Science AAAS

Using Neural Measures of Economic Value to Solve the Public Goods Free-Rider Problem

Ian Krajbich, *et al.*

Science **326**, 596 (2009);
DOI: 10.1126/science.1177302

fasciculus gracilis 14 days later. In the $PTP\sigma^{-/-}$ mice, axon extension into the lesion penumbra was significantly improved ($P < 0.002$; Fig. 4, D to F, and fig. S4). $PTP\sigma^{-/-}$ axons extended well into the region of inhibitory proteoglycan surrounding the lesion (Fig. 4, G to I). However, similar to the effect of chondroitinase (*10*), robust regeneration beyond the core of the lesion did not occur. This might, in principle, reflect partial redundancy with other PTPs in the LAR subfamily, and it would also be consistent with the known presence of other growth impediments such as the myelin inhibitors (*1*–*5*) and factors intrinsic to unconditioned neurons (*33*, *34*). The results in this regeneration model system demonstrate a role for PTPσ in mediating the axonal response to the inhibitory CSPG-rich scar in a spinal cord lesion in vivo.

It has long been recognized that CSPG is one of the major inhibitors of neural regeneration; however, the mechanism has been poorly understood, and it has been unclear whether the mechanism even involves specific cellular receptors, limiting the options to tackle this important area by molecular approaches. Finding that PTPσ is a functional receptor that binds and mediates actions of CSPGs opens the door to new molecular approaches to understand CSPG action not only in regeneration, but also in development and plasticity. Our work on PTPσ also sheds new light on functions of the PTP family, and it will be interesting to know whether the other PTPs of the LAR subfamily may collaborate in nerve regeneration. The finding that a PTPσ fusion protein can detect lesion sites in the adult CNS not only sheds light on the biological role of PTPσ but also provides an injury biomarker, and thus a potential tool for research or diagnosis. Furthermore, the identification of a specific site on PTPσ that binds CSPG provides a lead for potential drug design to treat spinal cord injury. Alternative blocking approaches, such as soluble receptor ectodomains, could be used, and such approaches could potentially be combined with the blockade of other regeneration inhibitors. In addition to the possible treatment of spinal cord injury, the results here may be relevant to many other forms of neural injury as well as neurodegeneration that involves reactive astrogliosis. Identifying a functional receptor for a major class of regeneration inhibitors provides new pathways for research into mechanisms and therapeutic interventions to enhance regeneration or plasticity after nervous system injury.

Note added in proof: While this paper was in press, an additional characterization of the *PTPσ* gene knockout was published. Fry *et al.* (*35*) studied the corticospinal tract and reported regeneration after both surgical and contusive lesions, further contributing to the evidence in the present paper and previous studies that PTPσ acts in multiple areas of the nervous system and can play a key role in regeneration.

References and Notes

1. L. C. Case, M. Tessier-Lavigne, *Curr. Biol.* **15**, R749 (2005).
2. M. Domeniconi, M. T. Filbin, *J. Neurol. Sci.* **233**, 43 (2005).
3. B. P. Liu, W. B. Cafferty, S. O. Budel, S. M. Strittmatter, *Philos. Trans. R. Soc. London Ser. B* **361**, 1593 (2006).
4. P. Lu, M. H. Tuszynski, *Exp. Neurol.* **209**, 313 (2008).
5. G. Yiu, Z. He, *Nat. Rev. Neurosci.* **7**, 617 (2006).
6. J. Silver, J. H. Miller, *Nat. Rev. Neurosci.* **5**, 146 (2004).
7. K. E. Rhodes, J. W. Fawcett, *J. Anat.* **204**, 33 (2004).
8. V. J. Tom, M. P. Steinmetz, J. H. Miller, C. M. Doller, J. Silver, *J. Neurosci.* **24**, 6531 (2004).
9. J. M. Massey *et al.*, *J. Neurosci.* **26**, 4406 (2006).
10. E. J. Bradbury *et al.*, *Nature* **416**, 636 (2002).
11. W. B. Cafferty *et al.*, *J. Neurosci.* **28**, 11998 (2008).
12. J. D. Houle *et al.*, *J. Neurosci.* **26**, 7405 (2006).
13. T. Pizzorusso *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8517 (2006).
14. N. J. Tester, D. R. Howland, *Exp. Neurol.* **209**, 483 (2008).
15. D. M. Snow, V. Lemmon, D. A. Carrino, A. I. Caplan, J. Silver, *Exp. Neurol.* **109**, 111 (1990).
16. R. J. Gilbert *et al.*, *Mol. Cell. Neurosci.* **29**, 545 (2005).
17. R. J. McKeon, A. Hoke, J. Silver, *Exp. Neurol.* **136**, 32 (1995).
18. M. J. Chagnon, N. Uetani, M. L. Tremblay, *Biochem. Cell Biol.* **82**, 664 (2004).
19. K. G. Johnson, D. Van Vactor, *Physiol. Rev.* **83**, 1 (2003).
20. A. R. Aricescu, I. W. McKinnell, W. Halfter, A. W. Stoker, *Mol. Cell. Biol.* **22**, 1881 (2002).
21. A. N. Fox, K. Zinn, *Curr. Biol.* **15**, 1701 (2005).
22. K. G. Johnson *et al.*, *Neuron* **49**, 517 (2006).
23. J. McLean, J. Batt, L. C. Doering, D. Rotin, J. R. Bain, *J. Neurosci.* **22**, 5481 (2002).
24. P. S. Sapieha *et al.*, *Mol. Cell. Neurosci.* **28**, 625 (2005).
25. K. M. Thompson *et al.*, *Mol. Cell. Neurosci.* **23**, 681 (2003).
26. R. A. Asher *et al.*, *J. Neurosci.* **20**, 2427 (2000).
27. C. C. Chan, A. K. Wong, J. Liu, J. D. Steeves, W. Tetzlaff, *Glia* **55**, 369 (2007).
28. K. Haworth, K. K. Shu, A. Stokes, R. Morris, A. Stoker, *Mol. Cell. Neurosci.* **12**, 93 (1998).
29. M. Elchebly *et al.*, *Nat. Genet.* **21**, 330 (1999).
30. Y. Gao *et al.*, *Neuron* **44**, 609 (2004).
31. J. G. Flanagan *et al.*, *Methods Enzymol.* **327**, 19 (2000).
32. S. J. A. Davies *et al.*, *Nature* **390**, 680 (1997).
33. J. L. Goldberg, M. P. Klassen, Y. Hua, B. A. Barres, *Science* **296**, 1860 (2002).
34. S. Neumann, C. J. Woolf, *Neuron* **23**, 83 (1999).
35. E. J. Fry *et al.*, *Glia* 10.1002/glia.20934 (2009).
36. We thank M. Tremblay for the PTPσ mouse strain; D. Van Vactor, M. Osterfield, Y. Chen, P. Hess, P. Brittis, J. Tcherkezian, N. Preitner, and D. Nowakowski for discussions; and the Harvard Medical School Nikon Imaging Center staff for expert assistance with imaging.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5952/592/DC1
Materials and Methods
Figs. S1 to S4
References

26 June 2009; accepted 31 August 2009
10.1126/science.1178310

Using Neural Measures of Economic Value to Solve the Public Goods Free-Rider Problem

Ian Krajbich,[2] Colin Camerer,[1,2] John Ledyard,[2] Antonio Rangel[1,2]*

Every social group needs to decide when to provide public goods and how to allocate the costs among its members. Ideally, this decision would maximize the group's net benefits while also ensuring that every individual's benefit is greater than the cost he or she has to pay. Unfortunately, the economic theory of mechanism design has shown that this ideal solution is not feasible when the group leadership does not know the values of the individual group members for the public good. We show that this impossibility result can be overcome in laboratory settings by combining technologies for obtaining neural measures of value (functional magnetic resonance imaging–based pattern classification) with carefully designed institutions that allocate costs based on both reported and neurally measured values.

Public good allocation problems are pervasive in society. Examples in the government sector include the provision of national defense and environmental clean-ups. Examples in the private sector include hiring a security guard or improving common areas in a condominium association. These examples highlight two key features of public goods. First, since their benefits are nonexcludable, they are enjoyed by all members of the group, even those who do not help pay for them. Second, the optimal allocation of public goods depends on the group members' willingness to pay for them (*1*).

If the government (or group leadership) knew every individual's valuation for the good, the allocation problem would be straightforward: The government could compute the socially optimal level of the public good and then tax group members in proportion to the benefits that they receive in order to finance the cost of the good. In fact, in this case there are many possible fair rules for splitting the cost of the public good such that every individual's benefit from the public good is greater than his or her tax (*2*, *3*). Unfortunately, individual valuations for public goods are not directly observable by the government, which makes the allocation problem challenging. In particular, self-interested individuals have an incentive to understate those values, if they are asked directly for their valuations and know that their share of the cost will increase with their reported values. This is known as the free-rider problem, and it makes it very difficult in practice to accurately determine which public goods should be provided and how the costs should be shared. Countless experiments around the world have shown that the financial incentive to free-ride is pervasive and leads to allocations with a socially inefficient level of public good provision (*4*–*6*).

Social scientists have explored two different ways to limit the problems caused by free-riding. One approach investigates whether prosocial motives can be used to overcome the financial incentive to free-ride. For example, preplay communication and costly punishment of free-riders have been shown to ameliorate the problem in laboratory settings (*7*, *8*). Although the full capabilities of these types of institutions are not yet known, the body of evidence (*4*, *5*) suggests that prosocial motives are not always sufficient to eliminate free-riding behavior in all cultures (*9*). It is also unknown whether these motives are strong and pervasive enough to solve large-scale problems of practical interest.

The second approach has focused on designing institutions (known as "mechanisms") that make it advantageous for self-interested individuals to reveal their true values. A mechanism is a set of rules specifying the information that is collected from the group members and how that information is used to decide how much of the public good to produce and how to split the costs. The number of potential mechanisms for public good problems is very large. Fortunately, the mechanism design problem is greatly simplified by a result, known as the revelation principle (*10*–*12*), which states that for every mechanism with a desirable set of properties, there is a related mechanism that achieves the same outcomes but in which individuals are simply asked to reveal their values. This result is useful and important because it limits the search space to direct revelation mechanisms: If a desirable solution does not exist within this class, then it does not exist at all.

A large body of work in economics has sought to design revelation mechanisms satisfying four desirable properties. The first is social efficiency (SE), which requires that the optimal amount of the public good always be produced, meaning that the net benefit to the group is maximized. The second property is dominant strategy incentive compatibility (DSIC), which requires that the wealth-maximizing strategy for each member of the group is to reveal his or her true value, regardless of others' values or behavior. This property is desirable because truthful reporting is essential for determining the socially efficient level of the public good, and DSIC ensures that every subject has a financial incentive to do so regardless of his or her beliefs about the other group members. The third property is balanced budget (BB), which requires that the cost of the public good be completely covered by the members of the group. This property is desirable because it rules out the need for outside sources of funding. The fourth property is voluntary participation (VP), which requires that the expected value from participating in the mechanism be nonnegative for each individual, so that members do not have to be coerced into participating. A central result in economic theory is that that there is no set of rules satisfying all four desired criteria (SE, DSIC, BB, and VP) simultaneously (*13*). In response to this fundamental impossibility result, theorists and experimenters have explored mechanisms that relax some of the criteria, but those mechanisms constitute a less than ideal solution to the problem (*14*, *15*).

A key assumption behind the impossibility result is that the information used by the mechanisms is restricted to voluntarily reported values. However, a growing body of work in neuroscience has shown that it is possible to read subjective states with ranging degrees of accuracy (commonly 60 to 90%) using technology such as functional magnetic resonance imaging (fMRI) (*16*–*23*). This technology opens the door for a new class of mechanisms in which outcomes and payments depend both on individuals' reported values and on neural readings about their values. We refer to this new class of institutions as neurally informed mechanisms (NIMs).

To explore the technological feasibility of NIMs, we studied the public good allocation problem in a simple experimental setting. In each trial, subjects were randomly assigned to a group of size N = 5, 10, 15, 20, or 25 and were assigned either a low ($0 to $2) or high ($8 to $10) induced value for an abstract public good (*24*). The cost of this good was fixed at $5 × N. As is common in experimental economics, subjects were paid based on their payoffs in the experiment. Therefore, subjects were paid an amount equal to their value for the public good if it was produced, and zero otherwise. Subjects made decisions in 50 different trials and were paid based on their average payoff from all trials. The overall payoff for each trial depended on the subject's value, the tax he or she had to pay (described below), and whether or not the public good was produced. Under the NIM, the public good was produced only when the sum of the reported values was greater than its cost. The true values were independently and identically drawn from a uniform distribution so that on average it was efficient to produce the public good in only half of the trials.

The experimental task procedure and rules of the NIM were as follows. First, subjects were shown the parameters of the decision problem in the sequential order depicted in Fig. 1A (*24*) while undergoing whole-brain fMRI. Their trial-specific value for the public good was shown in isolation during an initial screen, which allowed us to use a nonlinear support-vector-machine classifier (SVM) to predict subjects' values (high or low) based only on their pattern of neural responses to the value screen (*24*). After seeing the group size and the total cost of the public good, subjects chose whether to report their true value for the public good (high or low). If the public good was produced, the NIM then used both the classifier predictions and the reported values to determine the taxes paid by each individual, as depicted in Fig. 1B. Subjects are penalized with a higher tax when their reported value differs from the classifier's prediction. Furthermore, the higher the prediction accuracy, the more likely it is that a lie will be detected.

In the supporting online material (*24*), we show that the NIM satisfies SE, DSIC, BB, and VP. Because the public good is produced only when the reported values exceed the cost, SE requires that every individual reveal his or her true value. Subjects' incentives to reveal their true values depend

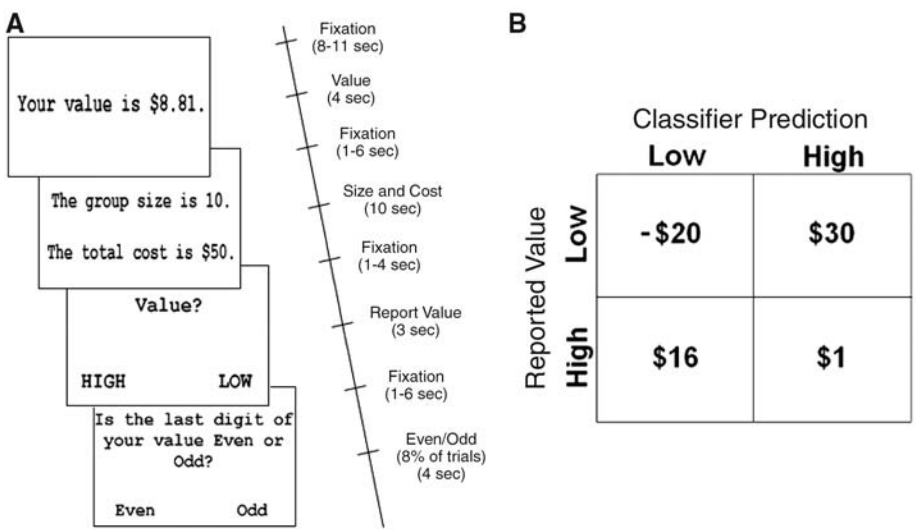


Fig. 1. (**A**) Timing of the experimental trials (top to bottom). (**B**) Tax paid by the subject in each trial as a function of the classifier's prediction and his or her reported type. Negative numbers denote transfers to the subjects.

[1]Computation and Neural Systems Program, California Institute of Technology, Pasadena CA 91125, USA. [2]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: rangel@hss.caltech.edu

on what they believe the accuracy of the classifier to be. Figure 2A depicts the difference in expected payoff between truth-telling and lying as a function of the classifier's accuracy. Low-value types are strictly better off revealing their true value for any classification rate between 50% (i.e., no decoding) and 100% (i.e., full decoding). In contrast, high-value types are strictly better off revealing their true value for any accuracy rate above 55% but have an incentive to lie for rates between 50% and 55%. This provides an intuition for why the mechanism satisfies DSIC, and thus SE, for classification rates above 55%. Figure 2B shows the total expected payoff from reporting truthfully in the NIM at different classifier accuracies, assuming that the other subjects are reporting truthfully and that they have high values 50% of the time. The expected payoff is positive for both value types at classification rates above 60%, which illustrates why VP is satisfied. Finally, BB is satisfied because by design the NIM distributes any financial surplus or deficit evenly between the players.

There was no feedback during the experiment, and subjects' values were classified afterward to determine their payments. Therefore, subjects made decisions based solely on their beliefs about the classifier's accuracy, which were assessed at the end of the experiment by debriefing. The rules of the NIM were explained to the subjects beforehand (*24*). In particular, they were told that in a previous experiment the same classification algorithm used here was able to predict values with an accuracy of 60%. Clear instructions about how the mechanism works are necessary to guard against comprehension mistakes, which would cloud interpretation of the results, and are considered a requirement by mechanism designers (*25*).

The 60 ± 2% (SEM) estimate for the classifier accuracy was based on an actual preliminary calibration experiment in which 14 subjects played a simple version of the NIM. In this experiment, the classified values played no role on outcomes and the subjects did not know that their values were being predicted (Fig. 3A) (*24*).

Figure 3 depicts the results of the experiment. The average classification accuracy was 56 ± 4% (SEM), insignificantly below the stated 60% rate (two-tailed $P = 0.33$) (Fig. 3A). We tested subjects' belief in the accuracy of the classifying technology by asking them to predict the classifier prediction rate for their own data and rewarding them based on the accuracy of their guess. During the debriefing period, subjects predicted a classification rate of 64 ± 2% (SEM), which is insignificantly different from the actual classification rate (two-tailed $P = 0.10$) (Fig. 3A). Most importantly, subjects revealed their true values nearly 100% of the time, consistent with the properties of the NIM at the subjects' predicted classification rates (Fig. 3B). Figure S15 shows that the frequency of truth-telling did not change during the experiment (*24*). Figure 3C compares the social surplus generated by the NIM, which is a measure of social efficiency, with two important benchmarks: (i) the social optimum that could be achieved if the government had full information and thus could always choose the socially efficient allocation and (ii) the theoretical average outcome generated by the best non-NIM mechanism satisfying BB, VP, and DSIC (*24*). The NIM generated 93% of the full-information so-

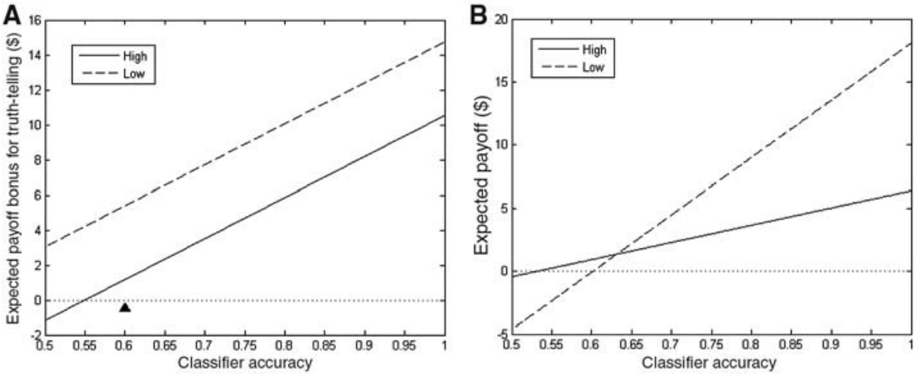
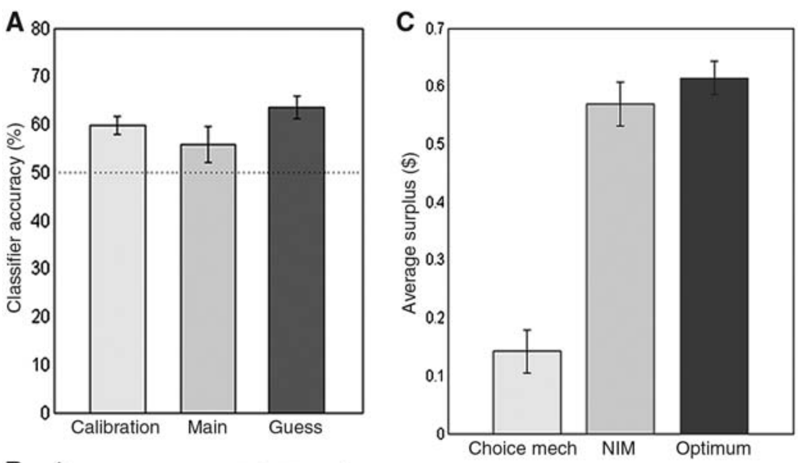


Fig. 2. (**A**) Expected benefit of truth-telling as a function of value type and classifier accuracy. For a particular classifier accuracy, the value of the curve indicates the difference in expected payoff between reporting truthfully and lying. Therefore, if the value is positive, then IC is satisfied and the subject should report his or her true value; if the value is negative, then IC is violated and the subject would earn more by misreporting his or her value. The arrow denotes the payoffs at the 60% accuracy rate used to describe the mechanism. A subject's decision is based on his or her beliefs about what the accuracy of the classifier will be and not on the realized accuracy after the experiment. (**B**) Total expected payoffs as a function of the actual classification accuracy of the mechanism for a subject who reveals his or her true type (*24*). For a particular classifier accuracy, the value of the curve indicates how much the subject can expect to earn, on average, if he or she reports his or her type truthfully. A positive value means that VP is satisfied; a negative value means VP is violated. Because the function is increasing with the accuracy rate, subjects have an incentive to cooperate with the experimenter to make the classifier as accurate as possible.



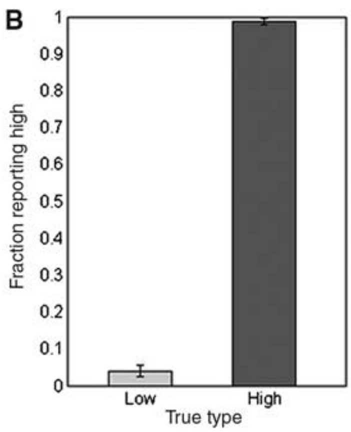


Fig. 3. (**A**) Mean accuracy rate of the classifier in the calibration ($N = 14$) and main experiments ($N = 10$), as well as the classification rate guessed by the subjects ($N = 10$), with standard error bars. Two-sided P values: calibration versus main = 0.41; calibration versus guess = 0.20; main versus guess = 0.10. (**B**) Individual reports as a function of true type ($N = 10$), with standard error bars clustered by subject. Low types misreported high 3.5 ± 1.6% (SEM) of the time, whereas high types reported high 99.5 ± 0.5% (SEM) of the time. (**C**) Average social surplus per individual ($N = 489$), a monetary measure of social efficiency, in the best non-neural mechanism [0.142 ± 0.037 (SEM)], the NIM [0.569 ± 0.038 (SEM)], and the best possible allocation under conditions of full information [0.614 ± 0.029 (SEM)]. One-sided P values: choice versus NIM = 10^{-18}; choice versus optimum = 10^{-23}; NIM versus optimum = 0.20.

cial optimum, as compared with 23% for the best theoretical non-NIM mechanism.

This study establishes the viability of NIMs in a simple experimental setting with two types and with experimentally induced valuations. Because NIMs constitute a considerable departure from previous institutions used to solve the public goods allocation problem, it is worth highlighting several of their key properties.

First, NIMs advance the theory and practice of mechanism design by combining economic theory with neural measurement technology. In the past, economists have considered mechanisms that only use the reported values from each group member to determine whether the public good is produced and how the costs are shared. Here, we show that it is possible to do substantially better by also employing fMRI measures that are reliably correlated with value.

Furthermore, the use of NIMs is not limited to fMRI technology. As shown in detail in (*24*), all that is needed for the NIM to work is the existence of some signal of value that is known to be informative, whatever its source. Thus, simple physiological measures (e.g., pupil dilation or facial electromyography) might be feasible as well.

Another attractive property of NIMs is that they do not depend on beliefs about the types or behavior of the other group members. Truth-telling and voluntary participation are both dominant strategies with these mechanisms. The only requirement is that subjects believe that their values can be predicted with sufficient accuracy by the technology. Therefore, NIMs might not be viable if subjects could interfere with the technology. Fortunately, NIMs have a built-in incentive for subjects to make the classifier predictions as accurate as possible, because subjects' expected payoffs are increasing with the prediction accuracy (Fig. 2B).

Finally, VP is an attractive feature of the NIMs because it ensures that the public good makes every individual better off, so the entire group has an incentive to support the use of the NIM. Mechanisms are deliberately required to satisfy this VP property to bolster widespread acceptance. However, VP can be harder to satisfy when individuals have substantial amounts of risk or loss aversion (*24*), although the problem is substantially reduced as the accuracy of the neural measurements improves. Thus, future technological advances should alleviate this problem.

To summarize, the free-rider problem has been a challenge for economics, public policy, and political science since the work of Adam Smith (*26*). The field of mechanism design made substantial progress during the 20th century. Unfortunately, a major contribution of the theory was to show that an ideal solution is not possible when institutions rely only on revealed values. We have shown that this problem can be overcome in simple public good settings by using fMRI to obtain informative signals of individuals' values and using those signals to induce truthful reporting. Our results take the first step in combining physiological measurements with carefully designed mechanisms to create better institutions for collective decision-making. Future theory and experiments will be needed to take this technology to more practical applications.

References and Notes

1. P. Samuelson, *Rev. Econ. Stat.* **36**, 387 (1954).
2. E. Lindahl, in *Classics in the Theory of Public Finance,* R. Musgrave, A. Peacock, Eds. (Macmillan, London, 1919), pp. 168–176.
3. H. Moulin, *Rev. Econ. Stud.* **61**, 305 (1994).
4. Y. Chen, in *The Handbook of Experimental Economics,* C. Plott, V. Smith, Eds. (Elsevier Press, Amsterdam, 2004).
5. J. Ledyard, in *Handbook of Experimental Economics,* J. Kagel, A. Roth, Eds. (Princeton Univ. Press, Princeton, NJ, 1995).
6. S. Gailmard, T. Palfrey, *J. Public Econ.* **89**, 1361 (2005).
7. E. Fehr, S. Gachter, *Nature* **415**, 137 (2002).
8. J. Falkinger, E. Fehr, S. Gachter, R. Winter-Ebmer, *Am. Econ. Rev.* **90**, 247 (2000).
9. B. Herrmann, S. Gachter, C. Thoni, *Science* **319**, 1362 (2008).
10. D. Diamantaras, E. Cardamone, K. Campbell, S. Deacle, L. Delgado, *A Toolbox for Economic Design*. (Palgrave McMillan, New York, NY, 2009).
11. A. Gibbard, *Econometrica* **41**, 587 (1973).
12. R. B. Myerson, in *The New Palgrave Dictionary of Economics Online,* S. N. Durlauf, L. E. Blume, Eds. (Palgrave Macmillan, 2008).
13. J. Green, J. Laffont, *Econometrica* **45**, 427 (1977).
14. T. Groves, J. Ledyard, *Econometrica* **45**, 783 (1977).
15. E. Maskin, *Rev. Econ. Stud.* **66**, 23 (1999).
16. J. V. Haxby *et al.*, *Science* **293**, 2425 (2001).
17. J. D. Haynes, G. Rees, *Curr. Biol.* **15**, 1301 (2005).
18. Y. Kamitani, F. Tong, *Nat. Neurosci.* **8**, 679 (2005).
19. K. N. Kay, T. Naselaris, R. J. Prenger, J. L. Gallant, *Nature* **452**, 352 (2008).
20. A. J. O'Toole *et al.*, *J. Cogn. Neurosci.* **19**, 1735 (2007).
21. L. Pessoa, S. Padmala, *Cereb. Cortex* **17**, 691 (2007).
22. S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, *Science* **310**, 1963 (2005).
23. J. T. Serences, G. M. Boynton, *Neuron* **55**, 301 (2007).
24. Materials and methods are available as supporting material on *Science* Online.
25. D. Abreu, H. Matsushima, *Econometrica* **60**, 1439 (1992).
26. A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (Regnery Publishing, Washington DC, 1776).

27. Financial support from NSF–Integrative Graduate Education and Research Traineeship Program (I.K.), Gordon and Betty Moore Foundation (C.F.C., A.R.), and Human Frontier Science Program (C.F.C.) is gratefully acknowledged.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1177302/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S3
References

4 June 2009; accepted 25 August 2009
Published online 10 September 2009;
10.1126/science.1177302
Include this information when citing this paper.