]

Each trial began with a 1 second fixation cross on a black background, plus a value derived from an exponential distribution with mean of 1 second, truncated at 6 seconds. On the next screen two fractals appeared on the left and right of a fixation cross that subtended 10 degrees of visual field. Participants were required to respond within 3 seconds by choosing a left or right button. The chosen stimulus was framed by a white square for 0.5 second plus a value derived from an exponential distribution with mean of 1 second, truncated at 6 seconds. Then the outcome allocation (i.e., outcome for self and outcome for other) was displayed for 2 seconds (Fig. 1). The order in which blocks were presented was pseudorandomized across participants.

4. Computational modeling

We monitored decision-related hemodynamic activity with functional magnetic resonance imaging, and subsequently modeled these data using two prediction error regressors generated by fitting participant choices to a reinforcement learning model of reward for oneself and a social partner (Fig. 1b). Within this hybrid model, the “self expected value” (EVS) and “other expected value” (EVO) are updated on a trial-by-trial basis through separate prediction errors (PES and PEO, respectively), and PEO is transformed to reflect the competitive or cooperative preference of a decision-maker. The transformation of PEO (Fig. 1b) is achieved by weighting the monetary outcome received by the other person by a γ parameter. For instance, if the other person receives \$70, then, for a competitive person (γ= −1) the algorithm will behave as if the outcome is negative (γ ×\$70 = −\$70); on the contrary, for a cooperative person (γ=1) a positive outcome will produce a positive prediction error. This simple formulation reinforces cooperative and competitive actions for cooperative and competitive subjects, respectively.

Differences between expected and experienced outcomes were modelled using a modified standard Q-learning algorithm described by Sutton & Barto (1998) and implemented in a similar instrumental probabilistic learning task by Pessiglione et al., (2006). As illustrated in Fig. 1b, at each decision outcome, the algorithm computes (i) an expected value of the stimulus chosen for the outcome received by the subject (EV_S) and (ii) a second expected value for the outcome received by the other subject (EV_O). EV_S is updated by the usual rule $EV_{S,t} = EV_{S,t-1} + \alpha_S(PE_{S,t})$, where α_S represents a learning rate parameter, $PE_{S,t}$ represents a prediction error defined as $V_{S,t} - EV_{S,t}$ and $V_{S,t}$ is the reward received by the subject at time t. Similarly, EV_O is updated by the rule $EV_{O,t} = EV_{O,t-1} + \alpha_O(PE_{O,t})$, where α_O again represents a learning rate parameter, $PE_{O,t}$ represents a prediction error defined as $\gamma(V_{O,t})$ $EV_{O,t}$ and $V_{O,t}$ is the reward received by the subject at time t, while γ takes the values of 1 or −1 in order to allow for both cooperative and competitive orientations. Thus, if γ=−1, then a positive outcome for the other person is subjectively perceived as a negative outcome for the decision maker. The probability of choosing one stimulus over another is estimated by

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

the softmax rule. For example, the probability of choosing stimulus A is estimated as Pa(t) = $\exp(EV_{NET,A}(t)/\beta)/(\exp(EV_{NET,A}(t)/\beta) + \exp(EV_{NET,B}(t)/\beta))$, where $EV_{NET,A} = EV_{S,A} + EV_{O,A}$ and $EV_{NET,B} = EV_{S,B} + EV_{O,B}$ and β is a temperature parameter.

We estimated gamma (γ) for each subject by focusing on the two conditions in which outcomes delivered to the other varied, while outcomes delivered to self were kept constant (conditions [self −70 / other −70 ; self −70 / other +70] & [self +70 / other −70 ; self +70 / other +70] illustrated in Figures 2c and 2d). For that reason, the estimation procedure followed two steps: in the first step we estimated learning rates ($\alpha_{S,}$ α_O), gamma (γ) and temperature (β) for each subjectin these two conditions only. The goal of this step was to extract the γ that best fit these two conditions. Subsequently, the estimated γ value was used as a constant for estimations in all other conditions; all other parameters were estimated within each block in order to maximize the likelihood of the model choosing participants' actual choices.

Allowing γ to take the values of −1 or +1 enabled the model to account for the social preference of individual participants. For example, when a competitive individual chose a stimulus that resulted in an unexpectedly positive outcome for their social partner, negative values of γ mean that the positive reward for the social partner translates into a negative update of the Q-value associated with the chosen stimulus. Average learning rates ($\alpha_{S,}$ α_O), temperature (β), mean negative log likelihood and pseudo-R^2 (Camerer and Ho, 1998) for each subgroup and condition are reported in SM Table 1. Pseudo-R^2 is defined as $(r-l)/r$, where *r* is the log likelihood of a model choosing randomly and *l* is the log likelihood of our model. The estimated pseudo-R^2 assessing model fit were comparable to previous reports modelling choice behavior (mean pseudo-R^2 for our model: .32; (Daw et al., 2006): .31; (Li and Daw, 2011): .35; (Rutledge et al., 2009): .18; (Simon and Daw, 2011): .28).

We also estimated the Akaike Information Criterion (AIC) for alternative models discussed below. We employed the following formula to estimate the AIC: $AIC = 2k-2Log(L)$, where k is the number of parameters and L is the log likelihood of each model, estimated across all trials and subjects. Lower AIC values indicate a better fit of the observed behavior. The number of parameters for each model is described in the Results section.

5. Results

5.1. Neural prediction error signal for self-value

To identify activity underlying prediction errors associated with outcomes for oneself (PE_S), we focused on four experimental conditions in which self-value differed for outcomes associated with A and B. In the first manipulation (Fig. 2a), choosing one stimulus resulted in *gains for oneself* (+70 points, ±5) and for the other participant (+70 points, ±5) with 80% probability, whereas it yielded *loss for oneself* (−70 points, ±5) and gain for the other participant (+70 points, ±5) with 20% probability. The alternative stimulus yielded the same allocations, but with the probabilities reversed (i.e., self-gain, other-gain with 20% probability and self-loss, other-gain with 80% probability). Note that both stimuli resulted in a positive outcome (+70) for the other participant. Thus, this first experimental manipulation is similar to previous studies of value-learning for oneself, as actions are guided by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

differences in value for the decision-maker only. Similarly, in a second condition (Fig. 2b) possible outcomes were either (+70 self / −70 other) or (−70 self / −70 other), thus allowing self outcomes to primarily guide learning. In two additional conditions (Fig. 2e and 2f), possible outcomes differed for oneself, as well as for the social partner.

When PE_S were parametrically regressed to the hemodynamic activity at the outcome of each decision, correlated neural activity was found in bilateral ventral striatum across all conditions (Fig. 3a, upper panel; coordinates for PE_S across all conditions combined: 4,10,−4, $P < .01$, FWE whole-brain-corrected), consistent with previous studies of self-interested probabilistic learning (Hare et al., 2008; Pessiglione et al., 2006). The robustness of PE_S-related activity in the striatum across conditions was confirmed by separately examining the statistical significance at peak voxels in ventral striatum within each condition separately (Fig. 3a, lower panel). This analysis confirmed significant PE_S-related activity in striatum for each condition.

5.2. Preference-dependent prediction error learning of other-value

Prediction errors for outcomes to others (PE_O) were modeled using a similar reinforcement learning algorithm as PE_S above, substituting other-value for self-value (Fig. 1b; see Supplementary Material). However, to incorporate cooperative and competitive orientation toward social partners, our algorithm weighted other-value by a parameter, γ, that took values of +1 or −1. The right panel of Fig. 1b illustrates the effect of the γ parameter. Consider the perspective of a decision-maker when a choice results in a social partner receiving +70, an amount that exceeds the expectations of the decision-maker. A cooperative individual, with $\gamma = +1$, would subjectively experience the result to be an unexpectedly good outcome, corresponding to a positive prediction error that updates EV_O to be more positive. In contrast, a competitive individual, with $\gamma = -1$, would subjectively experience the result as an unexpectedly bad outcome, corresponding to a negative prediction error that updates EV_O to be more negative. In this way, inclusion of the γ parameter enables the model to incorporate preference-dependent prediction errors, and these prediction errors update EV_O to reflect the subjective perspective of the decision-maker (not the social partner).

5.3. Support for the model

As it is the case for all computational approaches trying to explain neurobehavioral data, the space of possible models is very large. Thus, there is the possibility that an alternative model that either we have not considered or the present data do not easily accommodate could explain the underlying process in a better way. Here, we provide evidence supporting the present model in three ways: (i) we employ an external measure of social preference to provide external validation for the included γ parameter of our model; (ii) we relate a model-free estimate of social preference within our task to the included γ parameter estimates, (iii) we compare goodness-of-fits metrics of the present model to corresponding metrics of a number of alternative models described below.

5.3.1. External measures—External validity for the fitted γ parameter values was established through a non-learning choice task designed to assess social value orientation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(SVO; (McClintock and Liebrand, 1988), (Kelly and Stahelski, 1970; Kuhlman and Marshello, 1975; Sattler and Kerr, 1991; Van Lange and Kuhlman, 1994)). In this task, participants completed a series of psychophysics-based, adaptively updated dictator games (SM Fig. 1; SM Table 1). This procedure yields an estimate of the extent to which social agents prefer allocations that maximize the sum of the self and other outcomes ("Cooperative"), prefer allocations that maximize the signed difference between self and other outcomes ("Competitive"), or are indifferent to the outcomes of others ("Individualistic"), seeking to maximize only their own outcomes. The correspondence between the SVO metric and fitted values of γ are illustrated in Fig. 4a. Values of γ for SVO-determined competitive individuals were negative, values of γ for SVO-determined cooperative individuals were positive, and SVO-determined individualists had values of γ that did not differ from zero on average. Note that while individual estimates of γ take values of either −1 or 1, the average of the group of SVO-determined individualists did not differ from zero.

5.3.2. Model-free estimations—To confirm that fitted values of γ reflect social orientation expressed within the learning task, binary logistic regressions for each subject s data were estimated using self- and other-outcomes in the previous trial as predictors and choice ('stay' = 1; switch = 0) as the dependent variable. Beta coefficients associated with other-outcomes are plotted against average fitted values of γ in Fig. 4b. Individuals who were likely to 'switch' following a positive other-outcome in the logistic regression were also estimated to have negative values of γ in the hybrid model estimation, individuals who were likely to 'stay' following a positive other-outcome in the model-free analysis had positive values of γ, and individuals who were as likely to 'stay' as they were to 'switch' regardless of the other-outcome had values of γ that did not differ from zero on average.

5.3.3. Alternative models—The goodness-of-fit of the current model was also compared with a number of alternative models described below:

- Alternative model (i). A first possibility is to allow γ to freely vary between −1 and 1. We tested this model and we found that mean Akaike Information Criterion (AIC) values to be lower in that case (original model: 531.08 [(19 free parameters: 6 conditions ×3 parameters (α Self (learning parameter for self), α Other (learning parameter for other), β (temperature)) + one γ (social orientation weighting parameter or −1 or +1)).] vs. 533.5 (alternative model) (19 free parameters: 6 conditions ×3 parameters + one γ varying in the [−1, 1]). On a first reading the two AIC values seem to be close enough but we have to consider that the alternative model allows for γ to take all possible values (as compared to the original model which allows only two values). AIC is unable to capture this variability as it is sensitive only to the number of parameters and not the values they are allowed to take. Therefore, by *lex parsimoniae*, i.e. that models that recruit shorter computations are assigned higher probabilities and therefore preferred (Gauch, 2003) the judgment is in favor of the original model.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- Alternative model (ii) (18 free parameters: 6 conditions ×3 parameters). To assess the impact of γ on the behavioral fit of the model, AIC values were also compared for models with and without γ.The AIC for the model without γ was 558.6.

- Alternative model (iii) (13 free parameters: 6 conditions ×2 parameters + one γ) Another possibility is that instead of participants computing each value (self and other) separately, they actually compute a weighted combination of the self and other outcome values in a single step (i.e. there is one prediction error based on updating the value of the bundle: Vtot = Vself + γVother). This model also has higher AIC values (562.40 as compared to 531.5) and thus we believe that it is not appropriate for the present data.

- Alternative model (iv). (12 free parameters: 6 conditions ×2 parameters). Another possibility is that participants altogether ignore the value offered to the other person. This can be modeled by setting $\gamma = 0$ for the original model, which means that the "other" values are not used in the computation of the value of (and the associated probabilities of choosing) each stimulus. This model also has higher AIC values (535.24).

- Alternative model (v). (13 free parameters: 6 conditions ×2 parameters + one γ) Another possibility is that participants have the same learning α parameter for self and other values. This model also has higher AIC values (563.37).

5.4. Neural prediction error signal for other-value

To examine how decision-makers learn how their choices map onto outcomes for social partners, we focused on four manipulations in which other-value differed for outcomes A and B (Fig. 2c, 2d, 2e, 2f). In one of these manipulations (Fig. 2c), choosing one stimulus resulted in gain for oneself (+70) and gain for the other participant (+70) with P = 80%, and, with P = 20%, it yielded gain for oneself (+70) and loss for the other participant (−70). Again, the alternative stimulus yielded the same outcomes, but the probabilities were reversed. In another manipulation (Fig. 2d), outcomes to social partners were the same as described above, but outcomes to oneself were losses (−70) rather than gains. In two additional manipulations (Fig. 2e and 2f), both self-value and other-value varied across the two options, allowing the examination of both PE_S and PE_O simultaneously.

Preference-dependent PE_O were calculated as described above (Fig. 1b), and parametrically regressed to hemodynamic activity at the outcome of each decision. Strikingly, this analysis identified a strong correlate of PE_O in medial prefrontal cortex (MPFC), indicating that hemodynamic activity in this region reflects the updating of value expectations for others in a manner that takes into account the social preference of the decision-maker (Fig. 3b, upper panel; coordinates for PE_O across all conditions combined: 10,54,0, $P < .01$, FWE whole brain corrected). The robustness of PE_O-related activity in MPFC across conditions was further tested by separately examining statistical significance at peak voxels in MPFC within each condition separately (Fig. 3b, lower panel). This analysis confirmed significant PE_O-related activity in MPFC for each condition.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To further elucidate the PE_O signal we explored the MPFC response as a function of the valence of the PE_O and the SVO (Fig. 4c). To facilitate this analysis, we multiplied the PE_O-related hemodynamic response by each decision-maker's γ, thus transforming the PE_O signal to the perspective of the 'other' social partner. In doing so, the positive PE_O of the cooperative decision-maker (better than expected from the perspective of 'other') becomes comparable to the negative PE_O of the competitive decision-maker (again, better than expected from the perspective of 'other'). We subsequently subtracted hemodynamic responses of trials associated with unexpectedly positive outcomes to 'other' from hemodynamic responses of trials associated with unexpectedly negative outcomes to 'other', and plotted the resulting difference by SVO (Fig. 4c). The resulting heatmap reveals that cooperative participants have higher MPFC response when they experience an unexpected negative outcome for the other person (see also Table 1). This difference is gradually reversed as SVO decreases: competitive participants have higher MPFC response for unexpectedly positive outcomes for the social partner. Notice that for SVO close to zero, the difference between the two signals is negligible. Taken together, these data indicate that MPFC signals outcomes that are preference-incongruent.

An alternative hypothesis for the apparent indifference of individualists to PE_O is that individualists may be modulating their behavior depending on whether they themselves receive positive or negative outcomes. For example, it could be the case that an individualist seeks to minimize envy and guilt by both seeking negative outcomes for others when receiving negative outcomes for oneself (condition S−/O− vs. S−/O+) and seeking positive outcomes for others when receiving positive outcomes for oneself (condition S+/O− vs. S +/O+). To explore this possibility, we examined whether individualists remain indifferent to the outcome of others regardless of their own outcomes, or whether orientation towards others is dependent on self-value. Behaviorally, we found that SVO-defined individualists exhibited no preference for positive or negative outcomes to others, either when consistently receiving positive self-values (null hypothesis that individuals choose S+/O+ and S+/O− in Fig. 2a with equal frequency; p =.34, t=.97, df=33) or when consistently receiving negative self-values (null hypothesis that individuals choose S−/O− and S−/O+ in Fig. 2b with equal frequency: p = .82; t = .23, df =33). Similarly, we found no evidence that individualists respond differentially to PE_O in MPFC as a function of valence for either of these conditions (S+/O+ vs. S+/O−; S−/O+ vs. S−/O−). That is, the prediction error responses following either positive value outcomes to other and negative value outcomes to other did not differ in either condition (SM Fig. 4; SM Tables 4 & 5).

Finally, to quantify the suggested neural dissociation of PE_S and the preference-incongruent PE_O, we examined neural correlates of PE_S and PE_O in both striatum and MPFC (Fig. 3c). A repeated-measures 2×2 ANOVA revealed a significant interaction for REGION and PE (F_{71}=9.3; P < .005), confirming that the striatum and MPFC preferentially encode PE_S and PE_O, respectively.

6. Discussion

Across four experimental conditions, our results provide strong evidence that prediction error learning signals in MPFC are used to update value for social partners, in a way that is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

topographically and functionally distinct from prediction error signals used to update value for oneself. Crucially, the PE_O signal is strongly associated with individual differences in how decision-makers prefer to divide resources with social partners in environments that do not require learning. That is, while value learning for others was well characterized by the same reinforcement learning process that guides non-social reward learning, the direction and magnitude of the neural signal was strongly determined by the social goals of the decision-maker: cooperative agents show increased MPFC activity in response to 'negative' outcomes for social partners, while competitive agents show increased MPFC activity in response to good outcomes for social partners.

The localization of PE_O-related activity across the four tasks (maximal at 10,54,0) was found to fall within anterior rostral medial prefrontal cortex (arMPC), and previous work suggests a critical role of arMPC in modeling mental states of other people, including how actions impact social partners (Amodio and Frith, 2006; Bzdok et al., 2013; Coricelli and Nagel, 2009; Decety and Sommerville, 2003; Gallagher and Frith, 2003; Hare et al., 2010; McCabe et al., 2001; Van Overwalle and Baetens, 2009; Winston et al., 2002; Zaki and Mitchell, 2011). Recent studies have implicated MPFC in goal-directed choices that involve social agents (Behrens et al., 2008; Hampton et al., 2006; Nicolle et al., 2012; Suzuki et al., 2012; Yoshida et al., 2010), whereas other studies have focused on the role of striatum in evaluating rewards delivered to others (Harbaugh et al., 2007; Hsu et al., 2008; Suzuki et al., 2012). The current results are consistent with (Suzuki et al., 2012), suggesting that MPFC is recruited while learning to simulate the choices of others in addition to learning for oneself.

Another study (Apps et al., 2013) reported that anterior cingulate cortex activity correlated with prediction error signals when monitoring (expected or unexpected) outcomes delivered to a second person. In contrast to the present study, participants in Apps et al. were not responsible for the action leading to the outcome of the second person. Taken together, these studies potentially suggest that other-value prediction error signals are computed separably for outcomes for which one is responsible, and outcomes for which one's decision does not determine the outcome, and that only the former is sensitive to the motivational orientation of the decision-maker.

Individual differences in the direction and magnitude of the PE_O identified here highlight the importance of polymorphic social orientation for neurocomputational models of social decision-making ((Kuhlman and Marshello, 1975; McClintock and Liebrand, 1988; Van Lange and Kuhlman, 1994), (Bowles and Gintis, 2004; Fehr and Fischbacher, 2002; Kuhlman and Wimberley, 1976; Kurzban and Houser, 2005)). The current data indicate that social orientation strongly modifies learning-related value-representations for others when making decisions that impact others. Strikingly, in additional analyses we find no evidence that other-value is represented independent of the social preference of the decision-maker during social learning.

The present design does not exhaustively examine how Vs and Vo are integrated in the human brain. For example, alternative models have been suggested (see for instance Van Lange (1999)) that include fairness or (in)equity considerations. Such models have been constructed to predict behaviour across a wider space of allocations, where the self- and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

other- values vary considerably. For the allocations used here, where the Self or Other values are allowed to vary within a limited space, the alternative models would make very similar predictions. Yet, all previous models assume that Vo is transformed (usually by multiplying with a constant, as in our case) and this transformation represents the social orientation of the agent. Thus, here, we identify how the outcome received by the other is weighted using a typical PEST procedure; we find that the weighting is also employed in the learning process and corresponds to the external metrics (PEST). Finally, our data suggest that MPFC responses produce prediction error signals that mirror this weighting.

It could be suggested that the brain responses identified here, especially the MPFC PE_O signal, reflect a non-social perceptual or learning process, and further work will be required to test this possibility. For example, the MPFC signal could reflect updating of numeric values more generally, rather than the value-based outcome to a social partner. While the current design does not eliminate this possibility, the current data demonstrate this signal to be systematically related to the social preferences of participants. Thus, if the MPFC signal indeed reflected more general learning process, it nevertheless appears to be employed to guide decisions in a social preference-dependent manner.

While the current study focuses on learning when making decisions impacting others, related work has investigated how information from others impacts self-interested choices and associated outcomes. Behrens and colleagues (Behrens et al., 2008) and Burke and colleagues (Burke et al., 2010) found that social information and personal experience are combined to influence self-interested decisions and associated outcomes through separable learning signals. In contrast with this work, however, the PE_O signal identified in our present work is used to preferentially represent social-orientation weighted outcomes for others; critically, this information is used to reinforce the values of actions, akin to prototypic a-social reinforcement learning paradigms. To our knowledge, this is the first study to isolate such a signal. It is therefore expected that further studies will be required to elucidate the exact nature of the preference-dependent PE_O signal, as well as its relationship to PE_S. Indeed, initial accounts of the classic PE_S signal were similarly and necessarily incomplete, and led to the development of a broad field of neurocomputational signals underlying reward-guided learning for oneself. We expect that the present data will open a variety of research questions, including those examining the precise spatio-temporal and computational properties of this signal, possible alterations following pharmacological manipulations, disease or stress, the impact of various social norms on its amplitude, and alterations in the neurocomputational signal in more complex social contexts.

It is noteworthy that the neural learning signals identified here independently encode positive and negative reward prediction errors for ‘self’ and ‘other’ outcomes, respectively. In the case of reward for oneself, the receipt of an unexpectedly positive outcome leads to an increase in the value associated with the reward-predicting cue, and activity in ventral striatum mirrors this increase. However, in the case of rewards for others, the receipt of unexpected preference-congruent outcomes (i.e., better than expected), leads to decreased activity in MPFC activity (Fig. 4 and Table 1). The functional significance of learning self- and other-value through opponently-valenced learning signals here is unclear. One possibility is that decision processes seeking to satisfy multiple goals are hierarchically

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

updated, such that action-value pairs associated with primary goals (e.g., maximizing self-value) are updated through positive PE, while updating based on secondary goals (e.g., maximizing value for others) are updated through negative PE. Thus, cooperative individuals may primarily seek and update good outcomes for themselves and avoid and update bad outcomes for others second; while competitive individuals again primarily seek and update good outcomes for themselves first, avoiding good outcomes for others second. While this speculation cannot be confirmed within the current task, it suggests a number of questions regarding how multiple channels of information are combined and reconciled when making a social decision and how such outcomes are simultaneously updated.

4.1. Conclusions

Taken together, these data reveal neural computations fundamental to learning in social environments, where it is critical to take into account the impact of our actions to others. A significant conclusion of the present study is the necessity to incorporate the often ignored fact that populations and their social orientations are polymorphic – thus individual differences might exist and actually mask interesting phenomena. The present results bring together the tradition of reinforcement learning, which examines how humans adapt in dynamic environments, and behavioral game theory, where agents take actions that impact others. We choose to study a basic question: how outcomes are evaluated without the presence of any strategic component. The results suggest that monetary values are early on transformed. This conclusion informs the study of more complex strategic interactions as it implies that the game matrixes (for instance in a prisoner s dilemma game) are transformed to the so-called “effective matrixes” (where outcomes are weighted by social orientation). Finally, we believe that the present study, along with other similar studies (Decety and Lamm, 2007; Kishida and Montague, 2012; Wolpert et al., 2003) sets the basis for the further development of the field of social computational neuroscience, where social actions can be formally described by computational models and neurobiological mechanisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the National Institutes of Health (DA036017) and Office of Research and Development, Rehabilitation Research and Development Service, Department of Veterans Affairs (D7030R). We thank P. Chiu, S. LaConte, T. Lohrenz, and P.R. Montague for comments on this manuscript. G.C. and B.K.C. jointly designed the study; G.C. implemented the design and collected behavioral and imaging data; G.C. and B.K.C. jointly conducted the analysis; G.C. and B.K.C. jointly prepared the manuscript. The authors declare no competing financial interests.

References

Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and so cial cognition. Nat Rev Neurosci. 2006; 7:268–277. [PubMed: 16552413]

Apps MA, Green R, Ramnani N. Reinforcement learning signals in the anterior cingulate cortex code for others’ false beliefs. Neuroimage. 2013; 64:1–9. [PubMed: 22982355]

Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 2005; 47:129–141. [PubMed: 15996553]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF. Associative learning of social value. Nature. 2008; 456:245–249. [PubMed: 19005555]

Bowles S, Gintis H. The evolution of strong reciprocity: cooperation in heterogeneous populations. Theoretical population biology. 2004; 65:17–28. [PubMed: 14642341]

Burke CJ, Tobler PN, Baddeley M, Schultz W. Neural mechanisms of observational learning. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:14431–14436. [PubMed: 20660717]

Bzdok D, Langner R, Schilbach L, Engemann DA, Laird AR, Fox PT, Eickhoff SB. Segregation of the human medial prefrontal cortex in social cognition. Front Hum Neurosci. 2013; 7:232. [PubMed: 23755001]

Camerer C, Ho TH. Experience-Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-Variation. Journal of mathematical psychology. 1998; 42:305–326. [PubMed: 9710553]

Chang LJ, Smith A, Dufwenberg M, Sanfey AG. Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. Neuron. 2011; 70:560–572. [PubMed: 21555080]

Coricelli G, Nagel R. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:9163–9168. [PubMed: 19470476]

Crockett MJ, Clark L, Hauser MD, Robbins TW. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:17433–17438. [PubMed: 20876101]

D'Ardenne K, McClure SM, Nystrom LE, Cohen JD. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science. 2008; 319:1264–1267. [PubMed: 18309087]

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

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441:876–879. [PubMed: 16778890]

Decety J, Lamm C. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist. 2007; 13:580–593. [PubMed: 17911216]

Decety J, Sommerville JA. Shared representations between self and other: a social cognitive neuroscience view. Trends in cognitive sciences. 2003; 7:527–533. [PubMed: 14643368]

Delgado MR, Li J, Schiller D, Phelps EA. The role of the striatum in aversive learning and aversive prediction errors. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2008; 363:3787–3800. [PubMed: 18829426]

den Ouden HE, Daunizeau J, Roiser J, Friston KJ, Stephan KE. Striatal prediction error modulates cortical coupling. J Neurosci. 2010; 30:3210–3219. [PubMed: 20203180]

Fehr E, Fischbacher U. Why social preferences matter - The impact of non-selfish motives on competition, cooperation and incentives. Economic Journal. 2002; 112:C1–C33.

Fehr, E.; Krajbich, I. Social Preferences and the Brain. In: Glimcher, PW.; Fehr, E., editors. Neuroeconomics. 2. Vol. Chapter 11. Academic Press; San Diego: 2014. p. 193-218.

Fliessbach K, Weber B, Trautner P, Dohmen T, Sunde U, Elger CE, Falk A. Social comparison affects reward-related brain activity in the human ventral striatum. Science. 2007; 318:1305–1308. [PubMed: 18033886]

Gallagher HL, Frith CD. Functional imaging of 'theory of mind'. Trends in cognitive sciences. 2003; 7:77–83. [PubMed: 12584026]

Galvan A, Hare TA, Davidson M, Spicer J, Glover G, Casey BJ. The role of ventral frontostriatal circuitry in reward-based learning in humans. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2005; 25:8650–8656. [PubMed: 16177032]

Gauch, HG. Scientific Method in Practice. Cambridge University Press; Cambridge: 2003.

Hampton AN, Bossaerts P, O'Doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2006; 26:8360–8367. [PubMed: 16899731]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Harbaugh WT, Mayr U, Burghart DR. Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science. 2007; 316:1622–1625. [PubMed: 17569866]

Hare TA, Camerer CF, Knoepfle DT, Rangel A. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2010; 30:583–590. [PubMed: 20071521]

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2008; 28:5623–5630. [PubMed: 18509023]

Haruno M, Frith CD. Activity in the amygdala elicited by unfair divisions predicts social value orientation. Nature neuroscience. 2010; 13:160–161. [PubMed: 20023652]

Hsu M, Anen C, Quartz SR. The right and the good: distributive justice and neural encoding of equity and efficiency. Science. 2008; 320:1092–1095. [PubMed: 18467558]

Jocham G, Klein TA, Ullsperger M. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. J Neurosci. 2011; 31:1606–1613. [PubMed: 21289169]

Kelly HH, Stahelski AJ. Social interaction basis ofcooperators' and competitors' beliefs about others. Journal of Personality and Social Psychology. 1970; 16

Kishida KT, Montague PR. Imaging models of valuation during social interaction in humans. Biol Psychiatry. 2012; 72:93–100. [PubMed: 22507699]

Krugel LK, Biele G, Mohr PNC, Li SC, Heekeren HR. Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proceedings of the National Academy of Sciences. 2009; 106:17951–17956.

Kuhlman DM, Marshello AF. Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. Journal of Personality and Social Psychology. 1975; 32:922–931. [PubMed: 1185519]

Kuhlman DM, Wimberley DL. Expectations of Choice Behavior Held by Cooperators, Competitors, and Individualists across 4 Classes of Experimental Game. Journal of Personality and Social Psychology. 1976; 34:69–81.

Kurzban R, Houser D. Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:1803–1807. [PubMed: 15665099]

Li J, Daw ND. Signals in human striatum are appropriate for policy update rather than value prediction. J Neurosci. 2011; 31:5504–5511. [PubMed: 21471387]

Lurie S. A parametric model of utility for two-person distributions. Psychological Review. 1987; 94:42–60.

McCabe K, Houser D, Ryan L, Smith V, Trouard T. A functional imaging study of cooperation in two-person reciprocal exchange. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:11832–11835. [PubMed: 11562505]

McClintock G, Liebrand W. Role of interdependence structure, individual value orientation, and another's strategy in social decision making: A transformational analysis. Journal of Personality and Social Psychology. 1988; 55:396–409.

Montague PR, King-Casas B, Cohen JD. Imaging valuation models in human choice. Annual review of neuroscience. 2006; 29:417–448.

Murphy RO, Ackermann KA. Social value orientation: theoretical and measurement issues in the study of social preferences. Pers Soc Psychol Rev. 2014; 18:13–41. [PubMed: 24065346]

Nicolle A, Klein-Flugge MC, Hunt LT, Vlaev I, Dolan RJ, Behrens TE. An agent independent axis for executed and modeled choice in medial prefrontal cortex. Neuron. 2012; 75:1114–1121. [PubMed: 22998878]

O'Connell G, Christakou A, Haffey AT, Chakrabarti B. The role of empathy in choosing rewards from another's perspective. Front Hum Neurosci. 2013; 7:174. [PubMed: 23734112]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature. 2006; 442:1042–1045. [PubMed: 16929307]

Ramnani N, Elliott R, Athwal BS, Passingham RE. Prediction error for free monetary reward in the human prefrontal cortex. Neuroimage. 2004; 23:777–786. [PubMed: 15528079]

Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW. Dopaminergic drugs modulate learning rates and perseveration in Parkinson’s patients in a dynamic foraging task. J Neurosci. 2009; 29:15104–15114. [PubMed: 19955362]

Sattler DN, Kerr NL. Might Versus Morality Explored - Motivational and Cognitive Bases for Social Motives. Journal of Personality and Social Psychology. 1991; 60:756–765.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

Seymour B, O’Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS. Temporal difference models describe higher-order learning in humans. Nature. 2004; 429:664–667. [PubMed: 15190354]

Simon DA, Daw ND. Neural correlates of forward planning in a spatial decision task in humans. J Neurosci. 2011; 31:5526–5539. [PubMed: 21471389]

Sutton, R.; Barto, A. Reinforcement Learning: An Introduction. MIT Press; Cambridge: 1988.

Suzuki S, Harasawa N, Ueno K, Gardner JL, Ichinohe N, Haruno M, Cheng K, Nakahara H. Learning to simulate others’ decisions. Neuron. 2012; 74:1125–1137. [PubMed: 22726841]

Tricomi E, Rangel A, Camerer CF, O’Doherty JP. Neural evidence for inequality-averse social preferences. Nature. 2010; 463:1089–1091. [PubMed: 20182511]

Van Lange PAM, Kuhlman DM. Social value orientations and impressions of partner’s honesty and intelligence: A test of the might versus morality effect. Journal of Personality and Social Psychology. 1994; 67:126–141.

Van Overwalle F, Baetens K. Understanding others’ actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage. 2009; 48:564–584. [PubMed: 19524046]

Winston JS, Strange BA, O’Doherty J, Dolan RJ. Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nature neuroscience. 2002; 5:277–283. [PubMed: 11850635]

Wolpert DM, Doya K, Kawato M. A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci. 2003; 358:593–602. [PubMed: 12689384]

Yoshida W, Seymour B, Friston KJ, Dolan RJ. Neural mechanisms of belief inference during cooperative games. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2010; 30:10744–10751. [PubMed: 20702705]

Zaki J, Mitchell JP. Equitable decision making is associated with neural markers of intrinsic value. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:19761–19766. [PubMed: 22106300]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- Social actors must learn how their actions impact other people (‘others’)
- Prediction error (PE) signals of outcomes for others are found in medial prefrontal cortex
- Size and direction of these social learning signals depend on social preferences of the actors (competitive or cooperative)
- PE signals for others are distinct from PE signals tracking value for oneself

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

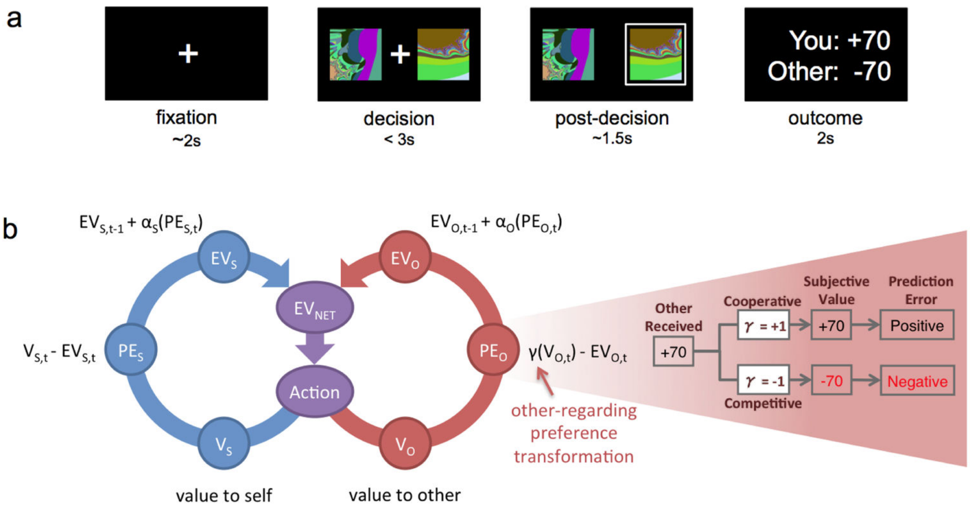

Fig. 1. Social Value Learning task and learning model

(a) Each trial began with a fixation cross (~2 secs) indicating the onset of a new trial. Two unique fractal stimuli representing two decision options were displayed until a decision was submitted by keypress (limited to 3 secs). The chosen stimulus was subsequently framed for ~1.5 secs, after which the outcomes for the decision-maker and the social partner were revealed for 2 secs. In each of six conditions, each participant made 30 choices between two options associated with probabilistic gain (or loss) for the decision-maker, as well as probabilistic gain (or loss) for a social partner.

(b) Hybrid learning model of self-value and preference-dependent other-value. Choices produced an outcome for the actor and a different outcome for the social partner simultaneously. Following typical reinforcement learning algorithms, rewards received for self (blue circle) update the expected value (EV) of a choice at time t via prediction errors (PE) weighted by a learning rate (α). Rewards that are delivered to the social partner (red circle - ‘other’) are also updated by the same mechanism with the difference that the value is subjectively transformed (pink inset) according to social preferences, represented by the γ coefficient. For example, a competitive orientation will transform a positive outcome to a negative subjective value, thus producing a negative prediction error.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

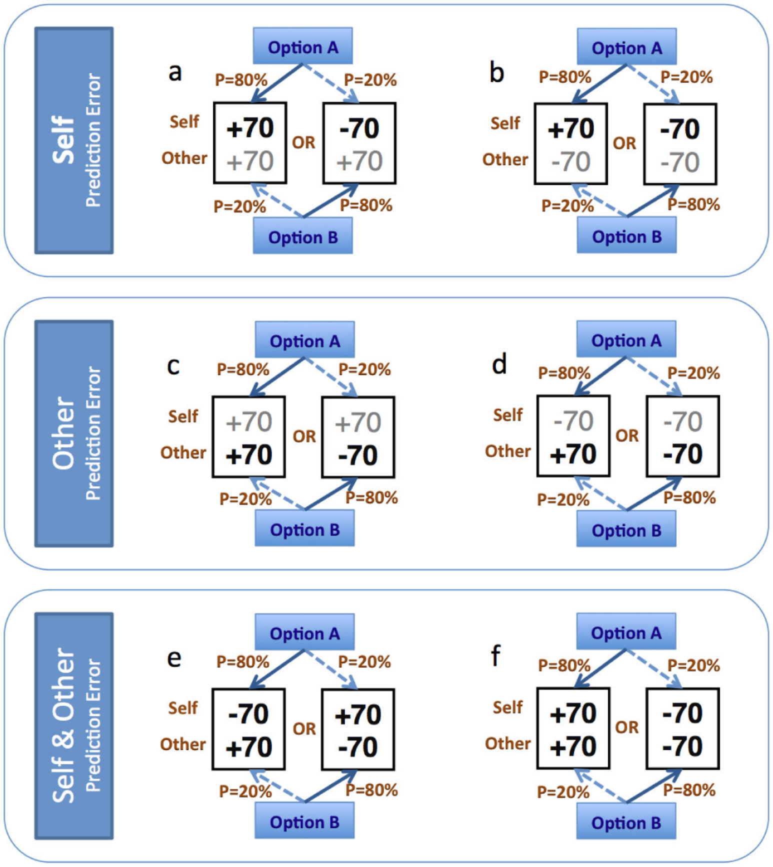

Fig. 2. Conditions of the experiment testing for ‘Self’ and ‘Other’ Prediction Error

In each condition participants learn by trial and error the following contingencies: (a) Self PE (1): Option A offers +70 to the decision maker and +70 to the Other participant [S+/O+] with P=80% or −70 to the decision maker and +70 to the Other participant [S−/O+] with P=20%. Option B offers the same outcomes but with opposite contingencies [S+/O+] with P=20% or [S−/O+] with P=80%. Notice that both options offer positive outcomes to Other; therefore, choice behavior and learning is primarily guided by Self PE.

(b) Self PE (2): Option A offers +70 to the decision maker and −70 to the Other participant [S+/O−] with P=80% or −70 to the decision maker and −70 to the Other participant [S−/O−] with P=20%. Option B offers the same outcomes but with opposite contingencies. Notice that both options offer negative outcomes to Other.

(c) Other PE (1): Option A offers +70 to the decision maker and +70 to the Other participant [S−/O−] with P=80% or +70 to the decision maker and −70 to the Other participant [S−/O−] with P=20%. Option B offers the same outcomes but with opposite contingencies. Notice that both options offer positive outcomes to Self; therefore, choice behavior and learning is primarily guided by Other PE.

(d) Other PE (2): Option A offers −70 to the decision maker and +70 to the Other participant [S+/O+] with P=80% or −70 to the decision maker and −70 to the Other participant [S+/O−] with P=20%. Option B offers the same outcomes but with opposite contingencies. Notice that both options offer negative outcomes to Self.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(e) Self and Other PE (1): In this and the next condition both Self and Other PE vary. Option A offers −70 to the decision maker and +70 to the Other participant [S−/O+] with P=80% or +70 to the decision maker and −70 to the Other participant [S+/O−] with P=20%. Option B offers the same outcomes but with opposite contingencies.

(f) Self and Other PE (2): Option A offers +70 to the decision maker and +70 to the Other participant [S+/O+] with P=80% or −70 to the decision maker and −70 to the Other participant [S−/O−] with P=20%. Option B offers the same outcomes but with opposite contingencies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

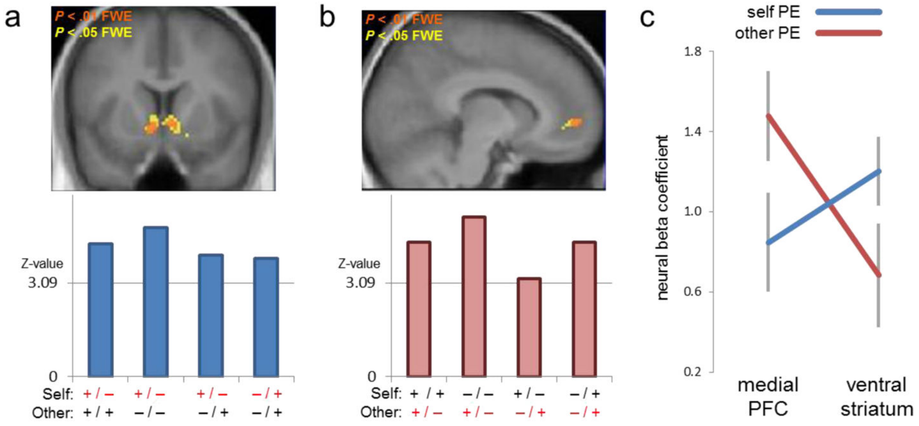

Fig. 3. Self and Other Prediction Error

(a) Prediction error signal for self-value. In each of 72 subjects, prediction errors associated with updating of self-value (PE_S) were estimated across four experimental conditions and subsequently regressed to hemodynamic activity. The four conditions are depicted in Fig. 2a, 2b, 2e, 2f, and correspond to the conditions in which self-value differed between the two available options. Upper panel: Consistent with previous reports of PE-related hemodynamic activity, a random-effects analysis across the four conditions revealed PE_S-related activity in ventral striatum (coordinates 4,10, −4; $P < .005$, FWE whole-brain-corrected). Lower panel: Z-values in ventral striatum are plotted separately for each of the four conditions (most significant voxel in striatum plotted for each condition: 6,4, −4; −8,20, −2; 16,6, −8; 28, −82, −22, respectively). For example, the first bar corresponds to the condition in which the two options yield either a positive or negative outcome to Self and always a positive outcomes to Other, while the last bar corresponds to the condition in which the two options yield positive or negative outcomes for both Self and Other. Z-Values for each of the four conditions exceeded the threshold for $P < .001$ (equivalent Z-value: 3.09), indicating striatal PE_S signal to be evident both within and across conditions.

(b) Prediction error signal for other-value. Preference-dependent prediction errors associated with updating of other-value (PE_S) were estimated across four experimental conditions and subsequently regressed to hemodynamic activity. The four conditions are depicted in Fig. 2c, 2d, 2e, 2f, and correspond to the conditions in which other-value differed between the two available options. Upper panel: A random-effects analysis across the four conditions revealed PE_O-related activity in medial prefrontal cortex (MPFC; coordinates 10,54,0; $P < .001$, FWE whole-brain-corrected). Lower panel: Z-values in MPFC (coordinates 4,60,14; 12,58,12; 4,58,2; 10,50,2) are plotted separately for each of the four conditions. Z-Values for each of the four conditions exceeded the threshold for $P < .001$ (equivalent Z-value: 3.09), indicating MPFC PE_O signal to be evident both within and across conditions.

(c) Beta values representing fitted responses to self [blue] and other [red] PE in medialprefrontal cortex (coordinate 10,54,0) and ventral striatum (coordinate 4,10,-4). The interaction of Region x PE type is significant ($F_{71} = 9.3$; $P < .005$).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

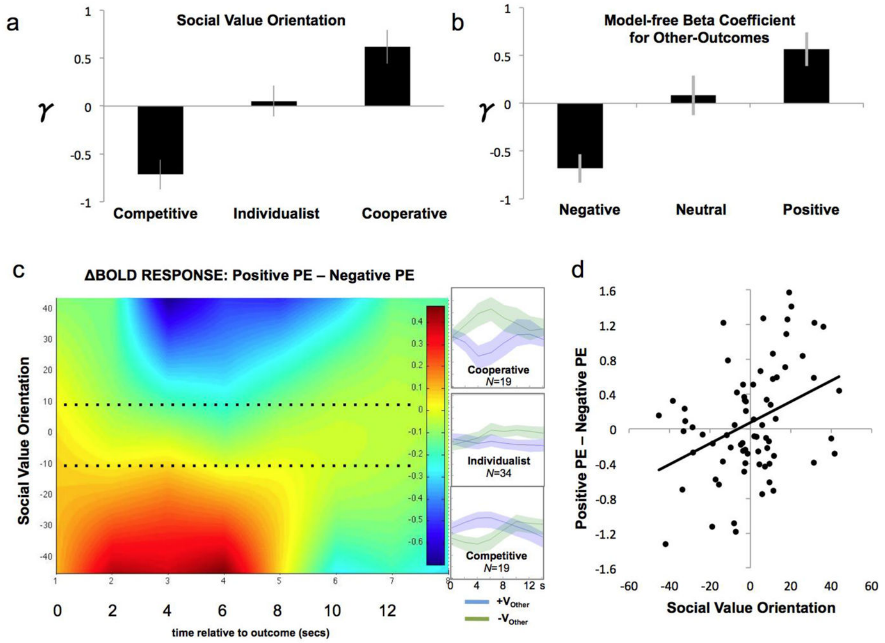

Fig. 4. Prediction error for "Other Values" (V_O)

(a) We divided the group of the 72 participants into three groups based upon an independent assessment of Social Value Orientation (see Methods). In this separate task, participants chose between allocations of an endowment between oneself and an anonymous social partner, thus revealing the Cooperative, Individualistic, or Competitive orientation of each participant. We used the mean SVO ± ½ std. deviation as boundary criterion. For each SVO-determined group, the average γ (i.e., fitted parameter determining social preferences in the hybrid learning model) is plotted on the y-axis. Positive γ is consistent with a cooperative orientation, while negative γ is consistent with a competitive orientation.

(b) For each subject we tested a binary logistic regression model with the 'outcome to self' and 'outcome to other' received on the previous trial as predictors; the choice (stay or switch) in the present trial as the dependent variable. We then created three groups based on the estimated betas associated with $\mathbf{V_O}$, using the mean beta value ± ½ std. deviation as boundary criterion. Positive beta values correspond to a higher probability of choosing "stay" when $\mathbf{V_O}$ of the previous trials is positive. The average γ (i.e., fitted parameter determining social preferences in the hybrid learning model) is plotted on the y-axis for each group.

(c) Differential BOLD response to positive and negative PE as a function of Valence, Social Value Orientation and time. X-axis represents time since onset of outcome screen. Y-axis represents Social Value Orientation. The heatmap represents fitted BOLD response to positive PE minus negative PE in 72 subjects. To facilitate comparison, we multiplied the PE_O-related hemodynamic response by each decision-maker's γ, thus transforming the PE_O signal to the perspective of the 'other' social partner. Thus, the positive PE_O of the cooperative decision-maker (better than expected from the perspective of 'other') becomes comparable to the negative PE_O of the competitive decision-maker (again, better than expected from the perspective of 'other'). Insets on the right depict peristumulus time

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

histograms to positive and negative PE for the three groups. Dotted line represents the boundaries separating the three groups.

(d) Correlation (*r* = .37; P < .005) between social value orientation (X-Axis) and difference in BOLD response 4 secs after the onset of the outcome screen (Y-axis).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Relationship of behavioral PE_O to neural response in MPFC. Note MPFC activity is negatively related to the modeled PE_O, regardless of social preference

Subject‘s Social	**Outcome**	**Behavioral**	**MPFC**
Preference	to Other	PE_O	Response
Cooperative	Positive $ (+)	Positive (+)	Negative (−)
Cooperative	Negative $ (−)	Negative (−)	Positive (+)
Individualistic	Positive $ (+)	Zero (±0)	Negligible
Individualistic	Negative $ (−)	Zero (±0)	Negligible
Competitive	Positive $ (+)	Negative (−)	Positive (+)
Competitive	Negative $ (−)	Positive (+)	Negative (−)