

Activity in ventromedial prefrontal cortex co-varies with revealed social preferences: evidence for person-invariant value

Jamil Zaki,¹ Gilberto López,² and Jason P. Mitchell²

¹Department of Psychology, Stanford University, Jordan Hall, Bldg 420, Jordan Hall, Bldg 420, Stanford, CA 94305, USA and ²Department of Psychology, Harvard University

Although altruistic and selfish behaviors seem fundamentally incommensurable humans regularly choose between them. One model of such choices suggests that individuals ascribe a common form of subjective value to their own outcomes and those of others. To test this ‘person invariance’ hypothesis, we asked individuals to choose between allocating varying amounts of money to themselves or to a partner. Participants’ choice patterns provided an estimate of the relative value they placed on their own and others’ gains. These estimates were used to isolate neural activity correlating with the subjective value of gains irrespective of the recipient (self or other) during a separate set of trials in which rewards were offered only to the self or partner. Activity in ventromedial prefrontal cortex scaled with this person-invariant value parameter, consistent with earlier demonstrations that this region supports common value computation. These data suggest that individuals reduce the value associated with their own and others’ experiences to a common subjective scale, which is used to guide social decision-making.

Keywords: prosociality; altruism; value; ventromedial prefrontal cortex; decision-making

By almost any metric, humans engage in an unparalleled amount of prosocial behavior, including helping unfamiliar individuals, sharing resources with others and cooperating in large groups. US citizens alone donate roughly 200 billion dollars to charities each year and produce the equivalent of another 200 billion dollars in labor through volunteering (Giving USA, 2006). Ethnographic observations demonstrate that people across a wide range of cultures spontaneously share resources with others during economic games (Henrich *et al.*, 2005). Even children as young as 18 months spontaneously help strangers (Warneken and Tomasello, 2006), and by 3 years of age consistently share resources with others (Hamann *et al.*, 2011).

Of course, humans are equally capable of self-serving behavior. Americans may donate billions of dollars to charities each year, but they also spend an equivalent amount on personal luxury goods, including jewelry, cosmetics and gourmet foods (Chamber of Commerce, 2007). Although people display some generous behavior on standard game theoretic paradigms, they typically keep a large majority of the resources available during such tasks, as opposed to making entirely equitable choices (Engel, 2011). And although young children may be considerably more likely to share resources than chimpanzees (Warneken and Tomasello, 2006; Hamann *et al.*, 2011), any observer of playground behavior can confirm that toddlers are hardly perfectly selfless creatures.

Given the human capacity for both generous and selfish behavior, how do individuals choose between improving their own well-being and prosocially attending to the welfare of others? After all, the outcomes associated with generous behavior are qualitatively different from those associated with selfish behavior. How, for example, does a person compare the value of donating to charity against the value of dining out at an expensive restaurant? Both may be valuable, but in deeply incompatible ways.

This dilemma strongly resembles a class of problems common to decision-making more generally. All organisms must frequently choose among incommensurable outcomes, such as whether to devote energy to procuring food, finding a mate or maintaining a nest. Both classic (Cabanac, 1992) and modern (Montague and Berns, 2002) theories suggest a clear solution to this dilemma, namely, that organisms transform the value of disparate outcomes onto a common scale that allows outcomes to be directly compared. More recently, researchers have suggested that this transformation function is supported by the ventromedial prefrontal cortex (vmPFC), a region involved in computing the subjective value of multiple reward types (Izquierdo *et al.*, 2004; Rangel and Hare, 2010; Schoenbaum and Esber, 2010; Grabenhorst and Rolls, 2011; Rushworth *et al.*, 2011). For example, activation of this region predicts participants’ decisions between disparate classes of rewarding stimuli, such as food and money (Chib *et al.*, 2009; Levy and Glimcher, 2011).

These findings suggest a potential answer to the question of how people arbitrate between their own and others’ welfare. Under this hypothesis, an individual might compare opportunities to act prosocially (e.g. donating money to another person) against opportunities to act selfishly (e.g. keeping that money for himself) by translating the value of both self- and other-oriented outcomes onto a common scale; behavioral studies by Andreoni and Miller (2002) suggest that common value calculations indeed support interpersonal decision-making. As with other instances of common value calculation, such ‘person-invariant value’ would likely be reflected in vmPFC activity.

To provide a direct test of these hypotheses, we scanned participants using functional magnetic resonance imaging (fMRI) while they made a series of decisions about whether to allocate sums of money to themselves or another person. These choices were modeled in a manner broadly consistent with economic theories of ‘revealed preferences’ (Samuelson, 1947) to estimate the relative subjective value participant ascribed to rewards given to themselves (self-oriented outcomes), as compared with rewards given to the other person (other-oriented outcomes). This information, in turn, allowed us to translate self- and other-oriented outcomes into a single, person-invariant parameter

Received 15 August 2012; Accepted 3 January 2013

Advance Access publication 12 January 2013

We thank T.J. Eisenstein and Adam Waytz for advice and assistance. This work was supported by a grant from the Templeton Foundation for Positive Neuroscience

Correspondence should be addressed to Jamil Zaki, Department of Psychology, Stanford University, Stanford, CA 94305, USA. E-mail: jzaki@stanford.edu

representing the common subjective value of both outcome types. We then used this parameter to predict neural response in a separate set of trials on which rewards were offered only to participants themselves or to another person. If choices to act prosocially vs selfishly draw on a reduction of self- and other-oriented outcomes onto a single-value parameter, then this common parameter should correlate with the response of brain regions associated with computing value, most notably, the vMPFC.

In contrast, if interpersonal choices represent a fundamentally different form of decision problem, we should find that no neural response (or, perhaps, responses in different brain regions) co-varies with this person-invariant parameter. For instance, neuroscientific investigations of prosocial choice in both children (Steinbeis *et al.*, 2012) and adults (Spitzer *et al.*, 2007) demonstrate that selfless—as compared with selfish—decision-making sometimes recruits areas of lateral pre-frontal cortex associated with inhibition of pre-potent responses. These data suggest an alternative to person-invariant value: altruistic and selfish value decision-making should instead reflect the operation of disparate cognitive processes and neural systems. The current paradigm provided us with a test of these competing predictions.

METHODS

Participants

Nineteen right-handed participants (12 male, $M_{age} = 23.2$, range = 18–36, $s.d. = 5.5$) with no history of psychiatric or neurological disorders completed the study in exchange for monetary compensation. Informed consent was obtained in accordance with the regulations of the Committee on the Use of Human Subjects at Harvard University. Four participants made no prosocial choices; as such, their data could not be modeled, resulting in a final data set of 15 participants (9 male, $M_{age} = 21.8$, range = 18–33, $s.d. = 4.2$). Data from a distinct, but partially overlapping subset of trials in these participants’ sessions are described elsewhere (Zaki and Mitchell, 2011).

Protocol

We combined an approach developed by Padoa-Schioppa and Assad (2006, 2008) to identify neural correlates of common value coding in monkeys with a game theoretic paradigm, in order to test the prediction that individuals represent selfish and altruistic outcomes using a common value metric. Prior to scanning, participants were introduced

to a confederate whom they believed was a second participant. An experimenter informed both the actual participant and the confederate that the study had been designed to examine how individuals form impressions about other people and that one of the participants would enter the scanner to make decisions about the other participant, while the second participant completed unrelated tasks outside the scanner. The experimenter did not mention an economic decision-making task. Through an ostensibly random (but actually fixed) assignment, the participant was assigned to enter the scanner.

After entering the scanner, participants completed a short social judgment task described elsewhere (Mitchell *et al.*, 2006; Waytz *et al.*, 2012) and were then introduced to a modified ‘dictator game’ (Hoffman *et al.*, 1994). Participants were told that, during this game, they would make repeated decisions about whether to allocate money to themselves or to the other participant (hereafter: the ‘receiver’) and that five of their decisions, chosen at random, would be enacted. Importantly, participants were told that they would not have any further contact with the receiver, and that the receiver did not know the participant was completing the dictator game. These features of the task were designed to minimize external pressures to act generously, and more cleanly tap participants’ actual valuation of selfish and generous actions (Zaki and Mitchell, 2011).

Following the instruction period, participants completed 210 rounds of the dictator game. Each trial began with an offer phase, in which the participant saw a photograph of herself or himself and a photograph of the receiver (each taken immediately prior to the start of the experiment). Photographs appeared on opposite sides of the screen, and the location of each was counterbalanced across trials. Each photograph was associated with a monetary offer printed immediately above it; for example, ‘\$1.00 for’ above the photograph of the participant and ‘\$1.50 for’ above the photograph of the receiver. After a jittered interval of 1.5–5.0 s, the word ‘Decide’ appeared onscreen, and participants were given 2 s to choose between the options. Participants’ choice appeared as a red box surrounding the chosen option, which was displayed onscreen for the remainder of the choice period (see Figure 1 for trial structure). Trials were separated by a jittered inter-trial interval of 1–6 s, during which participants passively viewed a fixation cross.

The amounts that each person stood to gain varied across trials, but always adhered to one of a set of six ratios specifying the relation between the two monetary amounts: 3:1, 2:1, 3:2, 4:3, 5:4 and 1:1.

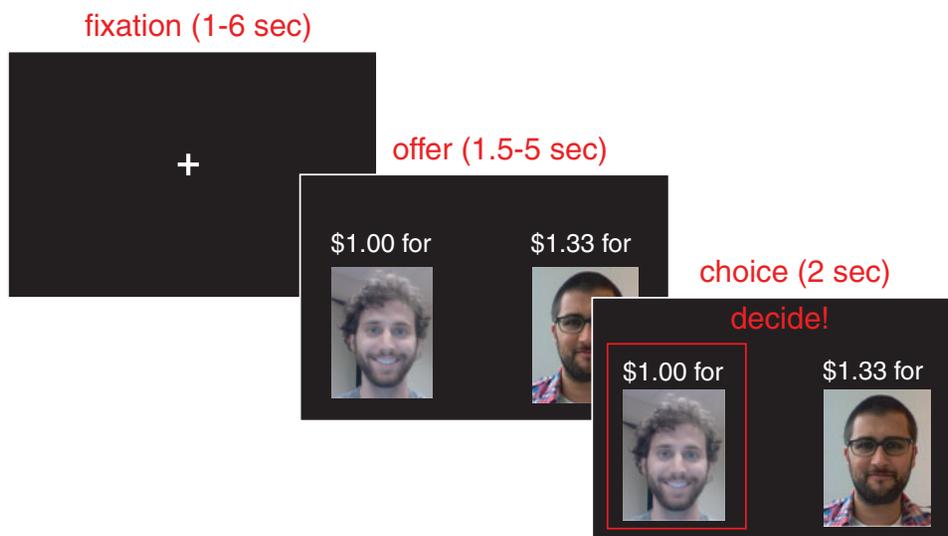


Fig. 1 Participants made repeated choices between allocating amounts of money to either themselves or their partner. The magnitude of potential other and self gains were always related by one of a set of ratios, ranging from 1:3 to 3:1.

For each trial, a random value between \$0.10 and \$3.00 was chosen, and a second value was determined by transforming the first value according to the ratio that applied during that trial. These two amounts were then randomly paired with the two targets (the participant and the receiver). For example, if the amount of \$1.50 was selected and the ratio was 2:1, then the other choice presented would be \$0.75. Note that each ratio could thus produce two relations between the amounts that the receiver and the participant stood to gain. If \$1.50 was assigned to the receiver and \$0.75 was assigned to the participant, this would produce a 2:1 ratio between possible other/self gains (i.e. the receiver would stand to gain twice as much as the participant himself). If the opposite assignment was made, the other/self ratio would instead be 1:2. As such, 11 total ratios could relate potential other to self gains: 1:3, 1:2, 2:3, 3:4, 4:5, 1:1, 5:4, 4:3, 3:2, 2:1 and 3:1. The maximum amount that either the participant or receiver could gain on any trial was \$9.00, and trials were organized such that the total amounts of money available to the participant and the receiver over the course of the entire study were comparable. The choice paradigm comprised 15 trials adhering to each of the 11 ratios.

In addition to these choice trials, the paradigm included a separate set of 15 ‘pure self’ and 15 ‘pure other’ reward trials. During ‘pure self’ trials, participants decided between a non-zero of money for themselves (e.g. \$1.00) vs \$0.00 for the receiver. During ‘pure other’ trials, participants decided between \$0.00 for themselves or a non-zero amount of money for the receiver. Thus, these choices represented ‘costless’ rewards for the participant and the receiver, respectively. Unsurprisingly, participants allocated money to the person who stood to gain something (either self or other) on >95% of such ‘pure’ trials, suggesting that they did indeed perceive these choices as pure self and other gains. Finally, we included, but did not model, 15 non-reward trials in which participants chose between \$0.00 for themselves and \$0.00 for the receiver.

Behavioral analysis

To estimate the relative subjective value that participants placed on their own and the receiver’s outcomes, we calculated the proportion of trials in which each participant chose to donate money to the receiver rather than allocate it to herself or himself at each of the 11 other/self reward ratios (i.e. all choices excluding ‘pure self’, ‘pure other’ and non-reward trials). These proportions were then fit to a cumulative distribution function (CDF) that described that participant’s choice patterns, defined as:

$$f_{cdf}(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - \mu}{\sqrt{2\sigma^2}} \right) \right]$$

where ‘erf’ represents the CDF’s error in fitting participants’ choices, μ is the calculated indifference ratio and σ^2 is the variance or width of the CDF. The MATLAB function `fminsearch` was used to iteratively find values for μ and σ that best fit each participant’s choice patterns. Once the appropriate CDF was identified, we solved for the indifference ratio, the value of x that solved $f(x) = 0.5$. This ‘indifference ratio’ represents the other/self reward ratio at which participants were equally likely to act selfishly or altruistically (Figure 2a).

We then used these indifference ratios to construct a ‘person-invariant value’ parameter. This single parameter was designed to capture the subjective value with which each participant would be expected to imbue both self- and other-oriented rewards, based on that participant’s choice patterns. This person-invariant value metric was then used as a parametric regressor to predict BOLD responses during the offer stage of the independent set of ‘pure self’ and ‘pure other’ trials. The critical step in constructing this parameter involved transforming the raw dollar amounts given to the self and other during these pure

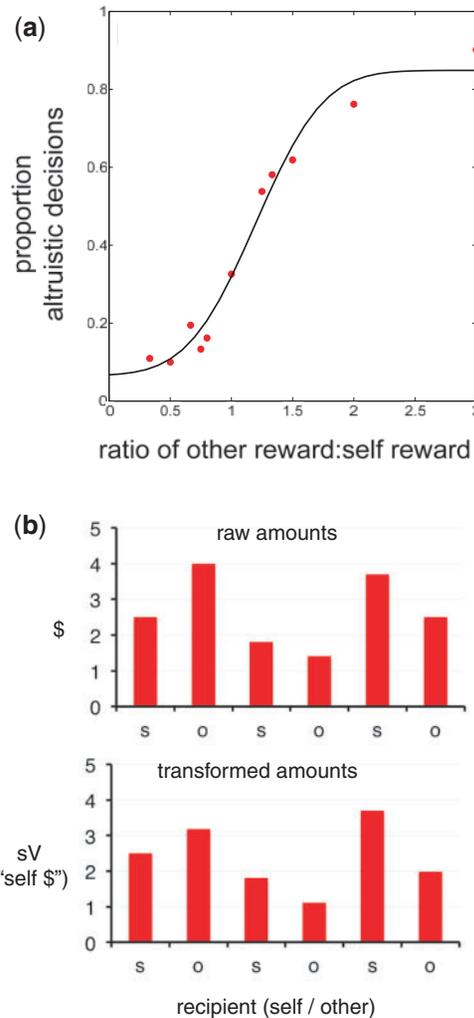


Fig. 2 (a) Cumulative distribution functions were fit to the proportions of altruistic choices participants made at each other/self reward ratio. These functions specify an ‘indifference ratio’ at which each participant would be expected to act selfishly or generously equally often. The group average indifference ratio (presented here) was \$1.26:\$1.00. (b) Each participant’s indifference ratio was used to construct a single-person-invariant regressor, which represented the subjective value participants would ascribe to a set of trials on which they (self-oriented) or their partner (other-oriented) received a reward. Raw dollar amounts of these rewards were transformed into a person-invariant scale by dividing other-oriented reward amounts by each participant’s indifference ratio (in this fictional example, the indifference ratio is the group mean of 1.26:1.00). The resulting regressor maps rewards onto a common scale of ‘self dollars’: that is, the subjective value of both self-oriented and other-oriented gains in units of the value of \$1.00 given to the self.

trials into subjective value, using subject-specific indifference ratios as indices of how much each participant valued their own vs their partner’s outcomes.

For example, consider a hypothetical participant, whose choices suggested an other/self indifference ratio of 2.00:1.00, indicating that he valued his own gains twice as much as the receiver’s (that is, that he placed equal value on receiving \$1 himself vs donating \$2 to the other person). A person-invariant value parameter for this participant would reflect the subjective value this participant would experience during ‘pure self’ and ‘pure other’ reward trials; this value, in turn, would be calculated by transforming raw dollar amounts during these trials based on the participant’s indifference ratio. For example, on trials in which that participant or the receiver gained the following amounts: [(a) \$1.00_{self} (b) \$1.00_{receiver} (c) \$2.60_{self} (d) \$2.60_{receiver}], the transformed values would be calculated by dividing receiver-reward trials by the

indifference value of 2 to yield: [(a') 1.00, (b') 0.50, (c') 2.60, (d') 1.30]. In other words, a person-invariant value function would reflect the fact that the participant values his own gains twice as much as the receiver, by calculating the subjective value of raw other-oriented rewards as $\frac{1}{2}$ the subjective value of identical raw values given to the self (for a visualization of an example transformation, see Figure 2b). Broadly, this approach assumes that the subjective value—and activity in related neural structures—associated with self- and other-oriented outcomes should reflect the following: (i) the magnitude of those outcomes and (ii) the relative value an individual places on their own and others' outcomes.

Neuroimaging acquisition and analysis

Imaging data were collected on a 3-Tesla Siemens Trio scanner using a gradient-echo echo-planar pulse sequence (31 axial slices, 5 mm thick; 1 mm skip; TR = 2 s; TE = 35 ms; 3.75×3.75 in-plane resolution). A high-resolution T1-weighted structural scan (MP-RAGE) was collected prior to three functional runs (214 TRs each). Stimuli were presented onto a screen at the end of the magnet bore using the Psychophysics Toolbox for Matlab (Brainard, 1997). Participants viewed the screen via a mirror mounted on the head coil, and a pillow and foam cushions were placed inside the coil to minimize head movement.

MRI data were preprocessed and analyzed using SPM (Wellcome Department of Cognitive Neurology, London, UK). Functional data were time-corrected for differences in acquisition time between slices for each whole-brain volume and realigned to correct for head movement. Data were then transformed into a standard anatomical space (3-mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute). Normalized data were then spatially smoothed (6 mm full-width-at-half-maximum) using a Gaussian kernel. We concatenated images from all three runs before subsequent analysis.

General linear models

Our main analysis consisted of a GLM aimed at isolating brain activity tracking with a person-invariant common value metric during pure self and pure other trials. Importantly, this common value metric correlated with the raw monetary amount on offer during pure self and pure other trials (i.e. the 'objective' value of these trials). To account for possible collinearity between objective and subjective values, we included the raw monetary amounts awarded during each trial as a separate regressor of no interest. Our main GLM included the following regressors: (1) onset and duration of offers across both 'pure self' and 'pure other' trials, (2) a parametric modulator reflecting the transformed, person-invariant subjective value associated with each 'pure self' and 'pure other' trial, (3) a parametric modulator reflecting the raw objective value on offer during each 'pure self' and 'pure other' trial, (4–6) temporal derivatives for conditions 1–3, (7–9) linear trends capturing drift in each fMRI run and (10–12) step functions to control for any differences in mean activity across runs.

Thus, this analysis controlled for the objective magnitude of monetary gains, and, as such, isolated neural activity that was specifically correlated with the subjective value of self and other reward.

We also calculated two additional GLMs to capture brain activity scaling with 'only' subjective value or raw monetary prize amounts, allowing us to compare these models' ability to predict brain activity. The first of these ('subjective only GLM') included all the regressors from our main GLM, with the exception of the parametric modulator corresponding to raw monetary values and its associated temporal derivative. The second these ('objective only GLM') included all the regressors from our main GLM, with the exception of the parametric modulator corresponding to subjective, person-invariant value and its associated temporal derivative.

These three GLMs were estimated individually for each participant, and contrast images from each participant were subsequently entered into second-level models for all three analyses (our main GLM, the 'objective only' GLM, and the 'subjective only' GLM), treating participants as random effects. Brain regions fitting each model were identified using a statistical criterion of 55 or more contiguous voxels at a voxel-wise threshold of $P < 0.005$. These height and extent thresholds were selected on the basis of a Monte Carlo simulation implemented in Matlab, to correspond with an overall false positive rate of $< 5\%$, corrected for multiple comparisons (Slotnick *et al.*, 2003).

RESULTS

When given choices between non-zero amounts for both themselves and the receiver, participants acted generously by allocating money to the receiver on a substantial proportion of trials ($M = 33.8\%$), thus foregoing a sizeable percentage of their total potential earning ($M = 22.2\%$). Such rates of giving are consistent with previously documented behavior during the dictator game (Engel, 2011). The group-average indifference ratio was $\$1.26/\1.00 (range: $\$0.71/\1.00 – $\$2.72/\1.00 , s.d. = 0.67); in other words, on an average, participants' choices suggested that they valued $\$1.26$ for the receiver as much as $\$1.00$ for themselves (Figure 2a).

As described earlier (see 'Methods' section), these indifference ratios were used to calculate a 'person-invariant' parameter, a single scale representing the subjective value that each participant would be expected to ascribe to both self- and other-oriented rewards, based on that participant's unique choice pattern. We then used this parameter to predict neural response during both 'pure self' and 'pure other' trials. A whole-brain, random-effects analysis revealed a single-brain region in which activity scaled with this person-invariant value parameter, located in vmPFC. Similar clusters were identified regardless of whether person-invariant value was the only regressor in the model (the 'subjective only' GLM described in 'Methods' section; MNI coordinates: $x/y/z = -4, 50, -9$, peak voxel $t = 5.21$, $k = 62$ voxels) or was included as part of a multiple regression that separately modeled the raw monetary values of participant and receiver rewards (the 'main GLM' described in 'Methods' section; $x/y/z = -8, 48, -6$, peak voxel $t = 4.42$, $k = 60$ voxels; see Figure 3). This second model controlled for any effect of objective (as opposed to subjective) value in predicting brain activity, and suggests



Fig. 3 A whole-brain analysis isolated neural activity scaling with a person-invariant metric of subjective value associated with self 'and' other rewards—calculated using each participant's indifference ratio—in a separate set of trials. Only ventromedial prefrontal cortex (MNI coordinates: $-8, 48, -6$) significantly correlated with this value index.

that activity in vMPFC cannot be explained by co-linearity between subjective and objective value during the 'pure self' and 'pure other' trials.

In contrast, objective value—that is, the raw amount of money awarded to participants themselves and to the receiver during 'pure self' and 'pure other' trials, irrespective of participants' indifference ratio—did not predict activity anywhere in the brain, even at a lenient threshold of $P < 0.05$, uncorrected for multiple comparisons, and regardless of whether raw monetary value was the only regressor in the model or part of a multiple regression that also included the person-invariant value parameter. This pattern suggests that vMPFC does not represent the mere objective value of self- and other-oriented reward, but rather specifically scales with the 'subjective' value of these events, calculated on the basis of participants' patterns of interpersonal choice.

DISCUSSION

Individuals continually face choices between maximizing their own well-being or fostering the welfare of others. Because selfish and generous choices result in fundamentally incommensurate outcomes, it has been unclear how individuals evaluate such interpersonal decisions. Here, we demonstrate that individuals transform the value associated with self- and other-oriented rewards into a 'common currency' that supports a direct comparison of these disparate social outcomes. When monetary rewards were alternately allocated to a participant and to another person, neural response tracked with the predicted subjective value that participants placed on those outcomes on a single-value scale, as estimated from participants' selfish vs generous choice patterns. Importantly, this neural modulation was restricted to the vMPFC, a region strongly linked to the computation of subjective value in the service of decision-making (Padoa-Schioppa and Assad, 2006; Chib et al., 2009; Rangel and Hare, 2010; Schoenbaum and Esber, 2010; Grabenhorst and Rolls, 2011; Levy and Glimcher, 2011; Levy et al., 2011; Rushworth et al., 2011).

Relating self and other value computations

These data add to a growing body of work that weighs against models of prosociality and selfishness as fundamentally antagonistic. A number of commentators have suggested that humans are predominantly concerned with maximizing their personal welfare, but can sometimes suppress these self-serving reflexes in order to act prosocially (Camerer and Fehr, 2006; Knoch et al., 2006; Spitzer et al., 2007; DeWall et al., 2008; Steinbeis et al., 2012). On this view, self-serving and prosocial decisions should depend on fundamentally different psychological mechanisms, and drive activity in separable neural circuits.

However, recent observations suggest a fundamentally different view, under which self- and other-regarding behaviors may stem from the same underlying set of motivations and neural systems rather than from countervailing mechanisms. For example, mesolimbic dopaminergic targets—including vMPFC—respond robustly when individuals receive rewards, but also when they observe others experiencing such rewards and decide to act generously toward others (Moll et al., 2006; Harbaugh et al., 2007; Mobbs et al., 2009; Hare et al., 2010; Tricomi et al., 2010; Zaki and Mitchell, 2011; Dawes et al., 2012; Janowski et al., 2012; Morelli et al., 2012; Zaki and Ochsner, 2012; Zaki, in press).

These data support the surprising conclusion that the human brain responds similarly to positive outcomes to self and other, and suggest that common value computations could support interpersonal decisions. Here, we provide complementary data that support this emerging conceptual model, by combining methodological techniques designed to identify neural correlates of common value currency in vMPFC (Padoa-Schioppa and Assad, 2006; Levy and Glimcher, 2011; Levy et al., 2011) with a game theoretic paradigm. This approach produced strong converging evidence that vMPFC indeed performs

common value calculations across both self- and other-oriented outcomes, and that these calculations are reflected in participants' patterns of selfish and prosocial decision-making. More broadly, these findings suggest that—in much the same way as intrapersonal decisions (e.g. between two different goods)—interpersonal decision-making draws on the ability to 'translate' disparate outcomes into a common value metric through which outcomes can be compared directly against one another.

These data draw a number of additional connections between the rich neuroeconomic literature on value-based decision-making on the one hand, and social behaviors on the other hand. First, neuroscientific models of decision-making distinguish between value associated with 'outcomes' (e.g. a desired stimulus, event or object) and with 'actions' (e.g. choosing that stimulus over another one). Studies of brain activity during the act of decision-making suggest that vMPFC activity tracks value associated with actions (e.g. the value of an outcome minus the costs of obtaining it; Kable and Glimcher, 2009; Wunderlich et al., 2009; Rangel and Hare, 2010). However, other work suggests that neural signals related to outcome values, assessed during decision-making (Wunderlich et al., 2010; Grabenhorst and Rolls, 2011) and even in the absence of the need for any choice (Levy et al., 2011) can be used independently to predict decisions. Our work joins other efforts (Harbaugh et al., 2007) in translating this finding to the interpersonal domain. Specifically, we used participants' prosocial and selfish decisions to model brain activity during a set of trials in which rewards were offered only to themselves or another person. Although these trials retained the structure of choices between two outcomes, they functionally served as opportunities to model neural signals related to 'pure' rewards received by the self or other. The strong and selective relation between vMPFC activity during these trials and participants' choices during an independent set of trials suggests that outcome value can be used to predict decision-making in both intra- and interpersonal contexts.

Second, although prior work has documented overlapping engagement of vMPFC during self- and other-oriented value computation (Hare et al., 2010; Tricomi et al., 2010; Janowski et al., 2012), one outstanding question is the relative 'magnitude' of vMPFC signal accompanying to selfish and selfless outcomes. Here, we provide data concerning this magnitude, by demonstrating that vMPFC activity scales with the relative value that individuals place on their own, vs others' outcomes. For instance, vMPFC will respond with equal magnitude to self- and other-oriented reward if an individual—based on his choices—equally values his own and someone else's outcomes, but not if an individual holds a different subjective value profile. The relative magnitude of self- and other-oriented signals in vMPFC should further shift with manipulations that modulate the subjective value of others' outcomes, such as others' deservingness (Hare et al., 2010) and similarity to the self (Mobbs et al., 2009). Broadly, this suggests that vMPFC signals in social contexts flexibly track parameters associated with subjective value, as they do in non-social contexts (Tom et al., 2007; Padoa-Schioppa, 2009).

Third, our work suggests that—as in non-social contexts—the underlying utility of social outcomes can be derived from a series of binary decisions. Whereas our work borrows from the logic of 'revealed preferences'—use of individuals' choices to isolate their subjective valuation of stimuli—revealed preferences are typically examined in non-social contexts, for instance by studying preferences between different forms of goods according to purchase decisions. However, this same logic can be adapted to interpersonal choices. Here, we systematically varied the 'price' of giving to others on a trial-by-trial basis, allowing us to use participants' choices to reconstruct the strength of their preference for giving (i.e. how much they would pay for the opportunity to 'buy' a dollar given to someone else). Consistent with Andreoni and Miller's

(2002) behavioral findings, we demonstrate that this approach allows us to predict vMPFC signals associated with value.

Of course, the observation that vMPFC represents the value associated with self- and other-oriented rewards on a common scale does not imply that such outcomes are experienced as subjectively identical. As would be expected from a region capable of evaluating disparate types of outcomes, the vMPFC receives inputs from a range of distinct neural structures that provide different kinds of information relevant for guiding behavior, such as an animal's need for food, water or sleep (Carmichael and Price, 1996; Rolls, 2000; Ongur *et al.*, 2003). Likewise, when evaluating outcomes that benefit others, the vMPFC may operate over inputs from areas involved in social cognition, such as the medial prefrontal cortex or superior temporal sulcus (Tankersley *et al.*, 2007; Masten *et al.*, 2011; Waytz *et al.*, 2012). Indeed, recent evidence suggests that decisions to make charitable donations produce functional coupling between vMPFC and the superior temporal sulcus (Hare *et al.*, 2010; see also Zaki *et al.*, 2007 for a similar connectivity pattern when observing others in pain).

CONCLUSION

Although at first blush, selfish and prosocial behavior seem deeply incommensurable, these disparate outcomes may in fact result from a common representation of subjective value, instantiated by brain regions that evaluate personal outcomes such as food or money. It appears that individuals are similarly able to reduce the value of self- and other-oriented outcomes onto a single-person-invariant value scale. This ability likely plays a critical role in the constant decisions individuals must make between serving themselves and helping others.

REFERENCES

- Andreoni, J., Miller, J. (2002). Giving according to GARP: an experimental study of rationality and altruism. *Econometrica*, 70(2), 737–53.
- Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–6.
- Cabanac, M. (1992). Pleasure: the common currency. *Journal of Theoretical Biology*, 155(2), 173–200.
- Camerer, C.F., Fehr, E. (2006). When does “economic man” dominate social behavior? *Science*, 311(5757), 47–52.
- Carmichael, S.T., Price, J.L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology*, 371, 179–207.
- Chib, V.S., Rangel, A., Shimojo, S., O’Doherty, J.P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *The Journal of Neuroscience*, 29(39), 12315–20.
- Chilean and American Chamber of Commerce. (2007). *Snapshot of the U.S. Luxury Goods Market*. Philadelphia, PA: Chilean and American Chamber of Commerce.
- Dawes, C.T., Loewen, P.J., Schreiber, D., *et al.* (2012). Neural basis of egalitarian behavior. *Proceedings of the National Academy of Sciences*, 109(17), 6479–83.
- DeWall, C.N., Baumeister, R.F., Gailliot, M.T., Maner, J.K. (2008). Depletion makes the heart grow less helpful: helping as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulletin*, 34(12), 1653–62.
- Engel, J. (2011). Dictator games: a meta study. *Experimental Economics*, 14, 583–610.
- Giving USA. (2006). Indianapolis, IN: Giving Institute.
- Grabenhorst, F., Rolls, E.T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences*, 15(2), 56–67.
- Hamann, K., Warneken, F., Greenberg, J.R., Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature*, 476(7360), 328–31.
- Harbaugh, W.T., Mayr, U., Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, 316(5831), 1622–5.
- Hare, T.A., Camerer, C.F., Knoepfle, D.T., Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, 30(2), 583–90.
- Henrich, J., Boyd, R., Bowles, S., *et al.* (2005). “Economic man” in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795–815; discussion 815–755.
- Hoffman, E., McCabe, K., Shachat, K., Smith, V. (1994). Preferences, property rights and anonymity in bargaining games. *Games and Economic Behavior*, 7, 346–80.
- Izquierdo, A., Suda, R.K., Murray, E.A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience*, 24(34), 7540–8.
- Janowski, V., Camerer, C., Rangel, A. (2013). Empathic choice involves vmPFC value signals that are modulated by social processing implemented in IPL. *Social Cognitive and Affective Neuroscience*, 8(2), 201–8.
- Kable, J.W., Glimcher, P.W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, 63(6), 733–45.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829.
- Levy, D.J., Glimcher, P.W. (2011). Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *Journal of Neuroscience*, 31(41), 14693–707.
- Levy, I., Lazzaro, S.C., Rutledge, R.B., Glimcher, P.W. (2011). Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *Journal of Neuroscience*, 31(1), 118–25.
- Masten, C.L., Morelli, S.A., Eisenberger, N.I. (2011). An fMRI investigation of empathy for ‘social pain’ and subsequent prosocial behavior. *Neuroimage*, 55(1), 381–8.
- Mitchell, J.P., Macrae, C.N., Banaji, M.R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50(4), 1–9.
- Mobbs, D., Yu, R., Meyer, M., *et al.* (2009). A key role for similarity in vicarious reward. *Science*, 324(5929), 900.
- Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J. (2006). Human fronto-temple networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences*, 103(42), 15623–8.
- Montague, P.R., Berns, G.S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, 36(2), 265–84.
- Morelli, S.A., Rameson, L.T., Lieberman, M.D. (2012). The neural components of empathy: predicting daily prosocial behavior. *Social Cognitive and Affective Neuroscience*.
- Ongur, D., Ferry, A.T., Price, J.L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *Journal of Comparative Neurology*, 460(3), 425–49.
- Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. *Journal of Neuroscience*, 29(44), 14004–14.
- Padoa-Schioppa, C., Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090), 223–6.
- Padoa-Schioppa, C., Assad, J.A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience*, 11(1), 95–102.
- Rangel, A., Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, 20(2), 262–70.
- Rolls, E.T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, 10(3), 284.
- Rushworth, M.F.S., Noonan, M.A.P., Boorman, E.D., Walton, M.E., Behrens, T.E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, 70(6), 1054–69.
- Samuelson, P. (1947). *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press.
- Schoenbaum, G., Esber, G.R. (2010). How do you (estimate you will) like them apples? Integration as a defining trait of orbitofrontal function. *Current Opinion in Neurobiology*, 20(2), 205–11.
- Slotnick, S.D., Moo, L.R., Segal, J.B., Hart, J.Jr (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Research Cognitive Brain Research*, 17(1), 75–82.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, 56(1), 185–96.
- Steinbeis, N., Bernhardt, B., Singer, T. (2012). Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron*, 73(5), 1040–51.
- Tankersley, D., Stowe, C.J., Huettel, S.A. (2007). Altruism is associated with an increased neural response to agency. *Nature Neuroscience*, 10(2), 150–1.
- Tom, S.M., Fox, C.R., Trepel, C., Poldrack, R.A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, 315(5811), 515–8.
- Tricomi, E., Rangel, A., Camerer, C.F., O’Doherty, J.P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, 463(7284), 1089–91.
- Warneken, F., Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 311(5765), 1301–3.
- Waytz, A., Zaki, J., Mitchell, J. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *Journal of Neuroscience*, 32, 7646–50.
- Wunderlich, K., Rangel, A., O’Doherty, J.P. (2009). Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences*, 106(40), 17199–204.
- Wunderlich, K., Rangel, A., O’Doherty, J.P. (2010). Economic choices can be made using only stimulus values. *Proceedings of the National Academy of Sciences*, 107(34), 15005–10.
- Zaki, J. (in press). Cue integration: a common framework for physical perception and social cognition. *Perspectives on Psychological Sciences*.
- Zaki, J., Mitchell, J. (2011). Equitable decision making is associated with neural markers of subjective value. *Proceedings of the National Academy of Sciences of the United States of America*, 108(49), 19761–6.
- Zaki, J., Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls, and promise. *Nature Neuroscience*, 15(5), 675–80.
- Zaki, J., Ochsner, K.N., Hanelin, J., Wager, T., Mackey, S.C. (2007). Different circuits for different pain: patterns of functional connectivity reveal distinct networks for processing pain in self and others. *Social Neuroscience*, 2(3 and 4), 276–91.