The Right and the Good: Distributive Justice and Neural Encoding of Equity and Efficiency
Ming Hsu, *et al.*
Science **320**, 1092 (2008);
DOI: 10.1126/science.1153651

24. E. S. Russell, E. C. McFarland, E. L. Kent, *Transplant. Proc.* **2**, 144 (1970).
25. L. A. McNeill *et al.*, *Mol. Biosyst.* **1**, 321 (2005).
26. C. Peyssonnaux *et al.*, *J. Clin. Invest.* **117**, 1926 (2007).
27. D. Yoon *et al.*, *J. Biol. Chem.* **281**, 25703 (2006).
28. E. Beutler, *Science* **306**, 2051 (2004).
29. T. Tanno *et al.*, *Nat. Med.* **13**, 1096 (2007).

30. This work was supported by NIH (grant numbers AI054523 and DK53505-09) and by the Stein Endowment Fund.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1157121/DC1
Materials and Methods
SOM Text
Figs. S1 to S9
References

28 February 2008; accepted 17 April 2008
Published online 1 May 2008;
10.1126/science.1157121
Include this information when citing this paper.

The Right and the Good: Distributive Justice and Neural Encoding of Equity and Efficiency

Ming Hsu,[1]* Cédric Anen,[2]* Steven R. Quartz[2]†

Distributive justice concerns how individuals and societies distribute benefits and burdens in a just or moral manner. We combined distribution choices with functional magnetic resonance imaging to investigate the central problem of distributive justice: the trade-off between equity and efficiency. We found that the putamen responds to efficiency, whereas the insula encodes inequity, and the caudate/septal subgenual region encodes a unified measure of efficiency and inequity (utility). Notably, individual differences in inequity aversion correlate with activity in inequity and utility regions. Against utilitarianism, our results support the deontological intuition that a sense of fairness is fundamental to distributive justice but, as suggested by moral sentimentalists, is rooted in emotional processing. More generally, emotional responses related to norm violations may underlie individual differences in equity considerations and adherence to ethical rules.

Imagine driving a truck with 100 kg of food to a famine-stricken region. The time it would take you to deliver food to everyone would cause 20 kg of food to spoil. If you delivered food to only half the population, you would lose only 5 kg. Do you deliver the food to only half the population to maximize the total amount of food, or do you sacrifice 15 kg to help everyone and achieve a more equitable distribution?

This dilemma illustrates the core issues of distributive justice, which involves trade-offs between considerations that are at once compelling but that cannot be simultaneously satisfied. More generally, distributive justice concerns how individuals and societies allocate benefits and burdens in a just or moral manner, and it is central to social choice theory, moral psychology, and welfare economics (*1*–*3*). Despite the long history of work on distributive justice, however, its psychological and neural underpinnings remain poorly understood, much of it centering on two long-standing debates.

The first debate concerns the role of equity and fairness: Is it more just to maximize some overall good (such as well-being) independently of its distribution, or must its distribution satisfy certain criteria (such as equity), even if it results in less overall well-being? Utilitarian theories of justice, exemplified by Mill and Harsanyi, maximize the good, or efficiency. In its simplest form, this involves maximizing the sum of individual utilities, irrespective of equity (*4*). In contrast, deontological theories of distributive justice maintain that the right (e.g., equity) is prior to the good and that an action can maximize the good and yet be morally wrong if it violates a deontological principle, such as a rule, right, or duty. Contemporary proponents of deontological theories, most notably Rawls, observe the near universality of fairness norms and argue that this sense of fairness underlies institutions and society as a whole, thereby generating the notion of "justice as fairness" (*5*).

A second debate concerns the involvement of emotion in distributive justice. A prominent cognitivist tradition, including such philosophers as Plato and Kant, emphasizes the role of reason in resolving the trade-off between the right and the good, as do many contemporary thinkers including Rawls and Harsanyi (*4*, *5*). In psychology, a prominent cognitivist view suggests that a sense of justice emerged as a developmental consequence of formal and abstract cognition (*2*). An alternative tradition, including moral sentimentalists such as David Hume and Adam Smith, argues that distributive justice is rooted in emotions, such as sympathy and empathy (*6*, *7*).

Although these debates remain unresolved, recent works in related fields—including moral judgment and economics—provide converging evidence of the interplay between emotion and cognition (*8*, *9*), as well as the importance of fairness (*10*–*12*), in individual and social decision-making. Based on these findings, we hypothesized that distinct neural substrates may underlie the representation of equity and efficiency. First, we hypothesized that reward regions, such as the striatum, would be involved in encoding utility and efficiency. A wide variety of decision-making studies indicate the involvement of dopaminergic regions, such as the striatum, in reward computation and reward learning (*13*, *14*), including indirect rewards such as charitable giving and punishment of free-riders in public-goods games (*15*, *16*). More recent evidence has implicated nearby paralimbic regions, especially the septal-subgenual area, in altruism and social attachment (*17*, *18*).

Second, we hypothesized that emotional systems, particularly the insular cortex, would be involved in the encoding of inequity, as recent work has demonstrated the important role of the insular cortex in fairness and empathy (*9*, *19*, *20*). We also note the deep connection that exists in economic theory between decision-making under uncertainty and the measurement of inequity (*21*–*23*). This connection is of particular relevance in light of growing evidence that the insular cortex is involved in risky decisions and risk perception (*24*, *25*), as well as a possible separation in the encoding of reward and risk (*25*). Finally, we speculated that differential activation of reward and emotional regions may reflect their trade-off between efficiency and equity and that such differential activation may correlate with individual differences in these decisions.

To investigate the neural foundations of distributive justice, we employed a novel distribution task in conjunction with functional magnetic resonance imaging (fMRI). During functional brain imaging, 26 adult participants (nine males, mean age of 39.2 years, age range of 29 to 55 years) made decisions about how to allocate money to a group of children living in an orphanage in northern Uganda (*23*). Each group of three children was endowed by the experimenters with $5, the monetary equivalent of 24 meals per child. We denominated allocations in meals to give participants an approximation of the purchasing power of the monies being donated (*23*). In each trial, participants decided whether varying allocations of money, denominated in meals, would be taken away from either of two groups of children; the participant's choice was to decide from whom to take. Participants donated $87 on average (for a total of $2279) to the charity (*23*).

This design allowed us to parametrically vary the relative efficiency and equity of the allocations, providing a quantitative framework to evaluate participants' choices (fig. S3). Specifically, we used an inequity-aversion model in which individuals trade off between equity and efficiency (*26*). The additive nature of the model, together

[1]Beckman Institute for Advanced Science and Technology and Department of Economics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [2]Social Cognitive Neuroscience Laboratory, Division of Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena, CA 91125, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: steve@hss.caltech.edu

with the design of the experiment in which equity and efficiency were varied independently, allowed us to create orthogonal regressors and explore the possibility of separate neural encoding of these two variables (*23*). The dynamic nature of our task partitioned the temporal ordering of the trials (Fig. 1 and movies S1 and S2), allowing for an event-related fMRI analysis and a search for possible differential neural contributions at various decision-making stages, as time is a crucial variable in brain systems involved in reward and learning (*13*) and models of decision-making (*27*).

Behaviorally, the group inequity-aversion parameter estimate was $\alpha = 6.96 \pm 1.08$ (*23*). In addition, individual inequity-aversion estimates showed substantial variation (fig. S4 and table S3) (*23*), which further allowed us to use the estimated individual inequity-aversion attitude as a

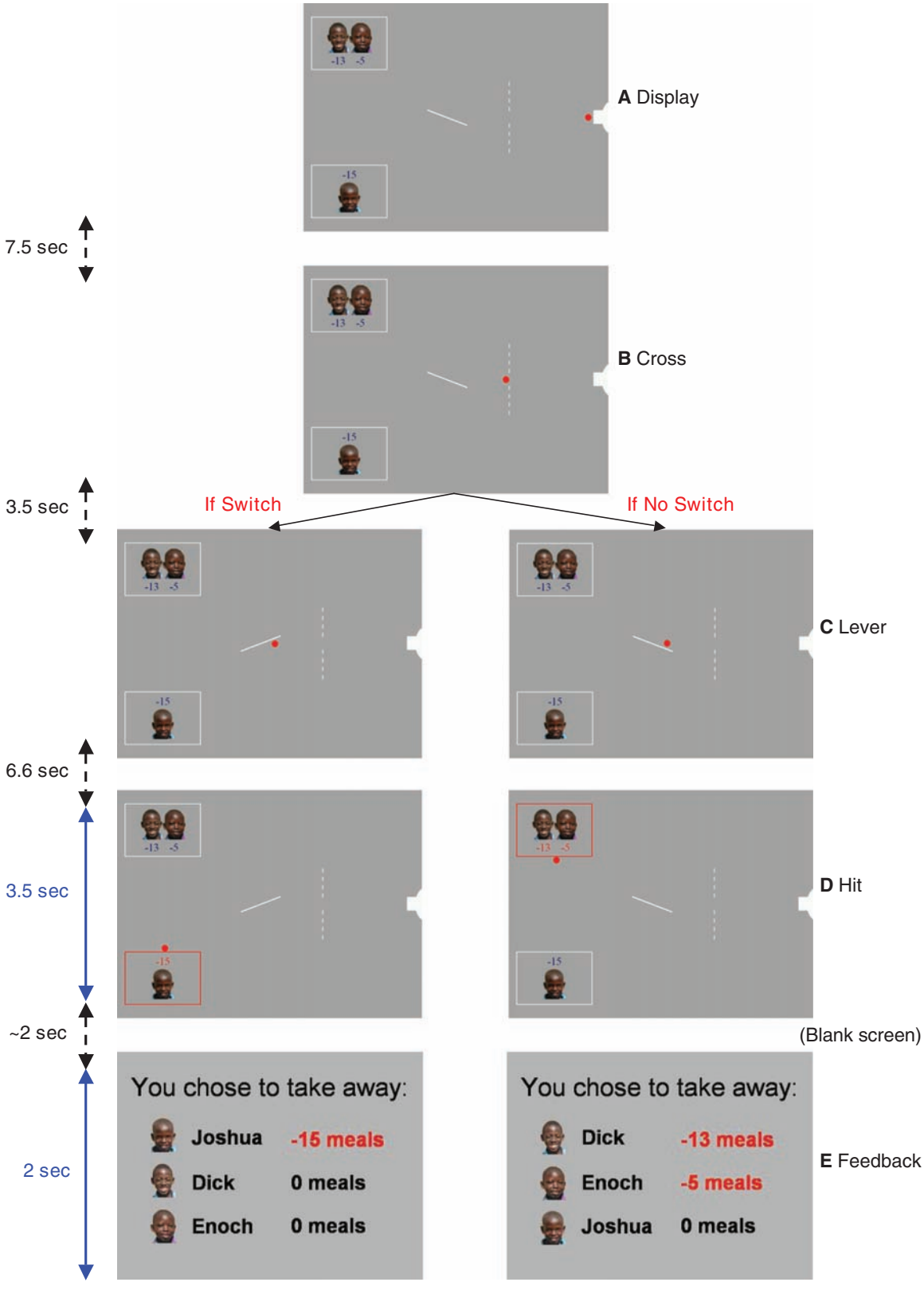

Fig. 1. Timeline and animation stills of experimental design. (**A**) Display: A projectile begins moving across the screen toward two groups of children. The number of meals each child can potentially lose (from an initial endowment of 24) is given next to the picture of the child. The position of the lever in the middle of the screen denotes the default group of children that will lose the meals. (**B**) Cross: After the projectile crosses the dotted line, the participant may switch the lever (Switch) to direct the projectile toward the other group of children. The participant may only switch the lever once and has 3.5 s to do so. (**C**) Lever: Once the projectile hits the lever, the participant can no longer switch it. The projectile continues to move toward the group of children from whom the participant has chosen/allowed the meals to be taken away. (**D**) Hit: The projectile touches the box surrounding the pictures; the box changes color and remains highlighted for 3.5 s. (**E**) Feedback: After a blank screen of random duration (uniformly distributed from 1 to 3 s), a feedback screen (2 s) informs the participant how many meals each child received. Subsequent trials are separated by a blank screen of random duration (uniformly distributed from 5 to 7 s).

between-participant measure in the neuroimaging data analysis.

Functional imaging results were analyzed using standard regression techniques (*23*). An event-related design was used where regressors were included for the various events of the trials (Fig. 1). Interaction terms corresponding to efficiency and equity, or utility, were added as parametric regressors (*23*).

We first searched for regions that respond to both efficiency and inequity in the form of the hypothesized utility function (*23*). Figure 2A shows activation of a region overlapping with the caudate head and the septal-subgenual area, with respect to the marginal utility of participants' choices (ΔU) during the "Hit" event only (*23*). The activation was driven by both marginal efficiency (ΔM) and marginal inequity (ΔG) and is the only region that survives at the uncorrected threshold of $P < 0.001$ (Fig. 2A). Furthermore, because ΔU was calculated with the group-level inequity-aversion parameter α, the inequity-aversion model predicts that the coefficients would be negatively correlated with the individual inequity-aversion estimates. That is, individuals with higher neural responses to inequity would reject the inequitable allocation in favor of one that is more equitable. Figure 2B shows that this is indeed the case (Spearman $\rho = -0.419$, $P < 0.02$, two-tailed) (*28*).

Next, we looked for the hypothesized separation in neural regions encoding efficiency and equity. Figure 3A shows bilateral putamen activation with respect to M_C, the efficiency of chosen allocations, during the "Display" event only and again is the only region that survives at the uncorrected threshold of $P < 0.001$ (*23*). Unlike the ΔU region, however, the putamen is correlated only with efficiency rather than inequity (Fig. 3B) and is also not correlated with individual inequity-aversion parameters (table S3).

In contrast, we found that activity in the bilateral insular cortex is significantly correlated with ΔG during the Display (and also the "Switch") event (Fig. 4A and fig. S4). The amount of inequity reduced by the participant's choice is therefore a monotonic function of insular activation. Furthermore, activity in the insula is not significantly correlated with measures of efficiency (Fig. 4B). Strikingly, we found that the individual β values of ΔG are significantly negatively correlated with individual inequity-aversion parameters (Fig. 4C). Therefore, as in the ultimatum game (UG) (*9*), high insula activity is associated with passing over the inequitable allocation and choosing the equitable allocation [the allocation of (0, 0) in the UG] (fig. S7 and table S4). This also supports the more general proposed role of insula in norm violation (*29*) and the idea that individual differences arise from differing sensitivity to inequity norms. That is, participants who receive strong negative affective signals may be more sensitive to violating fairness norms and hence adhere more to deontological norm following, whereas those who do not are influenced primarily by efficiency. As persistent violation of social norms is symptomatic of conduct disorder, our finding of an affective basis to norm following may also have important implications for understanding the disruption in norm following in clinical populations (*30*).

Investigations of distributive justice reach back to the beginning of philosophy and have important social, political, and economic implications (*23*). Our results show how the brain encodes two considerations central to the distributive justice calculus and shed light on the cognitivist/sentimentalist debate regarding the psychological underpinnings of distributive justice. Specifically, the dissociation between the inequity regions on the one hand and the efficiency regions on the other supports the inequity-aversion model we employ, in that individual differences in choice behavior arise from

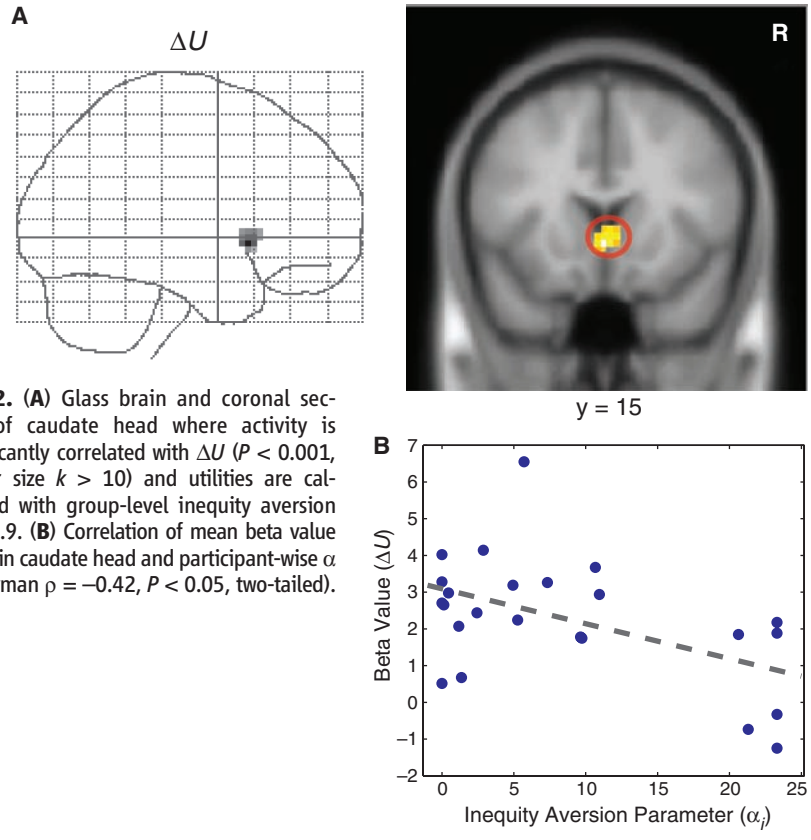

Fig. 2. (**A**) Glass brain and coronal section of caudate head where activity is significantly correlated with ΔU ($P < 0.001$, cluster size $k > 10$) and utilities are calculated with group-level inequity aversion $\alpha = 6.9$. (**B**) Correlation of mean beta value of ΔU in caudate head and participant-wise α (Spearman $\rho = -0.42$, $P < 0.05$, two-tailed).

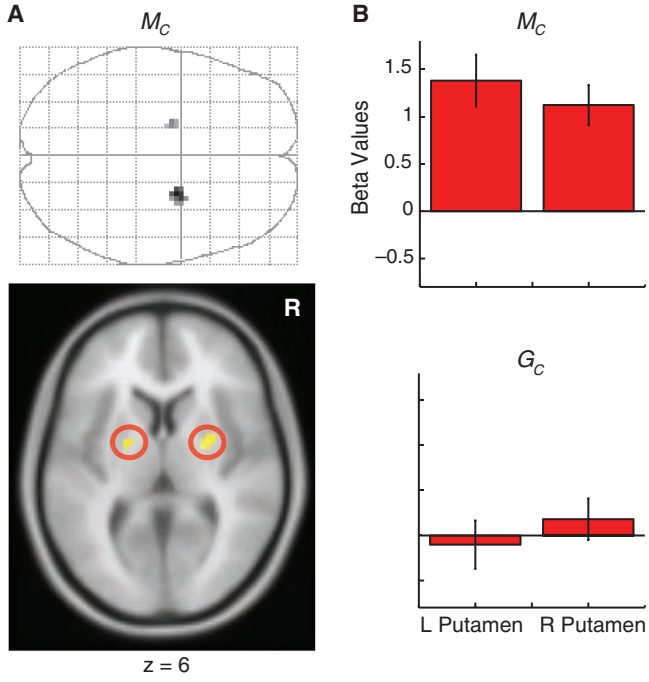

Fig. 3. (**A**) Glass brain and axial section of bilateral putamen where activity is positively correlated with M_C ($P < 0.001$, cluster size $k > 10$). (**B**) Dissociation between M_C and G_C in left and right putamen. G_C, chosen gini. Error bars indicate SEM.

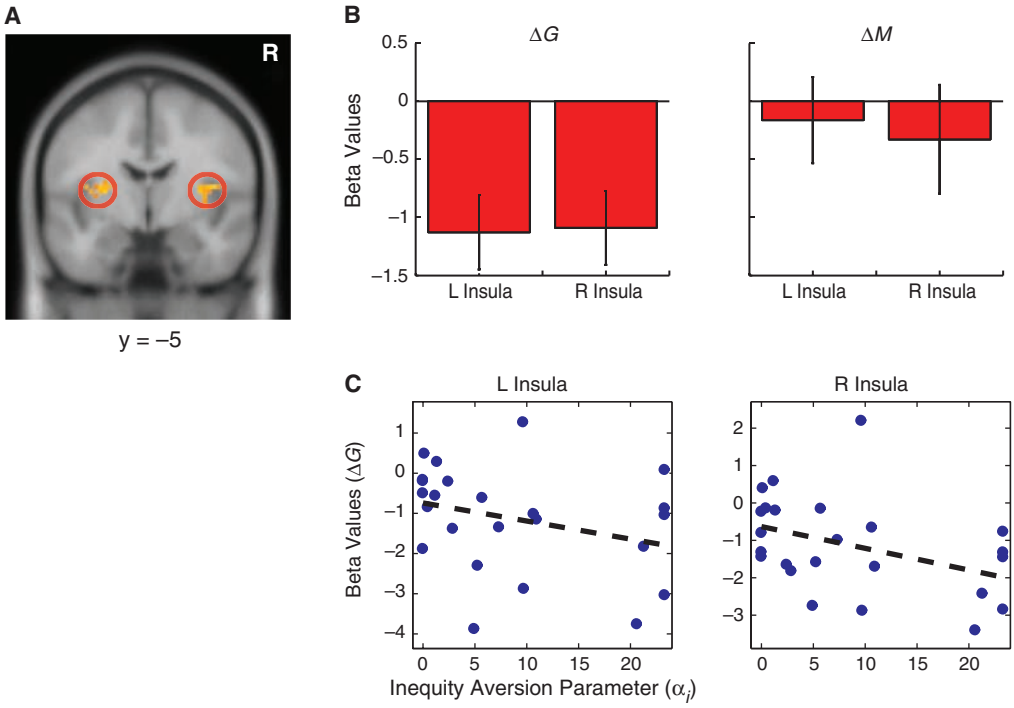

Fig. 4. (**A**) Activation in the bilateral insula during the Display event is negatively correlated with inequity measure ΔG ($P < 0.002$, cluster size $k > 10$). (**B**) Dissociation between ΔG and ΔM in the insula. Error bars indicate SEM. (**C**) Correlation of mean beta value in the insula and participant-wise α (right insula: Spearman $\rho = -0.41$, $P < 0.05$, two-tailed; left insula: $\rho = -0.36$, $P < 0.1$, two-tailed).

participants placing different weights upon inequity, as opposed to efficiency. In addition, we note the substantial degree to which activations reported above are neuroanatomically and temporally distributed across the relevant events of interest (tables S5 to S14). That is, M_C is activated in the putamen only during Display, ΔG is activated in the insula during Display and Switch, and ΔU is activated in the caudate/septal region during Hit. The anatomical separation implies that computation of the hypothesized inequity-aversion utility is distributed in the brain in much the same way that expected reward and risk are proposed to be distributed (*25*). In addition, the separation suggests a role for the region around the caudate head/septal-subgenual system in integrating multiple values into an social evaluatory signal, which is consistent with previous studies implicating this region in social attachment, trust, and charitable givings (*16*–*18*). The temporal separation supports the long-standing distinction between decision utility and experienced utility specifically, and multiple representations of utility in general (*27*). More broadly, our results support the Kantian and Rawlsian intuition that justice is rooted in a sense of fairness; yet contrary to Kant and Rawls, such a sense is not the product of applying a rational deontological principle but rather results from emotional processing, providing suggestive evidence for moral sentimentalism.

References and Notes

1. Aristotle, W. D. Ross, J. O. Urmson, *The Nicomachean Ethics* (Oxford Univ. Press, Oxford, 1980).
2. L. Kohlberg, *The Philosophy of Moral Development: Moral Stages and the Idea of Justice* (Harper and Row, San Francisco, CA, ed. 1, 1981).
3. A. K. Sen, *Collective Choice and Social Welfare* (Elsevier, New York, 1984).
4. J. C. Harsanyi, *Essays on Ethics, Social Behavior, and Scientific Explanation* (D. Reidel, Dordrecht, Holland, 1976).
5. J. Rawls, *A Theory of Justice* (Clarendon, Oxford, 1972).
6. A. Smith, *The Theory of Moral Sentiments* (Longman, Hurst, Rees, Orme, and Brown, London, ed. 11, 1812).
7. D. Hume, *A Treatise of Human Nature* (Clarendon, Oxford, ed. 2, 1978).
8. J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, J. D. Cohen, *Science* **293**, 2105 (2001).
9. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, *Science* **300**, 1755 (2003).
10. D. Engelmann, M. Strobel, *Am. Econ. Rev.* **94**, 857 (2004).
11. M. R. Delgado, R. H. Frank, E. A. Phelps, *Nat. Neurosci.* **8**, 1611 (2005).
12. E. Fehr, M. Naef, K. M. Schmidt, *Am. Econ. Rev.* **96**, 1912 (2006).
13. W. Schultz, *Nat. Rev. Neurosci.* **1**, 199 (2000).
14. J. P. O'Doherty, *Curr. Opin. Neurobiol.* **14**, 769 (2004).
15. D. J.-F. de Quervain *et al.*, *Science* **305**, 1254 (2004).
16. W. T. Harbaugh, U. Mayr, D. R. Burghart, *Science* **316**, 1622 (2007).
17. F. Krueger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20084 (2007).
18. J. Moll *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15623 (2006).
19. T. Singer *et al.*, *Nature* **439**, 466 (2006).
20. T. Singer *et al.*, *Science* **303**, 1157 (2004).
21. A. B. Atkinson, *J. Econ. Theory* **2**, 244 (1970).
22. Y. Amiel, F. A. Cowell, *Thinking About Inequality: Personal Judgment and Income Distributions* (Cambridge Univ. Press, Cambridge, 1999).
23. See supporting data on *Science* Online.
24. C. M. Kuhnen, B. Knutson, *Neuron* **47**, 763 (2005).
25. K. Preuschoff, P. Bossaerts, S. R. Quartz, *Neuron* **51**, 381 (2006).
26. We assume the utility function for participant i to be $u_i(\boldsymbol{x}) = \sum_{x \in \boldsymbol{x}} x - \alpha_i \cdot \mathrm{gini}(\boldsymbol{x})$, where $\boldsymbol{x}$ is the vector of allocations. Efficiency is the total number of meals donated, and inequity is measured by the "gini" coefficient. The parameter α captures the weighting placed upon inequity. We denote the chosen total meals as M_C, the unchosen meals M_U, and the marginal number of meals $\Delta M = M_C - M_U$. Terms for gini (G) and utility (U) are defined similarly (*22*).
27. D. Kahneman, P. P. Wakker, R. Sarin, *Q. J. Econ.* **112**, 375 (1997).
28. Although we controlled for gender and age effects in the regression model, we cannot exclude the hypothesis that this relation is driven in part by gender effects (*22*). Prominent gender difference in fairness behavior has been reported, among others, in (*12*).
29. P. R. Montague, T. Lohrenz, *Neuron* **56**, 14 (2007).
30. R. Loeber, J. D. Burke, B. B. Lahey, A. Winters, M. Zera, *J. Am. Acad. Child Adolesc. Psychiatry* **39**, 1468 (2000).

31. This project was supported by the David and Lucile Packard Foundation, the Gordon and Betty Moore Foundation, the John Templeton Foundation (S.R.Q.), and the Beckman Institute (M.H.). We thank the Canaan Children's Home for their cooperation; M. Stewart for research assistance; and T. Anastasio, K. Binmore, P. Bossaerts, C. Camerer, P. Glimcher, S. Mysore, R. Montague, and S. Williams for valuable suggestions and discussion of ideas.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1153651/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 to S14
References
Movies S1 and S2

3 December 2007; accepted 21 April 2008
Published online 8 May 2008;
10.1126/science.1153651
Include this information when citing this paper.