

Human fronto–mesolimbic networks guide decisions about charitable donation

Jorge Moll^{*†}, Frank Krueger^{*}, Roland Zahn^{*}, Matteo Pardini^{*‡}, Ricardo de Oliveira-Souza^{†§}, and Jordan Grafman^{*†1}

^{*}Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892-1440; [†]Cognitive and Behavioral Neuroscience Unit, LABS-D'Or Hospital Network, 2228-080, Rio de Janeiro, Brazil; [‡]University of Genoa Medical School, 16132 Genoa, Italy; and [§]Gaffrée and Guinle University Hospital and Philippe Pinel Institute, 20270-004 Rio de Janeiro, Brazil

Edited by Marcus E. Raichle, Washington University School of Medicine, St. Louis, MO, and approved September 7, 2006 (received for review May 31, 2006)

Humans often sacrifice material benefits to endorse or to oppose societal causes based on moral beliefs. Charitable donation behavior, which has been the target of recent experimental economics studies, is an outstanding contemporary manifestation of this ability. Yet the neural bases of this unique aspect of human altruism, which extends beyond interpersonal interactions, remain obscure. In this article, we use functional magnetic resonance imaging while participants anonymously donated to or opposed real charitable organizations related to major societal causes. We show that the mesolimbic reward system is engaged by donations in the same way as when monetary rewards are obtained. Furthermore, medial orbitofrontal–subgenual and lateral orbitofrontal areas, which also play key roles in more primitive mechanisms of social attachment and aversion, specifically mediate decisions to donate or to oppose societal causes. Remarkably, more anterior sectors of the prefrontal cortex are distinctively recruited when altruistic choices prevail over selfish material interests.

altruism | brain | moral | reward | social

Human altruism far exceeds the immediate bonds of kinship, even when no material or reputation gains are anticipated (1, 2). Recent studies in experimental economics have started to explore the neurobiological basis of cooperation in interpersonal exchanges (3–5). Altruistic choices regularly take place beyond interpersonal and economic realms, however. People often sacrifice material interests, time, and even physical integrity on behalf of societal causes, principles, and ideologies (6–8). Anonymous donation to charitable organizations is an outstanding example of this unique aspect of human altruism (8, 9), which relies on our ability to directly link motivational significance to abstract moral beliefs and societal causes. Evolutionary and neurobiological theories suggest that this ability was critically shaped during the last major step of human evolution in the cultural explosion of the Upper Paleolithic period (10, 11).

We investigated the neural mechanisms of charitable donations using functional magnetic resonance imaging (fMRI; see *Materials and Methods*). Nineteen participants chose to endorse or oppose societal causes by anonymous decisions to donate or refrain from donating to real charitable organizations (ORGs). The ORGs' missions were linked to a wide range of societal causes, including abortion, children rights, death penalty, euthanasia, gender equality, nuclear power, and war (see *Materials and Methods*). Importantly, the experimental design allowed us to probe the interplay of material interests and altruistic preferences. Participants were entitled to receive a substantial endowment of U.S.\$128, which would be obtained in full if they solely cared about their self monetary interests when making decisions. This amount would decrease, depending on how often they made altruistic choices (see *Materials and Methods* for an operational definition of altruistic decisions).

The experimental conditions of interest, defined on the basis of participants' "Yes" or "No" decisions to different payoff types, were: (i) pure monetary reward, (ii) noncostly donation, (iii) noncostly opposition, (iv) costly donation, and (v) costly

opposition (Fig. 1; see *Materials and Methods* and Figs. 5–8, which are published as supporting information on the PNAS web site, for details on stimuli and task procedure). Thus, although some decisions involved pure monetary rewards, and donation or opposition at no personal costs, other decisions entailed a conflict between participants' personal monetary interests and their motivations to donate to or to oppose causes. Importantly, ORGs were paired with randomized payoff types, and all ORGs were presented to each participant. At the end of the experiment, all ORGs and their causes were scored according to familiarity and associated moral emotion (compassion and anger; see *Supporting Methods*, which is published as supporting information on the PNAS web site). In addition, self-reported ratings of engagement in real-life voluntary charitable activities were obtained.

As long as humans can derive utility directly from the act of alleviating the suffering of another (8, 12), we predicted activation of the mesolimbic reward system both for decisions leading to pure monetary rewards and decisions to donate. We also expected that medial and lateral sectors of the orbitofrontal cortex, respectively, would mediate decisions to donate or to oppose causes, in line with the involvement of these regions in reward and punishment (13). Finally, we predicted that anterior prefrontal regions that have been implicated in moral judgments and prospective assessment of outcomes (11, 14–16) would be engaged by altruistic decisions that involved sacrificing material interests for societal causes.

Results

Behavioral analyses showed that all participants consistently made costly decisions, sacrificing an average of 40% (U.S.\$51; range = U.S.\$21–80) of their endowment. Participants took longer to make costly than noncostly decisions (Fig. 2*a*). Consistent with the role of moral emotions in judgment (11) and in helping behaviors (17), ratings of experienced compassion were higher for causes participants chose to donate to, whereas anger scores were higher for opposed causes (Fig. 2*b*). For details of response times and emotion scores across the main experimental conditions, see Table 1, which is published as supporting information on the PNAS web site.

Author contributions: J.M. and J.G. designed research; J.M. and M.P. performed research; J.M., F.K., and R.Z. analyzed data; and J.M., F.K., R.Z., R.d.O.-S., and J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: ORGs, charitable organizations; VTA, ventral tegmental area; BA, Brodmann's area; IOFC, lateral orbitofrontal cortex; fMRI, functional magnetic resonance imaging.

Data deposition: The neuroimaging data have been deposited with the fMRI Data Center, www.fmridc.org (accession no. 2-2006-122A7).

¹To whom correspondence should be addressed at: Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Building 10, Room 5C205, MSC 1440, Bethesda, MD 20892-1440. E-mail: grafmanj@ninds.nih.gov.

© 2006 by The National Academy of Sciences of the USA

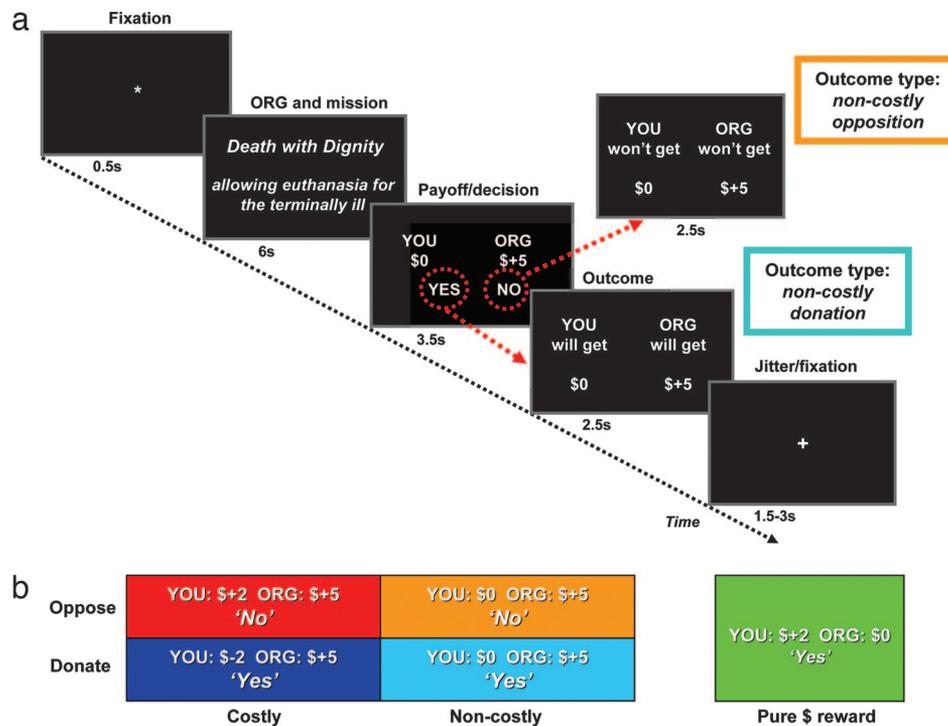


Fig. 1. Donation task and behavioral results. (a) Task design. The name and mission statement of an ORG was presented, followed by the payoff type (decision phase), and then by the outcome phase. Depending on the trial, Yes or No decisions to different payoff types had different monetary consequences to the participant and/or to the ORG ("outcome types"; see *Materials and Methods* and Figs. 5–8). In this example, a (YOU: \$0 ORG: \$+5) payoff is shown. (b) The conditions of interest derived from the main outcome types and comprised costly opposition [red; No to (YOU: \$+2 ORG: \$+5)], noncostly opposition [orange; No to (YOU: \$0 ORG: \$+5)], costly donation [dark blue; Yes to (YOU: \$-2 ORG: \$+5)], noncostly donation [light blue; Yes to (YOU: \$0 ORG: \$+5)], and pure monetary reward [green; Yes to (YOU: \$+2 ORG: \$0)]. Altruistic or costly decisions included costly donation and costly opposition.

The midbrain ventral tegmental area (VTA), the dorsal striatum, and the ventral striatum were activated by both pure monetary rewards and decisions to donate (Fig. 3a; see Table 2, which is published as supporting information on the PNAS web

site), suggesting that donating to societal causes and earning money share anatomical systems of reward reinforcement and expectancy (18, 19). This finding is compatible with the putative role of the "warm glow" ("joy of giving") effect, the rewarding experience associated with anonymous donations (8). But are the neural correlates of monetary rewards and donations identical? To address this issue, we directly compared donation conditions (costly and noncostly) to pure monetary reward. This contrast revealed that activity in the subgenual area [including Brodmann's area (BA) 25] was highly specific for donations (Fig. 3b and c; see Table 3, which is published as supporting information on the PNAS web site). Interestingly, the ventral striatum (together with the adjoining septal region) was also more active for donations than for pure monetary rewards. Furthermore, ventral striatum activity was correlated with the number of decisions to donate that each participant made during the experiment (Fig. 3d; see *Supporting Methods*). These findings indicate that donating to societal causes recruited two types of reward systems: the VTA–striatum mesolimbic network, which also was involved in pure monetary rewards, and the subgenual area, which was specific for donations and plays key roles in social attachment and affiliative reward mechanisms in humans (20, 21) and other animals (22).

Although morality often promotes cooperation and helping, it also can steer hostility among individuals and social groups. Moral beliefs powerfully incite people to challenge others' values and ideologies (6, 7). Previous research consistently implicates the lateral orbitofrontal cortex (lOFC) in aversive mechanisms (13), including anger and moral disgust (11, 23). To test the role of the lOFC in more abstract forms of culturally mediated social disapproval, brain responses to participants' decisions to oppose causes were compared with pure monetary rewards. The lOFC (BA 11/47), including its transition to the anterior insula and

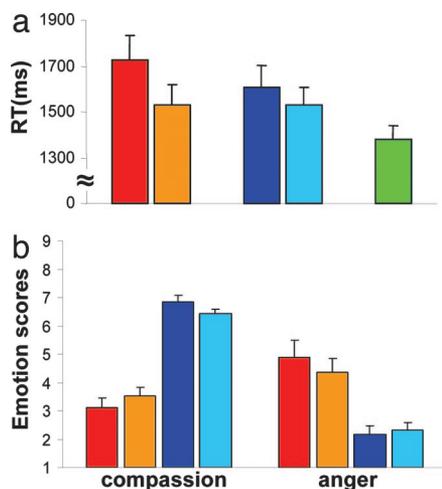


Fig. 2. Behavioral results. (a) Response times for main conditions of interest. Increased response times were associated with altruistic decisions (which included costly donation and costly opposition) as compared with noncostly decisions [$t(18) = 3.26$, $P < 0.005$]. Color code: costly opposition, red; noncostly opposition, orange; costly donation, dark blue; noncostly donation, light blue; and pure monetary reward, green. (b) Compassion and anger scores. Societal causes participants donated to received higher compassion scores [$t(16) = 7.84$, $P < 0.001$], whereas opposed causes scored higher in anger [$t(16) = 5.53$, $P < 0.001$]. Color code as the same as in a. Error bars indicate SEM.

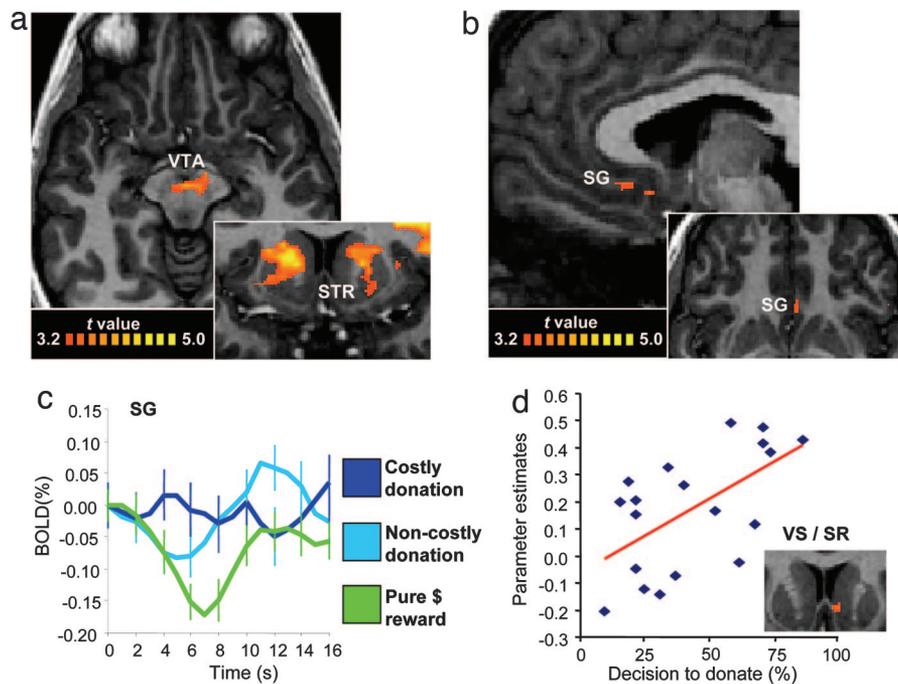


Fig. 3. Brain responses for monetary reward and donation. (a) Mesolimbic–striatal reward system, including the VTA and the dorsal and ventral sectors of the striatum (STR), activation for both pure monetary reward and noncostly donation (conjunction of pure reward vs. baseline and noncostly donation vs. baseline). (b) Subgenual area (SG) activation for decisions to donate (conjunction of costly and noncostly conditions) as compared with pure monetary reward. The subgenual area comprised the most posterior sector of the medial orbitofrontal cortex and the ventral cingulate cortex (BA 25) and the adjoining septal region structures. (c) Hemodynamic responses from the subgenual cortex for donation and pure monetary reward conditions. (d) Positive association between decision frequency of costly donation (how often each participant made costly donations) and ventral striatum/septal region parameter estimates (VS/SR; $x = -6$, $y = 11$, $z = 4$; $r = 0.58$; $P < 0.01$). BOLD, blood oxygenation level-dependent.

adjacent dorsolateral cortex, was activated by both costly and noncostly opposition (Fig. 4*a* and *c*; see Table 4, which is published as supporting information on the PNAS web site). Moreover, activity in the IOFC was modulated by how often participants decided to oppose societal causes (see *Supporting Methods*).

Decision making in real social environments requires balancing immediate motives against the long-term consequences of one's choices (11, 24). Previous work implicated anterior regions of the medial prefrontal cortex in goal representation (13–15, 25), altruistic punishment (5), prediction of future rewards (14, 18, 26), and implicit or explicit moral appraisals (11, 16, 27, 28). Our results indeed showed that costly decisions (choosing to costly donate or to costly oppose), which are altruistic in essence, were associated with activation of the anterior prefrontal cortex, including the frontopolar cortex and the medial frontal gyrus (BA 10/11/32; Fig. 4*b* and *d*; see Table 5, which is published as supporting information on the PNAS web site). Response time differences between costly and noncostly decisions did not correlate with anterior prefrontal activity, ruling out the possibility that these effects merely reflected decision difficulty. In contrast, the dorsal anterior cingulate cortex response, also observed for costly decisions, was correlated with response times, in agreement with its role in conflict and error monitoring (16, 29) (Fig. 4*b*). Finally, we probed the relationships between individual differences of self-reported engagement in real-life voluntary activities and brain activation patterns. Anterior prefrontal cortex activity to costly donation was highly correlated with engagement scores (Fig. 4*e*). This finding indicates that this region plays a key role in real-life altruistic behaviors, as suggested by a recent model of moral cognition (11).

Discussion

Our findings extend our knowledge of the neural bases of social cooperation from interpersonal economic interactions (3–5, 27,

30) to the realm of societal causes that are linked to culturally shaped moral beliefs. More specifically, they indicate that distinct neural systems underlie decisions to donate or to oppose societal causes: the mesolimbic reward system (VTA–striatum) provides a general reinforcement mechanism, the subgenual area and the IOFC mediate social attachment and aversion responses, and the anterior prefrontal cortex is crucial for representing more complex reinforcement contingencies related to altruistic decisions.

The importance of these fronto–limbic networks for human altruism concurs with their key roles in more basic social and motivational mechanisms. The mesolimbic system regulates overall reward reinforcement and prediction and is activated by a host of stimuli, including food, sex, drugs, and money (11, 14, 18). The subgenual area, which specifically was recruited by donations, comprises a primitive paralimbic, four-layered archicortex densely interconnected with the mesolimbic dopaminergic and dorsal raphe serotonergic pathways (31). This region plays a key role in controlling septo–hypothalamic function in social attachment and the release of the neuromodulators oxytocin and vasopressin (22, 32). Interestingly, recent studies showed that administration of oxytocin to humans increased trust and cooperation in economic interactions (32, 33). Further, the subgenual cortex and adjacent septal structures were activated when humans looked at their own babies and romantic partners (20, 21). The partially dissociable responses observed in the mesolimbic system and subgenual area indicate the existence of interlocking systems for self-serving monetary rewards and attachment to societal causes. Activity in the IOFC, in turn, was linked to opposing causes, a finding that converges with the role of this region in social aversion (3, 11, 23). This pattern is in general agreement with the medial to lateral functional specialization of the orbitofrontal cortex in reward and punishment

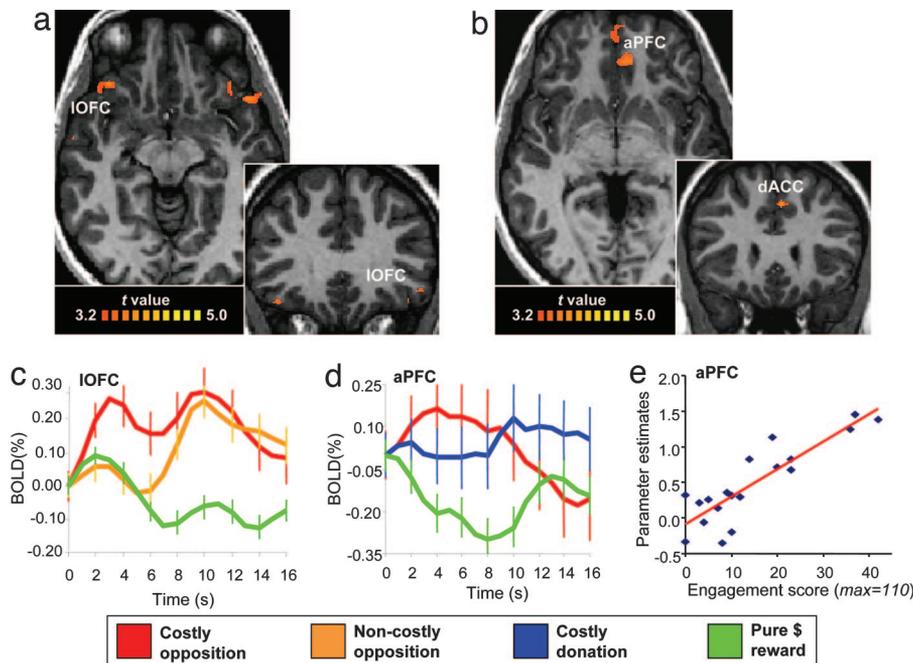


Fig. 4. Brain responses for opposition and costly decisions. (a) IOFC (BA 11/47) responses to decisions to oppose causes as compared with decisions involving pure monetary reward (conjunction of costly and noncostly conditions). Activity of the IOFC was modulated by decision frequency of costly opposition (peak: $x = -27, y = 35, z = -5; r = 0.76; P < 0.001$). (b) Comparison of costly decisions (sacrificing money either to donate or to oppose causes) to pure monetary rewards. Effects were observed in the anterior prefrontal cortex (aPFC), including the frontopolar cortex and the medial frontal gyrus (BA 10/11/32), and in the dorsal anterior cingulate cortex (dACC). The differences in response times between costly and noncostly decisions were correlated with parameter estimates from the dACC ($r = 0.46; P < 0.05$) but not from the frontopolar cortex ($r = -0.15; P = 0.53$) and the medial frontal gyrus ($r = -0.07; P = 0.75$). (c) Hemodynamic responses from the left IOFC for opposing causes. (d) Hemodynamic responses from the frontopolar cortex for costly decisions. (e) Relationship between self-reported engagement in real-life voluntary activities and aPFC activity to costly donation (peak: $x = -6, y = 25, z = -14; r = 0.87; P < 0.0001$). BOLD, blood oxygenation level-dependent.

representations (13, 34). We speculate that our capacity to feel attachment or aversion to societal causes might have emerged through similar gene-culture coevolution mechanisms as those proposed by the strong reciprocity theory (35). This premise would allow primitive reward, social attachment, and aversion neural systems to operate beyond the immediate spheres of kinship, thus enabling humans to directly link motivational value to abstract collective causes, principles, and ideologies (11). The observation that anterior prefrontal sectors were recruited by costly decisions indicates that when immediate self-interest and moral beliefs are at odds, altruistic decisions entail more complex event–outcome associations (11, 13, 34). This finding is supported by the role of this region in altruistic punishment, moral judgment, assessment of abstract future rewards, and long-term goals (5, 11, 13, 25, 36).

Taken together, these lines of evidence indicate that human altruism draws on general mammalian neural systems of reward, social attachment, and aversion. In the context of intertwined social and motivational contingencies, however, altruism tied to abstract moral beliefs relies on the uniquely developed human anterior prefrontal cortex.

Materials and Methods

Subjects. Nineteen healthy participants (10 men, 28.2 ± 6.2 years of age, education 17.4 ± 2.3 years; mean \pm SD) took part in the fMRI study. Before the fMRI experiment, a behavioral study involving 58 healthy volunteers (29 men, 33.3 ± 8.0 years of age, education 16.7 ± 2.2 years; mean \pm SD) was carried out to design and assess the stimuli and task procedures and to guide the selection of ORGs and societal causes for the fMRI experiment (see *Supporting Methods*). All participants were right-handed and native English speakers. Informed consent was obtained according to procedures

approved by the National Institute of Neurological Disorders and Stroke (NINDS) Internal Review Board. All participants were paid according to the NINDS standards.

General Task Design. A personal endowment of U.S.\$128 was made available for each participant in the fMRI experiment, which corresponded to the maximum amount they could obtain for themselves during the experimental task. Participants were told that additional experimental funds were available for ORG reimbursements, and they understood that their decisions on each trial would ultimately affect their personal endowment and the monetary benefits allocated to ORGs, depending on the payoff type. They were encouraged to make free choices and were guaranteed anonymity. Before scanning, participants browsed the full list of ORGs and mission statements and then were given a supervised 10-min practice session with the actual task (additional ORGs were used for this purpose). During each trial of the task, the name of an ORG and a short mission statement were displayed for 6 s. This step was followed by the combined payoff [i.e., the personal (YOU) and organizational (ORG)], which could be of four types: (YOU: \$+2 ORG: \$0), (YOU: \$0 ORG: \$+5), (YOU: \$-2 ORG: \$+5), and (YOU: \$+2 ORG: \$+5) (see Figs. 1 and 5–8). Next, a decision to accept (Yes) or reject (No) the combined payoff had to be made within 3.5 s by a button click with the index or the middle finger of the right hand. The outcome (e.g., “YOU will get: \$0” “ORG will get: \$+5”) was then presented for 2.5 s, followed by a jittered interval time. Each one of the four payoff types appeared 32 times during the experiment, randomly combined with 64 different ORGs. Each ORG appeared two times throughout the experiment (combined with different payoffs), totaling 128 trials (4 payoff types \times 64 ORGs \times 2 appearances). To enforce a decision in every trial, participants were informed that U.S.\$1 would be deducted

15. Wood JN, Grafman J (2003) *Nat Rev Neurosci* 4:139–147.
16. Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004) *Neuron* 44:389–400.
17. Eisenberg N (2000) *Annu Rev Psychol* 51:665–697.
18. Schultz W, Dayan P, Montague PR (1997) *Science* 275:1593–1599.
19. O’Doherty JP, Buchanan TW, Seymour B, Dolan RJ (2006) *Neuron* 49:157–166.
20. Bartels A, Zeki S (2004) *NeuroImage* 21:1155–1166.
21. Aron A, Fisher H, Mashek DJ, Strong G, Li H, Brown LL (2005) *J Neurophysiol* 94:327–337.
22. Young LJ, Wang Z (2004) *Nat Neurosci* 7:1048–1054.
23. Blair RJ, Morris JS, Frith CD, Perrett DI, Dolan RJ (1999) *Brain* 122:883–893.
24. Eslinger PJ, Flaherty-Craig CV, Benton AL (2004) *Brain Cognit* 55:84–103.
25. Coricelli G, Critchley HD, Joffily M, O’Doherty JP, Sirigu A, Dolan RJ (2005) *Nat Neurosci* 8:1255–1262.
26. Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002) *Neuron* 35:395–405.
27. Singer T, Kiebel SJ, Winston JS, Dolan RJ, Frith CD (2004) *Neuron* 41:653–662.
28. Moll J, de Oliveira-Souza R, Eslinger PJ, Bramati IE, Mourao-Miranda J, Andreiuolo PA, Pessoa L (2002) *J Neurosci* 22:2730–2736.
29. Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999) *Nature* 402:179–181.
30. Delgado MR, Frank RH, Phelps EA (2005) *Nat Neurosci* 8:1611–1618.
31. Freedman LJ, Insel TR, Smith Y (2000) *J Comp Neurol* 421:172–188.
32. Depue RA, Morrone-Strupinsky JV (2005) *Behav Brain Sci* 28:313–350; discussion 350–395.
33. Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) *Nature* 435:673–676.
34. Tremblay L, Schultz W (1999) *Nature* 398:704–708.
35. Boyd R, Gintis H, Bowles S, Richerson PJ (2003) *Proc Natl Acad Sci USA* 100:3531–3535.
36. Bechara A, Tranel D, Damasio H (2000) *Brain* 123:2189–2202.