nature neuroscience

Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice

Leor M Hackel[1], Bradley B Doll[2,3] & David M Amodio[1,2]

Humans learn about people and objects through positive and negative experiences, yet they can also look beyond the immediate reward of an interaction to encode trait-level attributes. We found that perceivers encoded both reward and trait-level information through feedback in an instrumental learning task, but relied more heavily on trait representations in cross-context decisions. Both learning types implicated ventral striatum, but trait learning also recruited a network associated with social impression formation.

© 2015 Nature America, Inc. All rights reserved.

Through positive and negative experiences, humans learn about other people and objects, and the value of these interactions guides their future decisions. Although learning through reward has been characterized by models of reinforcement[1], social contexts highlight a different form of learning, in which perceivers look beyond the reward of an interaction to encode more abstract and enduring traits[2]. For example, a colleague who shares resources may not only be associated with high reward value, but may also be seen as generous—a trait-level attribute suggesting she would be valued in other contexts as well, such as when choosing a companion for a social gathering. This higher level component of inference, found in classic psychological theories[3,4], is absent in models of reinforcement learning, yet may be especially important for social decision-making.

Trait learning involves the encoding of stable characteristics from observed behaviors[3–5], which in turn inform behavioral predictions and future decisions to interact with an individual across contexts[2,4]. This form of learning need not be limited to fellow humans; early theories of trait inference likened social dispositions to stable characteristics of objects that allow the prediction of their behavior across contexts[4], and recent research shows that trait-level social inferences and abstract inferences about nonsocial objects recruit overlapping neural regions[6]. Thus, while trait-level learning may be especially important for social decisions, it should inform decisions about both social and nonsocial objects beyond the reward-based learning explained by existing models.

To investigate the process through which traits are encoded beyond the reward value of people and objects during feedback-based learning, we examined neural activity using functional magnetic resonance imaging (fMRI) during a task that decorrelated trait and reward learning. Thirty-one participants played a game involving four human targets who had ostensibly participated in prior sessions and four computerized slot machines. Both human and slot machine targets could pay out from a pool of points, which were converted into money at the end of the experiment.

During a training phase, participants' interaction choices were followed by feedback indicating (i) the reward obtained from a target and (ii) the point pool available to the target. From this information, participants could infer the human or slot target's generosity (defined as the proportion shared) above and beyond reward value (defined as the magnitude shared). By design, average generosity was orthogonal to average reward (**Supplementary Fig. 1**).

Next, in the test phase, point pools available to each target were shown before choice, allowing us to test whether participants combined this contextual information with generosity knowledge during decision-making. Feedback was not provided, preventing further learning. Finally, to test whether learned generosity predicted choices in a different context, we assessed participants' preferences for a cooperative, noneconomic interaction with each human target.

Computational modeling of behavior indicated that participants learned both reward value (experienced outcomes) and generosity information (proportion of the allotted pool shared) and used a weighted combination of expectations derived from each to make choices across human and slot targets (see Online Methods and **Supplementary Tables 1** and **2**). However, participants relied more strongly on generosity information than reward value throughout the task: a weighting parameter w assessing the relative degree to which choice was driven by rewards ($w = 0$) or generosity ($w = 1$) was significantly greater than 0.50 (median = 0.70, Wilcoxon sign rank test, $z = -2.98$, $P = 0.003$). Moreover, when w was allowed to vary between task phases, estimates for training (median = 0.76) and test phases (median = 0.70) did not differ significantly ($z = -0.61$, $P = 0.54$). This pattern emerged even though the optimal strategy during training would be to ignore generosity information. Simulations revealed that relying solely on reward value would yield more points (mean (M) = 3,133.10, s.d. = 19.92) than relying equally on reward value and generosity (M = 2,753, s.d. = 54.76) or solely on generosity (M = 2,472.90, s.d. = 38.43), suggesting a propensity to encode trait-level information when the context affords such inferences.

We next tested whether neural activity in response to feedback reflected the encoding of generosity prediction errors (experienced generosity – expected generosity), independently of reward prediction errors (experienced reward – expected reward). Across human and slot trials, reward prediction errors correlated significantly with blood oxygenation level–dependent (BOLD) signal in the right ventral striatum (small-volume corrected, family-wise error–corrected P value (P_{FWE}) < 0.05; **Fig. 1a** and **Supplementary Table 3**), a region consistently associated with reward learning[7]. Strikingly, generosity prediction errors also correlated with BOLD signal in right

[1]Department of Psychology, New York University, New York, New York, USA. [2]Center for Neural Science, New York University, New York, New York, USA. [3]Department of Psychology, Columbia University, New York, New York, USA. Correspondence should be addressed to L.M.H. (leor.hackel@nyu.edu) or D.M.A. (david.amodio@nyu.edu).

Received 5 May; accepted 5 July; published online 3 August 2015; doi:10.1038/nn.4080

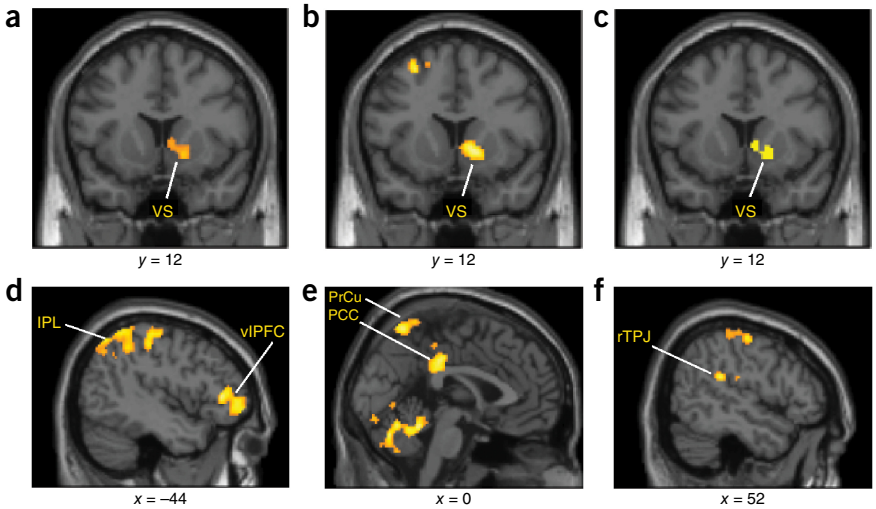

Figure 1 Statistical parametric maps showing neural regions where activity correlated with prediction errors during feedback ($n = 30$). (**a**,**b**) Activity in ventral striatum (VS) correlated with both reward prediction errors (**a**; $P_{FWE} < 0.05$, small-volume corrected) and generosity prediction errors (**b**; $P_{FWE} < 0.05$, whole-brain corrected). (**c**) Conjunction of **a** and **b**. (**d**–**f**) Generosity prediction errors further correlated with activity in a set of regions previously associated with social impression updating ($P_{FWE} < 0.05$, whole-brain corrected), including (**d**) left ventrolateral prefrontal cortex (vlPFC) and inferior parietal lobule (IPL), (**e**) posterior cingulate cortex (PCC) and precuneus (PrCu), and (**f**) right temporoparietal junction (rTPJ).

ventral striatum (whole-brain corrected, $P_{FWE} < 0.05$; **Fig. 1b**; conjunction shown in **Fig. 1c**), indicating that the striatum also encoded trait-level feedback, independent of rewards received, during instrumental learning.

Prediction errors for generosity, but not reward, correlated with an additional set of regions previously implicated in social impression updating (whole-brain corrected, $P_{FWE} < 0.05$; **Fig. 1d–f**), including left ventrolateral prefrontal cortex, bilateral inferior parietal lobule, posterior cingulate cortex, precuneus and right temporoparietal junction[8–11]. This pattern was observed for human and slot trials, and activity in these regions did not differ significantly by target (though it remains possible that differences could emerge using other methods), suggesting these regions perform computations for trait-level learning relevant to social contexts but not exclusive to them.

To understand how learned generosity information is incorporated into choice, we examined behavior and neural activity during the test phase. Behavioral analyses supported computational modeling results: across human and slot trials, participants were faster (**Fig. 2a**) and more likely (**Fig. 2b**) to choose targets who were higher on both prior reward value and generosity than the alternative target displayed onscreen (P values ≤ 0.001). Also consistent with the modeling results, generosity effects were significantly stronger than those of prior reward value (P values ≤ 0.004; we observed no reward × generosity interactions, P values > 0.49; see Online Methods and **Supplementary Tables 4** and **5**).

To use generosity knowledge during test-phase decisions, participants needed to combine pool amounts displayed onscreen with previously learned generosity information to generate stimulus–outcome value expectancies. In previous research, value signals relevant to decision-making have been associated with ventromedial prefrontal cortex (vmPFC)[12]. Therefore, we defined a region of interest in vmPFC that responded to the difference in point pools displayed for chosen minus unchosen targets (**Fig. 2c** and **Supplementary Table 6**), given that point pools directly represented the amount of money available. This region responded significantly to the difference in generosity between chosen and unchosen targets ($M = 22.82$, s.d. = 37.88, $t(29) = 3.30$, $P = 0.003$), suggesting that generosity information is combined with value signals when relevant to choice.

An analysis of individual differences corroborated this pattern: participants' generosity parameter estimates in vmPFC correlated with behavioral reliance on generosity information indicated by w ($r = 0.35$, $P = 0.03$, one-tailed; **Fig. 2d**). Thus, the representation of generosity knowledge in vmPFC explained variance in choice of targets on the basis of generosity.

A signature feature of trait learning is that it informs decisions across contexts[4]. Therefore, we tested the generalization of trait and reward learning to noneconomic social decisions. Participants were told, after the scan, that they might be invited back to complete a cooperative puzzle-solving task with another participant from the study, with no opportunity for winning additional money. Participants rated their preference for being partnered with each of the four human targets.

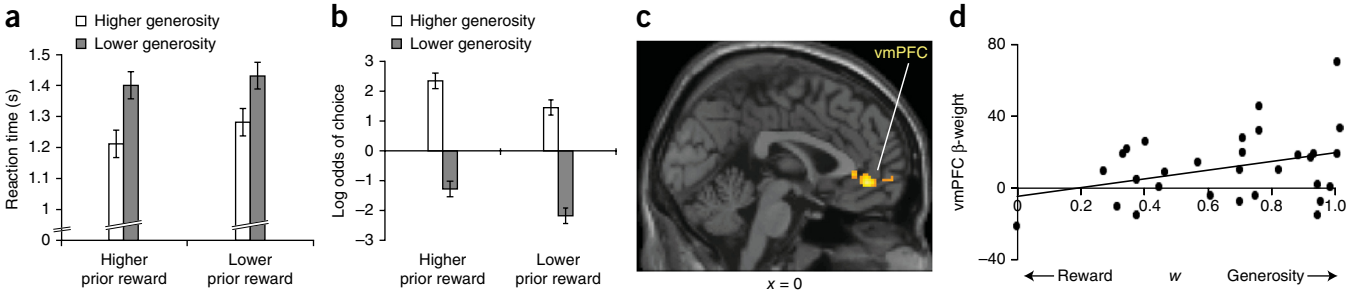

Figure 2 Test phase of task. Behavioral data points show predicted reaction times (**a**) and log odds of choice (**b**), ± standard error of coefficients, as estimated from multilevel linear and logistic regressions ($n = 31$). (**a**) Participants were faster to choose targets higher on prior reward value ($t(28.10) = -3.63$, $P = 0.001$) and generosity ($t(29.29) = -6.72$, $P = 0.2.16 \times 10^{-7}$) than an alternative. (**b**) Participants were more likely to choose targets higher on prior reward value ($z = 3.40$, $P = 0.0007$) and generosity ($z = 8.83$, $P = 2 \times 10^{-16}$) than an alternative. (**c**) A region of vmPFC responsive to points available for the chosen minus unchosen targets ($P_{FWE} < 0.05$, whole-brain corrected, $n = 30$) also responded to the generosity of chosen minus unchosen targets (region of interest analysis, $t(29) = 3.30$, $P = 0.003$, $n = 30$). (**d**) vmPFC responses to generosity correlated with a model-derived weighting parameter that estimated reliance on generosity in choice ($r = 0.35$, $P = 0.03$, one-tailed, $n = 30$).

© 2015 Nature America, Inc. All rights reserved.

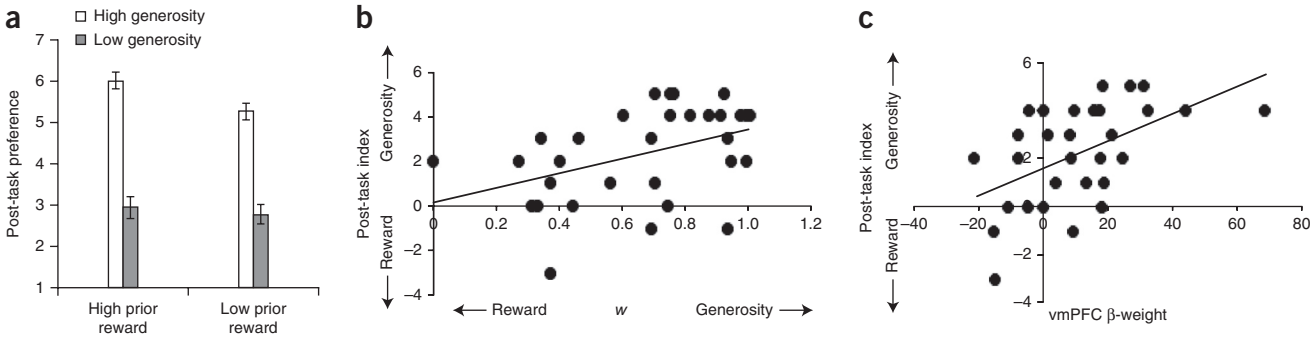

Figure 3 Post-task preferences for a future cooperative interaction featuring no economic rewards. (**a**) Participants ($n = 31$) preferred targets associated with high generosity ($F(1,30) = 100.2$, $P < 0.001$) and high prior reward ($F(1,30) = 5.32$, $P = 0.03$); the effects of generosity were stronger than those of prior reward ($F(1,30) = 38.36$, $P < 0.001$). Bars depict mean ± s.e.m. (**b**) Reliance on generosity in the economic task (indexed by weighting parameter w) correlated with post-task sensitivity to generosity versus prior reward ($r = 0.42$, $P = 0.02$, two-tailed, $n = 31$). (**c**) Parameter estimates for generosity in vmPFC correlated with sensitivity to generosity versus prior reward in post-task preferences ($r = 0.52$, $P = 0.003$, two-tailed, $n = 30$).

Participants strongly preferred to interact with more generous targets ($F(1,30) = 100.2$, $P < 0.001$, partial eta squared (η^2_p) = 0.77). Although they also preferred targets previously associated with high reward value ($F(1,30) = 5.32$, $P = 0.03$, $\eta^2_p = 0.15$), the effect of generosity was much larger ($F(1,30) = 38.36$, $P < 0.001$, $\eta^2_p = 0.56$) (**Fig. 3a**). The reward × generosity interaction was nonsignificant ($P = 0.18$).

Additionally, individual differences in reliance on generosity during the economic task (indicated by w) predicted reliance on generosity over reward information in post-task ratings (see Online Methods) ($r = 0.42$, $P = 0.02$; **Fig. 3b**). Furthermore, generosity parameter estimates in vmPFC during the economic task predicted post-task reliance on generosity over reward value ($r = 0.52$, $P = 0.003$; **Fig. 3c**). Hence, representations of trait knowledge coinciding with neural value representations supported trait-based social decision-making in a noneconomic context—a pattern of generalization that is the hallmark of a trait.

Together these findings demonstrate that trait information is encoded from feedback in addition to specific reward values and that trait information can dominate over reward information in decision-making. Notably, both forms of learning were associated with prediction error signals in ventral striatum. These findings extend our understanding of ventral striatum function beyond the encoding of model-based prediction errors[13], belief prediction errors[14] and hierarchical prediction errors[15], and beyond the influence of prior social knowledge and expectancies (for example, of fairness) on reward processing[16,17]. Although trait-level learning is particularly relevant to human interaction, our results suggest it pertains to any entity that affords a trait-like attribution.

This research also advances sociocognitive theories of impression formation. Trait learning has previously been examined through instructed or observational learning (for example, presenting trait-implying behavior descriptions), without feedback from social targets[2]. We demonstrated that impressions may also be formed through feedback-based instrumental learning. Indeed, the instrumental processes examined here implicated neural regions involved in reward processing and impression updating, but did not significantly correlate with activity in dorsomedial prefrontal regions linked previously to instructed trait learning[18]. Evidence for instrumental learning of traits complements previously studied trait learning processes, consistent with a multiple memory systems model of social cognition[19]. More broadly, these findings reveal the promise of integrating behavioral economics, reinforcement learning and computational neuroimaging to illuminate complex human decision-making processes[20].

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Rosenthal for assistance with data collection, and members of the New York University Social Neuroscience Laboratory and P. Mende-Siedlecki for comments on the manuscript. This work was funded by the New York University Center for Brain Imaging and by the US National Science Foundation (grant BCS 0847350).

AUTHOR CONTRIBUTIONS

L.M.H., B.B.D. and D.M.A. designed the research; L.M.H. collected the data; L.M.H. and B.B.D. analyzed the data; L.M.H., B.B.D. and D.M.A. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Sutton, R.S. & Barto, A.G. *Reinforcement Learning: An Introduction* (Cambridge Univ. Press, 1998).
2. Rim, S., Uleman, J.S. & Trope, Y. *J. Exp. Soc. Psychol.* **45**, 1088–1097 (2009).
3. Eysenck, H.J. *Dimensions of Personality* (Routledge & Kegan Paul, London, 1947).
4. Heider, F. *The Psychology of Interpersonal Relations* (New York, Wiley, 1958).
5. Jones, E.E. & Davis, K.E. *Adv. Exp. Soc. Psychol.* **2**, 219–266 (1965).
6. Baetens, K., Ma, N., Steen, J. & Van Overwalle, F. *Soc. Cogn. Affect. Neurosci.* **9**, 817–824 (2014).
7. Garrison, J., Erdeniz, B. & Done, J. *Neurosci. Biobehav. Rev.* **37**, 1297–1310 (2013).
8. Mende-Siedlecki, P., Baron, S.G. & Todorov, A. *J. Neurosci.* **33**, 19406–19415 (2013).
9. Mende-Siedlecki, P., Cai, Y. & Todorov, A. *Soc. Cogn. Affect. Neurosci.* **8**, 623–631 (2013).
10. Ma, N. *et al. Soc. Cogn. Affect. Neurosci.* **7**, 937–950 (2012).
11. Cloutier, J., Gabrieli, J.D.E., O'Young, D. & Ambady, N. *Neuroimage* **57**, 583–588 (2011).
12. Valentin, V.V., Dickinson, A. & O'Doherty, J.P. *J. Neurosci.* **27**, 4019–4026 (2007).
13. Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P. & Dolan, R.J. *Neuron* **69**, 1204–1215 (2011).
14. Zhu, L., Mathewson, K.E. & Hsu, M. *Proc. Natl. Acad. Sci. USA* **109**, 1419–1424 (2012).
15. Diuk, C., Tsai, K., Wallis, J., Botvinick, M. & Niv, Y. *J. Neurosci.* **33**, 5797–5805 (2013).
16. Delgado, M.R., Frank, R.H. & Phelps, E.A. *Nat. Neurosci.* **8**, 1611–1618 (2005).
17. Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. & Cohen, J.D. *Science* **300**, 1755–1758 (2003).
18. Amodio, D.M. & Frith, C.D. *Nat. Rev. Neurosci.* **7**, 268–277 (2006).
19. Amodio, D.M. & Ratner, K.G. *Curr. Dir. Psychol. Sci.* **20**, 143–148 (2011).
20. Robalino, N. & Robson, A. *Phil. Trans. R. Soc. Lond. B* **367**, 2224–2233 (2012).

© 2015 Nature America, Inc. All rights reserved.

ONLINE METHODS

Participants. We scanned 31 healthy, right-handed participants (22 female, mean age = 22.26). This sample size is within the standard range for the field (for example, see refs. 13–15). One participant was excluded from fMRI analyses due to severe artifacts in reconstruction most likely due to subject motion, leading to a total sample of 30 subjects in fMRI analyses. Because of scanner gradient malfunctions during two other scanning sessions, test phase data were collected in one of four runs for one subject and three of four runs for another subject; all available runs were included in analysis. (Excluding the two subjects with missing test phase data from test phase analyses did not change results.) All behavioral data were collected and all 31 participants were included in behavioral analyses. Participants gave informed consent in accordance with approval from the NYU University Committee on Activities Involving Human Subjects.

Pre-task instructions. Upon arriving to the scan center for a study entitled "Social Learning & Choice," participants read the following instructions:

> This is a sequential study. Participants are randomly assigned to either the role of Player A (Social Choice) or Player B (Social Learning). Participants in the Player A role make decisions that affect themselves and future participants in the Player B role. Participants in the Player B role learn about the decisions of a few previous Player A participants.
> If you are assigned to Player A, we will take a photo of your face against a white background; this photo will then be shown to later Player B participants as they learn about your choices. If you are assigned to the Player B role, no photo will be taken and no other participants will learn about your responses today.
> If you are comfortable having your photo taken and shown to other participants, you will be randomly assigned to either the Player A role or the Player B role. If you are uncomfortable having your photo taken and shown to other participants, you will automatically be assigned to the Player B role, which involves no photographs.

Participants then indicated their preferences; in reality, all participants were assigned to the Player B role through a rigged drawing.

Stimuli. Participants viewed photos of four white male faces[21] who were described as previous participants from the Player A role. Gender and race were kept constant across the four photos to avoid any cues to social group membership that could influence social judgments. The four photos were randomly assigned to one of the generosity/reward cells shown in **Supplementary Figure 1a**, ensuring that any potential variability in facial features (for example, attractiveness) was randomly distributed across conditions and could not influence effects of interest. Participants also viewed images of four schematic slot machines with different colors; again, each color of slot machine was randomly assigned to a generosity/reward level across subjects.

Training phase. Participants viewed the following instructions for the training phase of the task:

> Player B Instructions: In this task, you will play a game where you learn about four previous participants assigned to the Player A role. In this game, each of the Player A participants made a series of choices about how to divide up a pool of points between themselves and a future Player B participant (i.e., you). At the end of the game, points get exchanged for money; once you complete the game, you will find out the exchange rate (i.e., how much each point is worth in cents). Press any key to see pictures of the four Player A participants that have been selected for you to play with.
> On each decision, we made a different pool of points available to each Player A. Player A participants chose how much of that point pool to keep for themselves, and how much to donate to you. Of the many participants run in this and similar studies, Player A participants have shown a range of responses: sometimes they have a large pool but only share a small proportion; sometimes, they have a small pool and share a large proportion. How much you get depends on both the pool available to a Player A on a given trial, and how much they chose to share with you. On average, we find that Player B participants get 5 to 25 points per trial, and Player A participants tend to share 20% to 40% of the pool on average. However, these amounts vary (sometimes more, sometimes less), so you will have to learn about that as you go.
> GAME INSTRUCTIONS: You will learn about the four previous Player A participants we have selected for you to play with (all 4 are selected to be the same gender and similar ages). You will see the faces of two out of four on screen on each round of the game. Your job is to choose whom to play with on each round. After you decide, you will see two pieces of information: a) how much they chose to share with you on that round (labeled "shared"), and b) the pool of points that was available to that Decider on that round (labeled "out of"). Over the course of all rounds, you will accumulate points based on your decisions; at the end of the game, points will be exchanged for cash, on top of base payment. In other words, how much you get paid will depend on whom you pick.
> In addition, half the time you will also play the same game with 4 computerized "slot machines" with different colors. As with the human players, you will pick a slot machine, and then find out a) how much it pays out to you ("payout") and b) the pool available for that slot machine on that trial ("out of"). The points you get and proportions paid from the slot machines have been made comparable to the average range we observe in the Player A participants (on average, 5 to 25 points, with around 20% to 40% of the pool paid out; again, this will vary). Press any key to see pictures of the four slot machines.
> In sum: on each trial, you will see two faces or slot machines on the screen and pick which to play with. Indicate your choice using your index and middle fingers. Press your index finger to select the left choice, and your middle finger to select the right choice.
> You will have 2 seconds to respond on each trial. You must respond in that time for your answer to be recorded and to potentially win money.
> If you have any questions about the instructions, how payment works, or the experiment in general, let an experimenter know. Otherwise, you're ready to begin. Again, use your index finger (left choice) and middle finger (right choice) to make your responses.

Participants then completed 168 trials of the training phase, composed of 84 human trials and 84 slot trials. Human and slot trials were pseudorandomly interleaved in a different order for each subject, with the exception that no more than 8 trials of one type were allowed in a row, to avoid filtering out potential condition effects with a high-pass filter during fMRI analysis. Trials were divided into three functional runs of 56 trials each.

On each trial, participants saw two human faces or two slot machines; each possible pair of faces and of slots was viewed 14 times. Each face or slot was equally likely to appear on each side of the screen. During a 2-s choice epoch, participants indicated a choice by pressing their index finger (left choice) or middle finger (right choice) on a button box. Feedback was displayed for 3 s indicating the number of points received (labeled "shared" for human trials and "payout" for slot trials) as well as the point pool that had been available for the human or slot target chosen (labeled "out of"). Feedback was followed by a jittered inter-trial interval (1–8 s) to allow estimation of the hemodynamic response across trials.

The reward and generosity values for each target were generated using the average values displayed in the main text plus noise with s.d. = 10 for reward and s.d. = 0.075 for generosity, with the limitation that reward value had to be at least 2 points and generosity had to be at least 0.01 to ensure meaningful values.

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nn.4080

Values followed a (censored) normal distribution. Pool values displayed during training phase feedback were determined by dividing reward by generosity. Values were rounded as follows: (i) noisy reward values were rounded to the nearest integer, (ii) pool amounts were generated as [(rounded reward value)/(generosity)] and (iii) pool amounts were rounded to the nearest integer.

The task was implemented using the Psychtoolbox[22–24] for Matlab.

Test phase. Participants viewed the following instructions:

> You will now complete four more rounds of this game. However, there will be three differences.
> First, the point pool available for each option will be shown above each picture before you make your choice, so you can use it to make your decisions. Second, you will not receive feedback about how much you get from each choice. At the end, you will be informed how much you won today. Finally, you will have 4 seconds to make each choice.
> If you have any questions about these instructions, let the experimenter know. To begin, press any key now. Again, just use your index finger and middle finger to respond.

Participants then completed 240 trials of the test phase, divided into 120 human trials and 120 slot trials. As in the training phase, human and slot trials were pseudorandomly interleaved with the restriction that no more than 8 trials of one type were allowed in a row. Trials were divided into four functional runs of 60 trials each.

On each trial, participants again saw two human faces or two slot machines; each possible pair of humans or slots appeared 20 times. Each face or slot was equally likely to appear on each side of the screen. Above each target, a point pool was displayed representing the number of points a human had available to share or a slot had available to pay out. To determine point pools, a random integer from 10 to 100 was generated as the pool value for one of the targets. Next, this amount was multiplied by one of seven ratios to determine the point pool for the second target. The ratios (0.33, 0.67, 0.9, 1, 1.11, 1.5, 3) were designed to be symmetric around 1, allowing fine-grained expression of generosity knowledge. Each face pair and slot pair appeared an equal number of times at each ratio besides 1; each pair appeared an extra 6 times at a 1:1 ratio (that is, each pair was seen twice at each other ratio, and eight times at a 1:1 ratio). Participants made choices in a 4-s decision epoch, again pressing their index or middle fingers on a button box to indicate left or right choices, followed by a jittered inter-trial interval of 1–8 s.

Post-scan ratings. Immediately upon exiting the scanner, participants were brought to a computer terminal for post-scan ratings. Participants were told that they might be invited back for a second session involving a cooperative puzzle-solving task with a partner, and that this task would not involve the possibility of a monetary bonus. Finally, participants were told that the experimenters would try to match them with a partner based on their preferences, at which point participants were asked to rate how much they would like to be paired with each target seen in the main task on a 7-point scale (1 = not at all, 7 = definitely yes). No post-scan ratings were collected for slot machines, in keeping with the cover story. Participants then completed demographic measures.

Following these measures, participants were informed of the exchange rate of points to money (20 points = \$0.01), paid their winnings (average bonus = \$2.80) in addition to their base remuneration for completing the experimental session (\$30), thanked and debriefed.

fMRI data acquisition. Images were acquired with a 3-T Siemens Allegra head-only scanner. Functional images (TR = 2,000 ms; effective TE = 30 ms; flip angle = 82, 34 3-mm slices with a 0.45-mm gap for whole-brain coverage, matrix = 80 × 64; FOV = 240 × 192 mm; acquisition voxel size = 3 × 3 × 3.45 mm) were acquired using a customized multi-echo EPI sequence developed by the NYU Center for Brain Imaging to mitigate the effects of susceptibility artifacts in medial temporal and ventromedial regions. This sequence uses the dead time that precedes the readout in EPI sequences with normal TEs (~30 ms) to collect two more images at several echo times. The first is a low-resolution, fully sampled image (matrix size 80 × 16) with an echo time of 7.25 ms; the second is a partial-Fourier (5/8) image at full resolution with an echo time of 11.9 ms and with reversed phase-encoding gradients; and the third is a conventional fully sampled image with echo time of 32.05 ms. This scheme allows optimal detection of signal from fast decay (short T2*) regions and slower decay (long T2*) regions. Using the information from all three images, the raw data are fitted using an in-house algorithm developed in Matlab to obtain a spin density (ρ), decay rate (R2* = 1/T2*) and B_0 field maps for each acquired volume. The final image for each repetition is a weighted average of the signal ($\rho e^{-t/T2^*}$) from each voxel across the readout. The collection of one of the images with reversed phase-encoding gradient allows correction for the susceptibility artifacts even in the presence of significant subject motion during the run.

Five fixation scans were acquired at the start of each run and dropped from analysis to allow magnet equilibration. Slices were collected parallel to the AC–PC line. Finally, T1-weighted high-resolution anatomical images (MPRAGE, 1 × 1 × 1 mm) were acquired for each subject for registration and group normalization purposes.

Computational model. The task consisted of two stages: a training phase, in which participants learned about targets, and a test phase, in which they could express this knowledge. We adapted reinforcement models to this task using a hybrid reward and generosity learning model. Specifically, we use a model-free algorithm because each trial contained only one stage; although model-free versus model-based processing was not the theoretical concern of this study, future work could use sequential tasks to parse model-free and model-based processes, both of which may contribute to learning[25].

In the training phase, the hybrid model assumes that participants update a reward value Q for the chosen target on each trial t according to

$$Q_t = Q_{t-1} + \alpha\delta_{Rt} \tag{1}$$

where α is a free parameter representing a learning rate and δ_{Rt} is a reward prediction error defined as the difference between rewards received and rewards expected:

$$\delta_{Rt} = \text{Reward}_t - Q_{t-1} \tag{2}$$

Participants were informed in the instructions that players in their role tend to earn 5 to 25 points per trial on average; therefore, reward values were initialized to the mean of this range (15 points).

To allow generosity learning, the model allowed participants to update a generosity estimate G on each trial t according to

$$G_t = G_{t-1} + \alpha\delta_{Gt} \tag{3}$$

where δ_{Gt} represents a generosity prediction error defined as the difference between actual generosity and generosity expected:

$$\delta_{Gt} = \text{Generosity}_t - G_{t-1} \tag{4}$$

Generosity was defined quantitatively as the proportion of available points shared (by human players) or paid out (by slot machines). Since participants were informed that targets tend to share 20–40% per trial on average, generosity expectations were initialized to the mean of this range (0.30). We considered a model that included a unique learning rate for generosity, but this model did not provide a significantly better fit than using one learning rate. We therefore used one learning rate to stabilize parameter estimates.

An expected value based on generosity was defined as

$$GV = G \times \text{Pool} \tag{5}$$

where pool is an estimate of the number of points available. In training, since reward was initialized to 15 points and generosity was initialized to 0.30, pool expectations were set to 50 points (15/0.30). We considered a model in which pool estimates were allowed to update, but it provided a worse fit to the data despite an equivalent number of parameters.

The model allowed integration of generosity-based values and reward-based values according to

$$EV = w(GV) + (1 - w)Q \tag{6}$$

© 2015 Nature America, Inc. All rights reserved.

where w is a weighting parameter indicating how much participants rely on generosity values or reward values. A participant who fully relies on generosity values would have a weighting parameter $w = 1$, while a participant who fully relies on reward values would have a weighting parameter $w = 0$.

Finally, participant choices were modeled using a softmax choice function:

$$p_{i,t} = \frac{\exp\left(\beta_{\text{Train}} \times EV_{i,t}\right)}{\sum_j \exp\left(\beta_{\text{Train}} \times EV_{j,t}\right)} \quad (7)$$

where β_{Train} is an exploration parameter controlling stochasticity of choice in the training phase and $p_{i,t}$ is the probability of choosing option i (of j options) on trial t.

To model test phase behavior, final training phase estimates of Q and G were carried forward. Again, the model assumed that an integrated expected value was computed in the same manner as in equations (5) and (6), except that pool amounts were displayed on-screen instead of estimated from the average range. Additionally, choice was modeled using the softmax function in equation (7) using exploration parameter β_{Test}.

Altogether this hybrid model had four free parameters: α, w, β_{Train} and β_{Test} (**Supplementary Table 1**). Parameters were estimated using maximum a posteriori (MAP) estimation to optimize parameters across all choices, using priors of Gamma(1.2, scale = 5) applied to exploration parameters and Beta(1.1, 1.1) applied to learning rates and the weighting parameter[13,26]. We compared this hybrid model (model 1) to a set of simpler models using Bayesian model comparison:

Model 2. Reward only ($w = 0$). This model constrained the weighting parameter to zero, rendering it equivalent to a classic Q-learning model that assumes people choose on the basis only of cached values in training and test.

Model 3. Generosity only ($w = 1$). This model constrained the weighting parameter to one. This model is equivalent to pure generosity learning and generosity-based choice.

Model 4. Optimal choice ($w = 0$ in training, $w = 1$ in test). This model tested whether people followed the optimal strategy (described in the main text) of choosing on the basis of reward value in training and generosity × pool in test, by constraining w to 0 in training and to 1 in test.

Model 5. Reward in training, pool in test. This model provided a more realistic alternative to model 2 by allowing people to choose on the basis of reward values in training and pool values at test without having learned generosity. This model again constrained $w = 0$ in training and $w = 1$ in test, but enforced a generosity learning rate of zero. Therefore, test phase choice was determined by pool × average generosity with no generosity learning.

For each model, the best-fitting parameters were used to compute the Laplace approximation to the Bayesian model evidence[27]. Model evidence was compared using fixed-effects methods (Wilcoxon signed rank tests owing to non-Gaussian distributions; group log Bayes factor) and random-effects methods (Bayesian model selection using the spm_BMS function in Statistical Parametric Mapping software version 8 (SPM8; Wellcome Trust Center for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/))[28]. The hybrid model (model 1) had the best fit to the data across all metrics (**Supplementary Table 2**).

To generate time series for fMRI analysis of the training phase, the model was run forward for each participant using the mean of the best-fitting parameters across all subjects. This process has been used to stabilize noisy parameter estimates[29], and the mean of individual parameters provides an estimate of population parameters[30]. This process generated time series of reward prediction errors and generosity prediction errors in the training phase.

To simulate different strategies in the training phase, the hybrid model was simulated using the average best-fitting learning rate and temperature parameters. In different iterations, the weighting parameter w was set to 0 (pure reliance on reward value), 0.5 (equal reliance on reward value and generosity value) and 1 (pure reliance on generosity value). A total of 10,000 simulations were run for each w setting, and the mean and s.d. of points accumulated were recorded for each.

We analyzed participants' w parameter estimates using Wilcoxon sign rank tests, which did not assume normality of data.

fMRI data analysis. Preprocessing and analysis of fMRI data were performed using SPM8. Functional images were corrected for slice time acquisition, realigned to adjust for participant motion, co-registered to each subject's high-resolution anatomical image using linear transformations, normalized to MNI space using nonlinear transformation, resampled to 2 × 2 × 2 mm voxels and spatially smoothed using a 6-mm full-width at half-maximum Gaussian filter.

A general linear model (GLM) for the training phase included (1) onset of feedback on human trials and (2) onset of feedback on slot trials. Each feedback onset was modeled as an impulse and was parametrically modulated by a time series representing reward prediction errors and a time series representing generosity prediction errors (described above). Reward prediction errors and generosity prediction errors were orthogonal on average by design (human trials: average correlation = −0.005; slot trials: average correlation = 0.004), ensuring that any effects of generosity prediction error were above and beyond classic reward prediction error effects. SPM's default serial orthogonalization of parametric modulators was not used, so that effects would not depend on the order in which regressors were entered.

Further regressors of no interest included (3) onsets of choice epochs for human trials and (4) onsets of choice epochs for slot trials (both lasting the duration of the epoch), (5) onsets of choice and (6) feedback epochs for missed trials (both lasting the duration of the epoch) and (7–12) the six motion parameters obtained during realignment. A high-pass filter with cutoff period of 128 s was used.

First-level contrasts for each of the two parametric regressors (reward prediction error and generosity prediction error) were generated and entered into a second-level random effects analysis. Unless otherwise noted, all whole-brain analyses were corrected for multiple comparisons using a voxel-wise height threshold of $P < 0.005$ combined with an appropriate cluster extent to maintain a family-wise error (FWE) rate of $P < 0.05$, using Gaussian random field theory as implemented in SPM[31]. For the contrast of generosity prediction errors, the minimum cluster size as determined by SPM's defaults was 258 voxels. (A cluster of 226 voxels in left ventral striatum did not survive whole-brain correction for multiple comparisons.) We did not find significant activation negatively associated with generosity prediction errors.

Given strong a priori hypotheses about reward prediction errors in ventral striatum, reward prediction errors were examined using a small-volume correction within an anatomical mask of ventral striatum. This mask was generated by obtaining a mask of the caudate and putamen from the Automated Anatomical Labeling atlas[32] and maintaining only ventral portions, defined as $z < 2$ for putamen and $z < 7$ for caudate[33]. The mask included 981 voxels, and within this mask, a cluster of 161 voxels was deemed significant in SPM's small-volume correction procedure. For this contrast, we observed a cluster of 18 voxels in left ventral striatum that did not survive small-volume correction for multiple comparisons.

Since the test of reward prediction errors was more lenient, we tested whether any regions associated with generosity prediction errors also showed effects of reward prediction errors using small-volume corrections, to test that generosity prediction errors did recruit a unique network. Within each region significantly activated by generosity prediction errors, only ventral striatum showed significant activity correlating with reward prediction errors, suggesting that activity in the other regions was unique to generosity prediction errors. As a second test, we extracted average signal change from within each of these regions using a region of interest (ROI) approach (as opposed to small-volume correction); again, none of these regions other than ventral striatum correlated significantly with reward prediction errors. Finally, since we observed a cluster of 18 voxels (at voxel-wise $P < 0.005$) in left ventral striatum for reward prediction errors, we examined activation for reward prediction errors within regions associated with generosity prediction errors at this uncorrected threshold; again, we did not observe activation in other regions associated with generosity prediction errors.

We also tested for unsigned (absolute) prediction errors across reward and generosity. For reward, we found activation in regions previously associated with responses to surprising outcomes[34,35], including right inferior frontal gyrus, middle frontal gyrus and posterior parietal cortex (**Supplementary Table 3**). No regions showed decreasing activation with unsigned prediction errors (indicating certainty). For generosity, we did not observe significant activations

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nn.4080

to unsigned prediction errors, and only a cluster along superior occipital gyrus showed decreasing activity with unsigned prediction errors.

For the test phase, a GLM included (1) onsets for the decision phase of human trials and (2) onsets for the decision phase of slot trials. Trials were modeled as epochs lasting for the length of each trial's reaction time (that is, events began at trial onset and ended at the time a decision was entered). Each trial was modeled with six parametric regressors:

i. Point pool difference (chosen – unchosen options)
ii. Point pool average (chosen, unchosen options)
iii. Generosity difference (chosen – unchosen options)
iv. Generosity average (chosen, unchosen options)
v. Reward value difference (chosen – unchosen options)
vi. Reward value average (chosen, unchosen options)

The model therefore accounted for baseline levels of each variable as well as the difference between chosen and unchosen options. We used true underlying generosity and reward values, which allows the GLM weighting to reflect individual differences.

Pool differences (chosen minus unchosen values) were correlated with the average pool amounts as a result of the task design because pool amounts were determined by assigning a random integer to one option and multiplying this amount by one of a set of ratios to generate the second option. For this reason, SPM's default orthogonalization of parametric regressors was not used; orthogonalizing regressors would incorrectly assign set-level pool variance to the pool difference regressor. For example, a choice set with pool values of 90, 60 has a higher set-level value than a choice set with pool values 60, 40, and this variance should not be assigned to the regressor representing the difference. Avoiding orthogonalization therefore allowed a more appropriate test of the difference between chosen and unchosen.

Regressors of no interest included (3) onsets of choice epochs for missed trials (lasting the duration of the epoch) and (4–9) the six motion parameters obtained during realignment. A high-pass filter with a cutoff period of 128 s was used.

Since point pools displayed onscreen straightforwardly predicted differences in expected reward (mean correlation = 0.73), we created an ROI in vmPFC based on the pool point differences. First-level contrasts were generated for the regressor representing the difference of chosen minus unchosen pool amounts and were entered into a second-level random effects analysis. The minimum cluster size for a whole-brain correction was 281 voxels, based on SPM's defaults (as described above). We defined a functional ROI corresponding to the pool difference within vmPFC and extracted parameter estimates within this ROI for the contrast of the difference between chosen and unchosen generosity amounts, using the MarsBar toolbox[33].

If pool differences and generosity differences had been positively correlated, an ROI based on pool signals could provide a biased test of whether this region correlates with generosity. Two factors preclude this concern. First, the two regressors were not orthogonalized when they were simultaneously entered into the GLM, and so each should reflect unique variance. Second, the average correlations between pool differences and generosity differences were in the negative direction (mean correlation = – 0.15 for human trials, –0.10 for slot trials). Therefore, any residual shared variance would, if anything, bias the test against our hypothesis, making the test of positive generosity correlates in vmPFC more conservative. **Supplementary Table 7** displays the average correlations between all regressors in this analysis.

However, to further provide independence, we conducted a leave-one-subject-out cross-validation[36]. Specifically, we computed 30 GLMs, each leaving out one subject. For each GLM, we defined an ROI based on vmPFC activation for point pools at the group level. (Because each analysis included fewer subjects and was only used to define a region for an independent test, we used a slightly more liberal, uncorrected statistical threshold of $P < 0.005$ per voxel with a cluster extent of at least 200 voxels; see ref. 36 for a discussion of why this selection method does not lead to biased results.) We then used these ROIs to extract parameter estimates for generosity for the left-out subject. All results remained identical: we found a main effect of generosity differences, $t(29) = 3.00$, $P = 0.005$; a correlation between vmPFC and the w parameter, $r = 0.35$, $P = 0.03$ (one-tailed) and a correlation between vmPFC activation and the post-task index, $r = 0.50$, $P = 0.005$. To test robustness to assumptions of normality in Fisher's test for the correlations, we also computed bootstrap confidence intervals, which provided identical results.

We did not find significant neural correlates of the difference in prior reward value (chosen – unchosen) within the vmPFC ROI, and no regions survived correction for multiple comparisons in a whole-brain analysis. However, we did find that the average prior reward value of the chosen and unchosen targets was correlated with activation in a cluster extending across hippocampus and parahippocampal gyrus, posterior putamen and insula (**Supplementary Fig. 2** and **Supplementary Table 6**).

Test phase behavioral analyses. Test phase reaction times (RTs) were analyzed using multilevel modeling, which we used to predict trial-by-trial responses. As in test phase fMRI analyses, predictors included the difference between chosen and unchosen values of (1) displayed pool amounts, (2) generosity and (3) prior reward value. Fixed effects and random effects were included for all predictors, and all continuous predictors were z-scored. Target type was entered as a categorical predictor (–1 = human, 1 = slot). We also included the interactions of the generosity, prior reward value and target type terms. For the test of choice behavior, we used a multilevel logistic model that predicted the probability of choosing the target on the right as a function of the difference of values (right – left) for (1) pool, (2) generosity and (3) prior reward, again z-scored, again including target type and interactions, and again including fixed effects and random effects for each coefficient. Both analyses used objective underlying reward and generosity values to provide converging evidence to the computational model.

For both analyses, each continuous predictor was subject-mean centered, and analyses were performed using the lme4 and lmerTest packages for R[37–39]. To compute the predicted reaction times, log odds of choice and standard errors for high and low generosity and reward shown in **Figure 2**, fixed-effects predictors for reward and generosity were recentered 1 s.d. above and below the mean for each individual, the model was recomputed, and simple intercepts and their standard errors were recorded for each recentered model[40]. Pool amounts remained mean-centered to compute predicted values at different generosity and reward levels while accounting for pool. Reaction times less than 200 ms were excluded from analysis; removing this exclusion rule did not change results. Additionally, log-transforming reaction times for normality in the linear regression led to identical results. (Logistic regression does not assume normality of data.)

Next, to test the relative contributions of generosity and reward value to behavior, we contrasted the coefficients for each predictor against each other[41]. Since predictors were z-scored, β weights could be meaningfully compared without concerns about differences in units. For reaction times, since the two coefficients had slightly different degrees of freedom, we used the smaller of the two to provide a more conservative test.

In addition to effects reported in the main text, we report several related effects of interest in **Supplementary Tables 4** and **5**, including main effects of pool values, target type × generosity interactions, and simple effects for generosity and reward across target types. We analyzed simple effects by dummy coding the target type variable such that zero corresponded to either humans or slots.

Post-task analyses. To analyze post-scan social valuations, ratings for the four human targets were entered into a 2 × 2 repeated measures ANOVA (generosity: high, low × reward: high, low). Note that our approach here deviates from the multilevel regression approach we used elsewhere because participants made only one post-task rating for each unique combination of reward and generosity level. Means and s.d. are reported in **Supplementary Table 8**. We did not test for normality of data because ANOVA is generally robust to violations of this assumption[42].

Next we computed an index of individual sensitivity to generosity versus reward as follows. First, we computed a generosity sensitivity measure collapsing across reward value as

[average ratings for high generosity targets]
– [average ratings for low generosity targets]

© 2015 Nature America, Inc. All rights reserved.

We similarly computed a reward sensitivity measure collapsing across generosity levels as

[average ratings for high reward targets]
– [average ratings for low reward targets]

Finally, we computed a differential sensitivity measure as

[generosity sensitivity – reward sensitivity]

Note that differential sensitivity is mathematically equivalent to subtracting

[ratings for high generosity, low reward target]
– [ratings for low generosity, high reward target]

To test the relative strength of generosity versus reward on average, we conducted an F contrast of generosity sensitivity versus reward sensitivity; this test is equivalent to a one-sample t-test against zero on the differential sensitivity measure. Finally, to test robustness to assumptions of normality in Fisher's test for the correlation between the w parameter and the post-task index, we also used bootstrap confidence intervals, which provided identical results.

A **Supplementary Methods Checklist** is available.

21. Minear, M. & Park, D.C. *Behav. Res. Methods Instrum. Comput.* **36**, 630–633 (2004).
22. Brainard, D.H. *Spat. Vis.* **10**, 433–436 (1997).
23. Pelli, D.G. *Spat. Vis.* **10**, 437–442 (1997).
24. Kleiner, M., Brainard, D. & Pelli, D. *Perception* **36** (ECVP Abstract Supplement) (2007).
25. Doll, B.B., Simon, D.A. & Daw, N.D. *Curr. Opin. Neurobiol.* **22**, 1075–1081 (2012).
26. Decker, J.H., Lourenco, F.S., Doll, B.B. & Hartley, C.A. *Cogn. Affect. Behav. Neurosci.* **15**, 310–320 (2015).
27. Daw, N.D. *Decision Making, Affect, and Learning: Attention and Performance* **23**, 3–38 (2011).
28. Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J. & Friston, K.J. *Neuroimage* **46**, 1004–1017 (2009).
29. Wimmer, G.E., Daw, N.D. & Shohamy, D. *Eur. J. Neurosci.* **35**, 1092–1104 (2012).
30. Holmes, A.P. & Friston, K.J. *Neuroimage* **7**, S754 (1998).
31. Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C. & Evans, A.C. *Hum. Brain Mapp.* **1**, 210–220 (1994).
32. Tzourio-Mazoyer, N. *et al. Neuroimage* **15**, 273–289 (2002).
33. Brett, M., Anton, J.-L., Valabregue, R. & Poline, J.-B. *Neuroimage* **16**, S497 (2002).
34. d'Acremont, M., Schultz, W. & Bossaerts, P. *J. Neurosci.* **33**, 10887–10897 (2013).
35. O'Reilly, J.X. *et al. Proc. Natl. Acad. Sci. USA* **110**, E3660–E3669 (2013).
36. Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.-C. & Yantis, S. *Neuroimage* **50**, 572–576 (2010).
37. R Core Team http://cran.r-project.org/ (2014).
38. Bates, D., Maechler, M., Bolker, B. & Walker, S. *R Package Version 1.1-7* http://cran.r-project.org/package=lme4 (2014).
39. Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. *R Package Version 2.0-11* http://cran.r-project.org/package=lmerTest (2014).
40. Aiken, L.S. & West, S.G. *Multiple Regression: Testing and Interpreting Interactions* (Sage, 1991).
41. Cohen, J., Cohen, P., West, S.G. & Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Routledge, 2013).
42. Howell, D. *Statistical Methods for Psychology* (Cengage Learning, Belmont, California, USA, 2010).

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nn.4080