Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange
Brooks King-Casas, *et al.*
Science **308**, 78 (2005);
DOI: 10.1126/science.1108062

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

31. R. F. Anderson, Z. Chase, M. Q. Fleisher, J. P. Sachs, *Deep-Sea Res. II* **49**, 1909 (2002).
32. M. A. Brzezinski *et al.*, *Geophys. Res. Lett.* **29**, 1564 (2002).
33. D. A. Hutchins, K. W. Bruland, *Nature* **393**, 561 (1998).
34. M. Ikehara *et al.*, *Paleoceanography* **15**, 170 (2000).
35. E. Calvo, C. Pelejero, G. A. Logan, P. De Deckker, *Paleoceanography* **19**, 10.1029/2003PA000992 (2004).
36. P. J. Mueller, M. Cepek, G. Ruhland, R. R. Schneider, *Paleogeogr. Paleoclimatol. Paleoecol.* **135**, 71 (1997).
37. M.-A. Sicre *et al.*, *Org. Geochem.* **31**, 577 (2000).
38. P. Martinez *et al.*, *Org. Geochem.* **24**, 411 (1996).
39. D. Budziak *et al.*, *Paleoceanography* **15**, 307 (2000).
40. A. Sanyal, N. G. Hemming, G. N. Hanson, W. S. Broecker, *Nature* **373**, 234 (1995).
41. N. R. Catubig *et al.*, *Paleoceanography* **13**, 298 (1998).
42. J. Bijma, B. Hoenisch, R. E. Zeebe, *Geochem. Geophys. Geosys.* **3**, 10.1029/2002GC00038 (2002).
43. D. M. Anderson, D. Archer, *Nature* **416**, 70 (2002).
44. J. R. Toggweiler, *Paleoceanography* **14**, 571 (1999).
45. B. B. Stephens, R. F. Keeling, *Nature* **404**, 171 (2000).
46. J. R. Toggweiler, R. Murnane, S. Carson, A. Gnanadesikan, J. L. Sarmiento, *Global Biogeochem. Cycles* **17**, 1027 (2003).
47. D. E. Archer *et al.*, *Paleoceanography* **18**, 10.1029/2002PA000760 (2003).
48. W. Broecker *et al.*, *Global Biogeochem. Cycles* **13**, 817 (1999).
49. J. R. Toggweiler, A. Gnanadesikan, S. Carson, R. Murnane, J. L. Sarmiento, *Global Biogeochem. Cycles* **17**, 1026 (2003).
50. S. Levitus *et al.*, *Science* **292**, 267 (2001).
51. L. Bopp, C. Le Quéré, M. Heimann, A. C. Manning, P. Monfray, *Global Biogeochem. Cycles* **16**, 1022 (2002).
52. L. Xie, W. W. Hsieh, *Fish. Oceanogr.* **4**, 52 (1995).
53. M. E. Conkright *et al.*, *NOAA Atlas NESDIS 46* World Ocean Database 2001, vol. 5, *Temporal Distribution of Nutrient Profiles* (U.S. Government Printing Office, Washington, DC, 2002).
54. A. H. Orsi, T. Whitworth, W. D. Nowlin, *Deep-Sea Res. I* **42**, 641 (1995).
55. We thank A. Ridgwell, S. Kienast, I. Tegen, and two anonymous reviewers for helpful comments. We gratefully acknowledge the work of all authors cited in the Supporting Online Material for their invaluable data contributions. The ideas presented here emerged from stimulating discussions with I. C. Prentice and the participants of the Green Ocean Project: www.bgc-jena.mpg.de/bgc-synthesis/projects/green_ocean.

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5718/74/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S3

17 September 2004; accepted 31 January 2005
10.1126/science.1105375

Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange

Brooks King-Casas,[1] Damon Tomlin,[1] Cedric Anen,[3] Colin F. Camerer,[3] Steven R. Quartz,[3] P. Read Montague[1,2]*

Using a multiround version of an economic exchange (trust game), we report that reciprocity expressed by one player strongly predicts future trust expressed by their partner—a behavioral finding mirrored by neural responses in the dorsal striatum. Here, analyses within and between brains revealed two signals—one encoded by response magnitude, and the other by response timing. Response magnitude correlated with the "intention to trust" on the next play of the game, and the peak of these "intention to trust" responses shifted its time of occurrence by 14 seconds as player reputations developed. This temporal transfer resembles a similar shift of reward prediction errors common to reinforcement learning models, but in the context of a social exchange. These data extend previous model-based functional magnetic resonance imaging studies into the social domain and broaden our view of the spectrum of functions implemented by the dorsal striatum.

The expression and repayment of trust is an important social signaling mechanism that influences competitive and cooperative behavior (*1–6*). The idea of trust typically conjures images of complex human relationships, so it would seem to be a difficult part of social cognition to probe rigorously in a scientific experiment. Nevertheless, instances of trust can be stripped of complicating contextual features and encoded into economic exchange games that preserve its essential features (*7–9*). For example, in a game in which two players send money back and forth with risk, trust is operationalized as the amount of money a sender gives to a receiver without external enforcement (*9*). Such trust games now enjoy widespread use both in experimental economics (*10*) and neuroscience experiments (*11–17*).

To measure neural correlates of trust using functional magnetic resonance imaging (fMRI), we first made a simple modification to a single-exchange trust game in order to improve the ecological validity of the task (*10*). Specifically, we changed the single-round format to a multiround format in which the same two individuals (one designated the "investor," and the other the "trustee") played 10 consecutive rounds. This modification reflects the fact that significant social exchanges are rarely single-shot, and the assumption that algorithms in our brains are tuned to this fact (*1–6*). Thus, by adapting the multiround format, (i) trust becomes bidirectional, in that both the investor and trustee assume the risk that money sent might not be reciprocated by their partner; and (ii) reputation building can be probed, as players develop models of one another through iterated exchange (*10*, *11*). Participants were informed that individual rounds of the trust game would be implemented as follows: One player (investor) could invest any portion of $20 with the other player (trustee), the money appreciated (three times the investment), and the trustee then decided how much of the tripled amount to repay (Fig. 1) (*18*). Players maintained their roles throughout the entire 10-round game. Responses were encoded only in monetary units and player identities were never revealed, thus stripping away many of the confounding elements of context and communication known to influence trust (*10*). Volunteers were recruited from separate subject pools at Baylor College of Medicine in Houston, TX, and California Institute of Technology in Pasadena, CA. Volunteers were instructed identically, but separately, at each institution (instructors read a script describing the task).

We used event-related hyperscan-fMRI (h-fMRI) to monitor homologous regions of two subjects' brains simultaneously as they played the multiround trust game (*19*) (fig. S1). The motivating idea behind this approach is simple: To probe neural substrates of social interactions, we scan the brains of multiple subjects engaged in a social interaction. Social decision-making critically depends on internally represented models of social partners. In principle, such covert knowledge might be inferred from behavioral observations. However, behavioral signals are intrinsically lower dimensional than their underlying neural responses, and so behavior alone is an insufficient signal source for inferring neural representations. Put another way, an inference based only on the observable behavior of a social partner ignores many observable neural processes that give rise to that behavior. The measurement of both interacting brains directly sidesteps this problem and allows us to probe the cross correlation of internal models—replacing inference with a measurement.

Reciprocity predicts trust. Linear regression analyses of the behavior of 48 pairs of subjects identified reciprocity to be the strongest predictor of subsequent increases or decreases in trust (*20*). Reciprocity is defined as a fractional change in money sent across rounds by one

[1]Human Neuroimaging Laboratory, Department of Neuroscience, [2]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [3]Social Cognitive Neuroscience Laboratory, Division of Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: read@bcm.tmc.edu

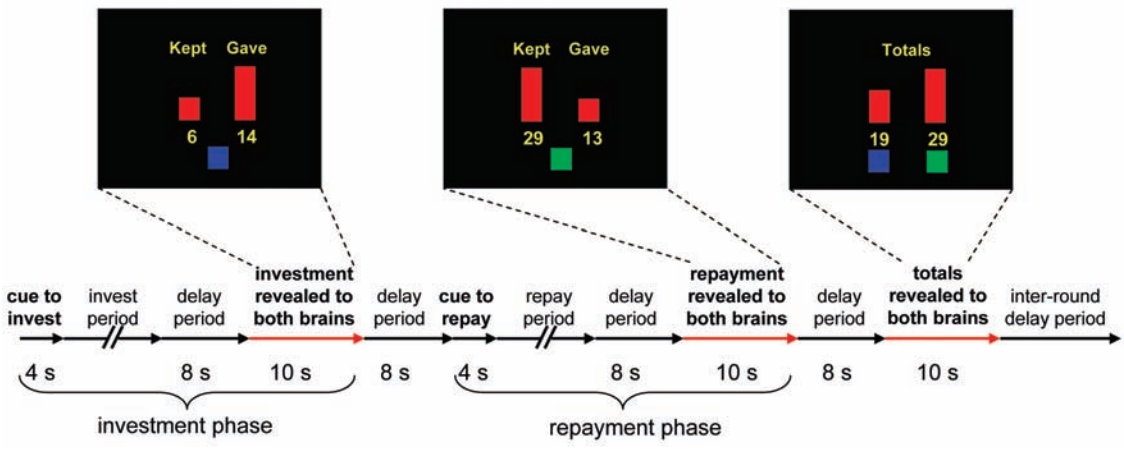

Fig. 1. Timeline for the two-person trust game. Trust experiments were carried out in 48 pairs of subjects. Each pair of subjects completed 10 consecutive trust exchanges. Each exchange began with a screen that indicated the beginning of the round, followed by a cue to invest. The investor then entrusted the trustee with any amount between 0 and 20 monetary units. During this first free response period, the trustee saw a blank screen for 8 s after the investor's decision was submitted. The investment was revealed to both players simultaneously. Amounts kept and given were represented both graphically (by a bar graph) and numerically. After the investor's decision was revealed, the trustee was then prompted to split three times the invested amount in any proportion between themselves and the investor. Eight seconds after the trustee repayment decision was submitted, the repayment was revealed to both players in the same graphical and numerical fashion. After another 8 s delay, the totals for the round were revealed using the same method. Rounds were separated by a variable 12- to 42-s interval. Except for the periods of free response, both players viewed the same visual stimuli simultaneously.

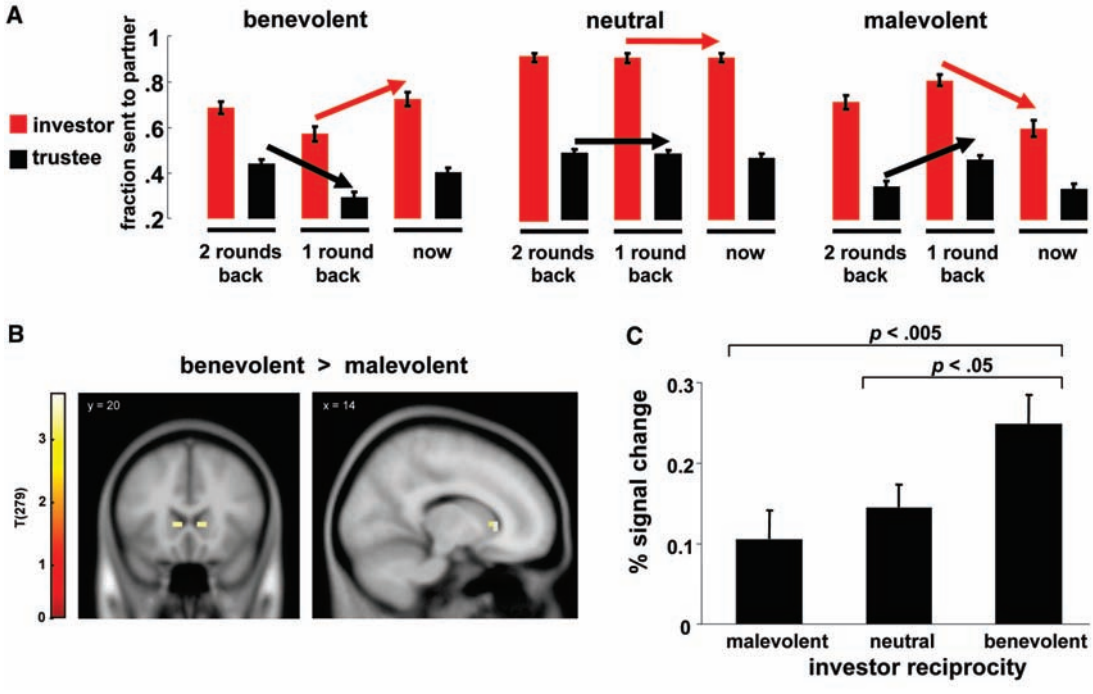

Fig. 2. Correlates of reciprocity in a multiround economic exchange. (**A**) Behavioral summary. Mean ± SE of investor (ΔI, red) and trustee (ΔR, black) behavior of rounds contributing to benevolent (n = 125), neutral (n = 134), and malevolent (n = 125) investor reciprocity categories. In each round j, investor reciprocity was defined as $r_j = \Delta I_j - \Delta R_{j-1}$; that is, the difference between the current change in payment ΔI_j by the investor in response to the previous change in repayment ΔR_{j-1} by the trustee. In the case of benevolent reciprocity, investors are being generous (sending more) in response to a defection by the trustee (decrease in repayment). Likewise, in the case of malevolent reciprocity, the investor repays the trustee's generosity (increase in previous repayment) with a breach of trust (*20*). (**B**) Response of trustee brain to investor reciprocity. A general linear model analysis identified four regions in the trustee brain that showed responses that were greater for the revelation of malevolent and benevolent investor reciprocity than for neutral reciprocity (*21*). Only one region, the head of the caudate nucleus, showed a response that was greater for benevolent relative to malevolent reciprocity (statistical parametric map shown alongside pseudo-color legend). No region showed greater responses to malevolent relative to benevolent investor reciprocity. (**C**) Region-of-interest analysis of head of caudate in trustee brain. Average activity 6 to 10 s after the investor's decision is revealed to trustee shows that the brain response to benevolent reciprocity was significantly greater from neutral (two-tailed t test, $P < 0.05$) and malevolent reciprocity (two-tailed t test, $P < 0.005$) (*21*).

player in response to a fractional change in money sent by their partner. This definition is simply an operationalized version of tit-for-tat, that is, a repayment in kind. Deviations from neutral reciprocity (perfect tit-for-tat) act as a strong social signal in the context of this game. In particular, strong deviation in investor reciprocity was the best predictor of changes in partner trust and became the primary focus of our analysis (*20*, *21*). Investor reciprocity on round j was quantified as $\Delta I_j - \Delta R_{j-1}$, where ΔI_j is the fractional change in investment from round $j - 1$ to j and ΔR_{j-1} is the last fractional change repayment ($R_{j-1} - R_{j-2}$).

Forty-eight subject pairs were scanned in this study (*21*), and we divided the exchanges into three approximately equal-sized groups: (i) benevolent reciprocity, (ii) neutral reciprocity, and (iii) malevolent reciprocity (*22*). These behavioral exchange data are summarized in Fig. 2A. For benevolent reciprocity, investors are actually being generous (sending more) in response to a defection by the trustee (decrease in repayment) (left panel). Conversely, for malevolent reciprocity, the investor repays the trustee's generosity with a breach of trust (right panel).

Using a general linear model analysis, we first sought trustee brain regions whose blood oxygenation level–dependent (BOLD) response was greater for benevolent or malevolent investor reciprocity than for neutral investor reciprocity (*21*). This analysis identified four significant regions: inferior frontal sulcus, superior frontal sulcus, thalamus, and inferior/superior colliculli (*23*). These findings

are consistent with a surprise signal—an unsigned response to deviations in the expected behavior of one's partner. A second analysis, comparing BOLD response for benevolent reciprocity to BOLD response for malevolent reciprocity, identified significant differences only in the head of the caudate nucleus (Fig. 2, B and C): (i) BOLD response was greater for instances of benevolent reciprocity relative to malevolent and neutral reciprocity; and (ii) responses to malevolent reciprocity did not differ from those to neutral reciprocity. These voxels were subsequently subjected to a region-of-interest (ROI) analysis (*21*).

"Intention to trust" signals. We expected to find a hemodynamic response in this ROI that correlated with the trustee's next choice to repay, and we expected that such signals might show strong cross-brain correlations. The reason for this expectation derived from the fact that reciprocity expressed by the investor ($\Delta I_j - \Delta R_{j-1}$) strongly predicted ($r = 0.56$) future changes in trust (repayment, ΔR_j) by the trustee. For example, benevolent reciprocity by the investor is expected to generate the intention to increase repayment (trust) in the brain of the trustee. A similar intention to decrease trust (repayment) would be expected in the trustee brain following malevolent reciprocity by the investor. Some part of the investor's brain should anticipate the neural consequences of changes in their own reciprocity on the trustee's brain; therefore, we also expected that such "intention to trust" signals would show strong cross-brain correlations. Indeed, they did.

Model building of partner: Cross-brain analysis. To carry out this analysis, we separated the hemodynamic responses in the caudate of the trustee brain into three groups according to whether their next repayment was larger, smaller, or the same as their last repayment. We were particularly interested in the net neural response to the intention to increase trust (repayment), because this act embodies risk on the part of the trustee and signals to the investor a degree of willingness to cooperate. We computed the net intent-to-trust signal in the ROI of the trustee caudate as

$$H(\text{increased repayment next round}) - H(\text{decreased repayment next round})$$

where H represents the hemodynamic response. Using this difference signal in the trustee brain, we computed cross-brain correlations with the investor brain and sought regions with the largest correlations. We were particularly interested in how the cross-brain correlations might change as the task developed and the subjects built better models of one another. Consequently, changes in this signal were examined across early (3 and 4), middle (5 and 6), and late (7 and 8) rounds using cross-brain and within-brain correlational analysis. Figure 3 illustrates the cross-correlograms of this signal with activity in two regions: the middle cingulate cortex (MCC) of investors and the anterior cingulate cortex (ACC) of trustees (*21*). The blue traces indicate that MCC activity in the investor brain and ACC activity in the trustee brain were most strongly correlated ($r > 0.59$) when the MCC signal was shifted forward in time by 14 s. The important point here is that the strongest cross-brain correlation did not shift significantly in time from early to late rounds; that is, neural responses in both brains to fiducial markers of the task did not change relative to each other. However, the peak of the cross-correlogram between investor MCC activity and the trustee "intention to trust" signal in the caudate showed a pronounced 14-s shift from early to late rounds (green traces). A similar finding resulted for the within-brain analysis of the trustee, using ACC activity and the same "intention to trust" signal in the caudate (red traces). These analyses show that a dramatic change in the relative timing of the measured BOLD signals was taking place either in the "intention to trust" signal of the trustee caudate or in both the trustee ACC and investor MCC. As shown in Fig. 4, the source of the shift is in the "intention to trust" signal of the trustee caudate.

Figure 4 shows the time traces of the hemodynamic responses in the head of the trustee caudate segregated according to future changes in trust (increases are shown in black,

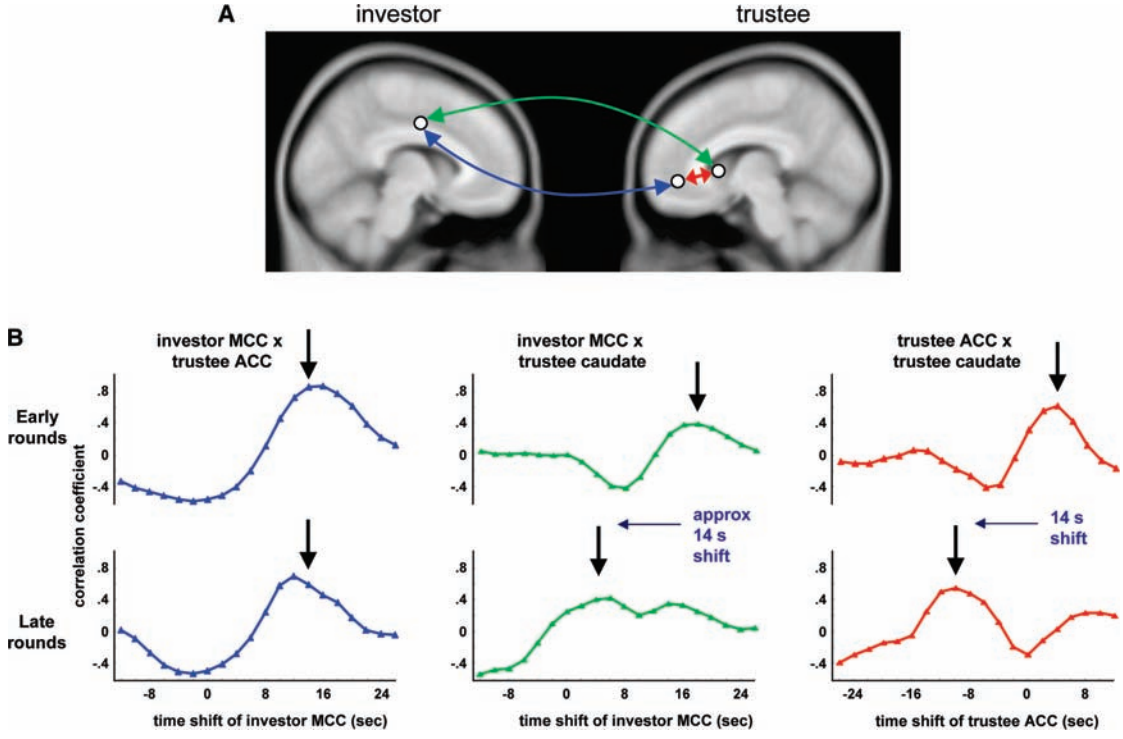

Fig. 3. Correlograms of the "intention to trust" with activity in investor MCC and trustee ACC. (**A**) Regions of correlation. The "intention to trust" signal in the trustee caudate was correlated within- and between-brains with regions that responded strongly to basic behavioral events within each round: The middle cingulate cortex (MCC) of the investor was strongly active when the investor lodged a decision, and the anterior cingulate cortex (ACC) of the trustee was strongly activated when an investor's decision was revealed (*21*). (**B**) Correlograms of caudate, ACC, and MCC. The caudate signal between rounds of increased and decreased repayment isolated an "intention to trust" signal in trustees. Average "intention to trust" signal was correlated with average ACC signal of trustee and average MCC signal of investors during the investment phase of each round (*21*) and is plotted with different time shifts. Correlograms are shown for early (rounds 3 and 4) and late (rounds 7 and 8) periods of the game. Blue traces indicate that the strongest cross-brain correlation for responses to basic behavioral events of the game did not shift significantly in time from early rounds to late rounds. The peak of the cross-correlogram between investor MCC activity and the trustee "intention to trust" signal in the caudate shows a pronounced 14-s shift from early to late rounds (green traces). A similar result is evident in the within-brain analysis of the trustee, using ACC activity and the same signal in the caudate (red traces).

decreases in red) (*21*). The amplitude and time effects associated with the 14-s time shift are shown in Fig. 4A and summarized in the bar graphs in Fig. 4B. In early rounds of the task (rounds 3 and 4), the peak of the response for intended increases in trust (i.e., an increase in next repayment) occurs after the investor's decision is revealed. In middle rounds (rounds 5 and 6), this response begins to drop back toward baseline and begins to grow at a time just before the revelation of the investor's decision. By late rounds (rounds 7 and 8), this peak is anticipatory and occurs before the revelation of the investor's decision. These data are consistent with a signal for intended increases in trust changing from being reactive to anticipatory and suggest that the trustee is building a model of the investor's likely next move. To test this model-building idea directly, we performed a separate version of the trust game and queried the trustees on each round about their expectation of the next investment.

Figure 5 illustrates the results of this additional experiment (n = 21 pairs, behavior only). On each round, both the investor and trustee were simultaneously prompted. The investor was cued to make their investment and the trustee was cued to guess the investor's decision (Fig. 5A). Timings were otherwise kept the same. The results of these experiments are summarized as the fraction of highly accurate guesses (to within ±$1) by the trustee as a function of round. Notice that the increase in the trustee's accuracy across rounds parallels the time during which the temporal transfer of the neural signal correlated with future increases in trust.

Discussion. We used an anonymous trust game in conjunction with event-related fMRI to probe neural correlates of the expression and repayment of trust between interacting human subjects. Important social relationships are rarely a single expression of trust between two strangers; thus, we made the game multiround instead of one-shot. Specifically, we sought to examine trust in a context in which (i) trust was expressed by both partners in the relationship, and (ii) trust could change over time and with experience (*25*).

Using a multiround trust game and a large sample of subjects (n = 48 pairs), we identified a social signal (reciprocity) expressed by the investor that strongly predicted changes in trust by the trustee. This social signal elicited two notable effects in the trustee brain: (i) brain regions whose activity correlated with large changes in reciprocity in a manner consistent with a surprise response; and (ii) a specific brain region, the head of the caudate nucleus, where the BOLD response was greater for benevolent reciprocity than for malevolent reciprocity. The strong relation between investor reciprocity and subsequent changes in trustee repayment led us to probe the "intention to trust" in the caudate nucleus. Rounds were segregated on the basis of whether trustees subsequently increased or decreased their repayment, representing a signal of the "intention to trust." Cross- and within-brain correlations of this intended-trust signal with neural responses to fiducial markers of the task (investment submitted and investment revealed) identified a remarkable temporal transfer of the "intention to trust" signal from a time just after the revelation of the investor's decision (a reactive signal) to a time just before this same revelation (an anticipatory signal). This shift suggested that the signal would correlate with the development of a model of the investor in the trustee's brain. To examine this latter possibility, we ran a separate behavioral experiment (n = 21 pairs) to test the trustee's ability to accurately guess (to within ±$1) the decision by the investor. The error rate of these accurate guesses dropped over the same time period during which the temporal transfer of the future trust signal shifted from reactive to anticipatory. This observation is consistent with the interpretation that the observed signals in the trustee caudate reflect the development of a reputation for their partner.

Lastly, we address an important detail about the amplitude differences between the caudate response to impending increases (black traces, Fig. 4) and impending decreases in trust (red

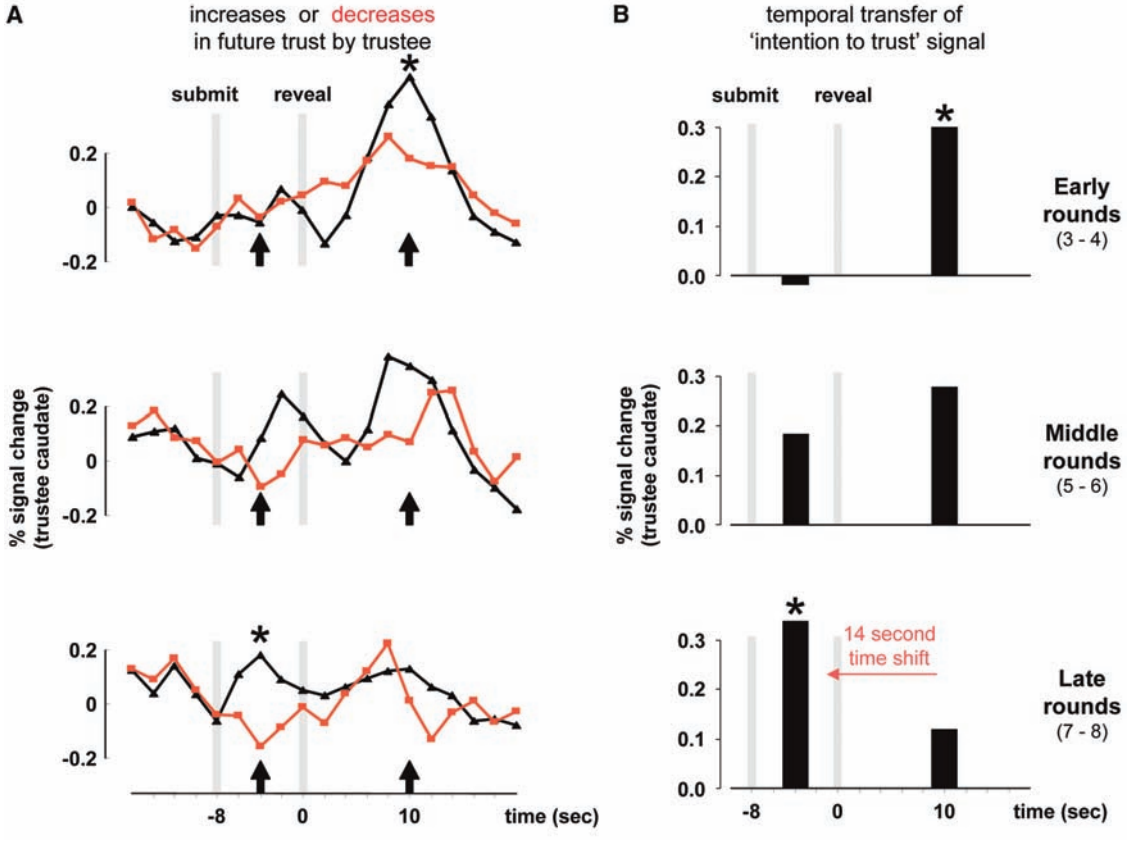

Fig. 4. Neural correlates of reputation building in trustee brain. (**A**) ROI time series. An ROI analysis was performed on voxels identified by the contrast illustrated in Fig. 2B (*21*). We segregated hemodynamic responses in response to the revelation of the investment (time = 0 s) according to the next decision made by the trustee (trustee's decision period begins at t = 22 s). Hemodynamic amplitudes for future increases in trust (ΔR > 5%; black trace) were greater (P < 0.05) than future decreases in trust (ΔR < −5%; red trace) in early rounds (top). As the game progressed (middle and bottom), the peak of this differentiated response underwent a temporal transfer from a time after the revelation of the investor's decision (t = 10 s; a reactive signal) to a time before this same revelation (t = −4 s; an anticipatory signal). Traces represent subsamples of 144 rounds in which repayment increased or decreased ≥5% (mean = 20; SD = 4.4). (**B**) ROI bar plot. The difference between the intention to increase trust [black trace of (A)] and the intention to decrease trust [red trace of (A)] is plotted for t = −4 s and t = 10 s. The 14-s temporal transfer from reactive to anticipatory is consistent with the development of a reputation for the investor within the trustee brain.

traces, Fig. 4). One explanation, supported by the behavioral data, is that increases in trust (ΔR) may have a greater effect on their partner's subsequent behavior (ΔI) than decreases in trust. If this were the case, an efficient computational system would devote more computational steps, and hence more energy, to deciding the magnitude of an increase in trust relative to a decrease. In this particular version of the trust game, increases in trust by the trustee were correlated positively with changes in investment on the subsequent round by the investor ($r = 0.27$) (fig. S6A). This was not true for decreases in trust, where there was no such correlation ($r = 0.00$) (fig. S6B). The absence of predictive information associated with a decrease in trust suggests that no analogous energetic investment should be made.

Taken together, these results suggest that the head of the caudate nucleus receives or computes information about (i) the fairness of a social partner's decision and (ii) the intention to repay that decision with trust. In early rounds of the game, the "intention to trust" is evident only after an investment is revealed. With experience, this signal shifts to a time preceding the revelation of the investment. This finding is reminiscent of analogous shifts of reward prediction error signals from reinforcement learning (*25*–*27*) that have recently been identified by fMRI in human caudate and putamen (*28*–*32*) and are thought to involve outputs of midbrain dopaminergic systems. These prediction error signals were identified using simple conditioning experiments in which lights predict the future delivery of rewards (e.g., squirt of juice or delivery of monetary return) (*33*, *34*). The scheme is simple: An initially neutral light is flashed; it causes no change in dopaminergic activity, but the later (surprising) arrival of juice causes a burst of activity in the dopamine neurons. Repeated pairing of light followed at a consistent time later by juice causes two dramatic changes: (i) The response to juice delivery drops back to baseline and (ii) a burst response occurs just after the light is flashed. This temporal transfer of the burst response to the light is thought to represent the future value predicted by the light. The simplicity of these experiments is somewhat beguiling.

The temporal transfer in the conditioning experiments is directly analogous to the temporal shift that we observe in the trustee brain as they build a model of the investor's response, but framed in the context of a social exchange. In the trustee brain, the analog to the light is the cue for the social partner to invest, and the "social juice" is change in investment. We know that positive changes in investment correlate with subsequent positive changes in repayment; a correlation that grows over the rounds of the task (fig. S5). Early in the exchange, the trustee's intention to increase trust occurs after revelation of the investor's decision to increase investment (Fig. 4A and fig. S5); that is, the increased investment is surprising. The intention to increase repayment therefore follows this revelation. As the game proceeds, this "intention to trust" response transfers to a time before the revelation of the investor decision to increase investment. The only open issue for this speculation is why the signal transferred to this particular time. There are several consistent predictors of the revelation of the investor's decision, but the signal backed up in time to occur just before this. This social prediction error interpretation is provocative and consistent but leaves this important question unanswered. The more general hypothesis is that the dopaminergic system can be used to establish more complex goal states ("rewards") and make more complex predictions through connections from prefrontal cortex onto midbrain and other subcortical structures (*35*).

It is possible that similar economic exchange tasks could be used to explore social processing deficits in a variety of neuropsychiatric disorders. These include populations that have faulty or missing capacities for building correct models of others (e.g., schizophrenia or autism spectrum disorders) (*36*, *37*), as well as individuals who misattribute motivations and intentions to others (e.g., borderline personality disorder) (*38*).

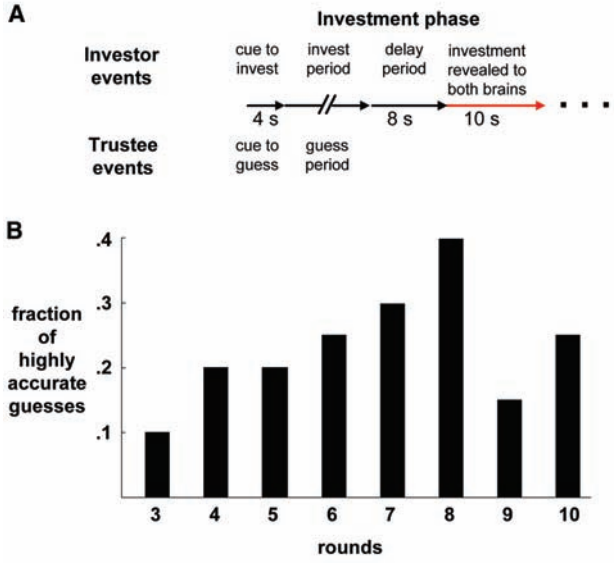

Fig. 5. Model building by trustee brain In a separate anonymous trust game (n = 21 pairs), trustees were queried to "guess the amount invested" just before the revelation of the investor's payment decision to both brains; otherwise, the task was identical to that of the original game (n = 48 pairs) from which scanning data were derived. (**A**) Timeline for queries to each player (investor and trustee). During the investment phase of the exchange, the trustees were prompted to guess the investor's decision. The trustee response to this query was not revealed to the investor. (**B**) Model building—highly accurate guesses by trustee of investor's next payment. A highly accurate guess was defined as ±1 monetary unit from the actual investment (±5%). These data show that a model of the investor's next move is available to the trustee by the middle to late rounds of the exchange and is not available in the early rounds.

References and Notes

1. R. L. Trivers, *Q. Rev. Biol.* **46**, 35 (1971).
2. R. Axelrod, W. Hamilton, *Science* **211**, 1390 (1981).
3. J. S. Coleman, in *Foundations of Social Theory* (Harvard Univ. Press, Cambridge, MA, 1990), pp. 177–179.
4. H. Rachlin, *Behav. Brain Sci.* **25**, 239 (2002).
5. R. Adolphs, *Nat. Rev. Neurosci.* **4**, 165 (2003).
6. E. Fehr, U. Fischbacher, *Nature* **425**, 785 (2003).
7. C. Camerer, K. Weigelt, *Econometrica* **56**, 1 (1988).
8. E. Fehr, G. Kirchsteiger, A. Riedl, *Q. J. Econ.* **108**, 437 (1993).
9. J. Berg, J. Dickhaut, K. McCabe, *Games Econ. Behav.* **10**, 122 (1995).
10. C. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton Univ. Press, Princeton, NJ, 2003), pp. 60–62.
11. K. McCabe, D. Houser, L. Ryan, V. Smith, T. Trouard, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11832 (2001).
12. D. J. de Quervain *et al.*, *Science* **305**, 1254 (2004).
13. J. Decetey, P. L. Jackson, J. A. Sommerville, T. Chaminade, A. N. Meltzoff, *Neuroimage* **23**, 744 (2004).
14. N. I. Eisenberger, M. D. Lieberman, K. D. Williams, *Science* **302**, 290 (2003).
15. P. Glimcher, A. Rustichini, *Science* **306**, 447 (2004).
16. J. Rilling *et al.*, *Neuron* **35**, 395 (2002).
17. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, *Science* **300**, 1755 (2003).
18. "$20" refers to 20 monetary units (MU). Subject payment at the end of the experiment varied between $20 and $40, depending on the number of MU the subject accumulated over 10 rounds. Payment schedule was as follows: <68 MU = $20, 68 to 133 MU = $25, 134 to 200 MU = $30, 201 to 300 MU = $35, and >300 MU = $40. Before the game, participants were informed that they would receive between $20 and $40, scaled by their performance. However, they had no knowledge of the step payoff function until the game was completed. Notice that the perfectly selfish Nash equilibrium strategy (in which the investor keeps all $20 each round) results in 200 MU; no subject adopted this strategy.
19. P. R. Montague *et al.*, *Neuroimage* **16**, 1159 (2002).
20. Investments (I) and repayments (R) were scaled by the amount available to be sent ($20 for I; three times the amount invested for R). See fig. S2 for a description of investments and repayments over the course of the game. Linear regressions identified significant predictors of change in trust for investors (ΔI_j) and trustees (ΔR_j). Three predictors of ΔI_j were examined: (i) previous repayment (R_{j-1}; $r = 0.02$), (ii) change in repayment (ΔR_{j-1}; $r = 0.10$), and (iii) previous trustee reciprocity ($\Delta R_{j-1} - \Delta I_{j-1}$; $r = 0.31$). Three predictors of ΔR_j were examined: (i) previous investment (I_j; $r = 0.10$), (ii) change in investment (ΔI_j; $r = 0.26$), and (iii) previous investor reciprocity ($\Delta I_j - \Delta R_{j-1}$; $r = 0.56$). Thus, reciprocity was a stronger predictor than either amount previously sent (I_j or R_{j-1}) or change in amount previously sent (ΔI_j or ΔR_{j-1}). However, it is noteworthy that reciprocity expressed by the investor ($r = 0.56$) was more strongly related to change in trust than reciprocity expressed by the trustee ($r = 0.26$). This difference is likely accounted for by an asymmetry in the structure of the exchange: In each round, the investor can accumulate money ($20 endowment) without the cooperation of the trustee,

whereas the trustee is wholly dependent on the investor's cooperation. This dependency of the trustee on the investor likely results in greater responsivity by the trustee to changes in investor reciprocity.
21. A description of methods is available as supporting material in *Science* Online.
22. Each dyad contributed eight behavioral events to this analysis (48 pairs × 8 rounds = 384 rounds). Investor reciprocity cannot be calculated for the initial two rounds and was excluded. The 384 rounds had a mean ± SD of −0.01 ± 0.35, skewness of −0.19 (SE = 0.12), and kurtosis of 2.55 (SE = 0.25). Rounds were divided into approximately equal-sized categories: 125 malevolent reciprocity rounds ($x < -0.025$), 134 neutral reciprocity rounds ($-0.025 \leq x \leq +0.05$), and 125 benevolent reciprocity rounds ($x > +0.05$). For additional description of reciprocity categories, see figs. S3 and S4.
23. Regions with ≥10 significant voxels were identified using *t* tests. *Z* values and statistical parametric mapping (SPM) coordinates for each region are available in table S1.
24. The correlation of change in investment (ΔI_j) and subsequent change in repayment (ΔR_j) grew as experience between players accrued (fig. S5).
25. P. Dayan, L. F. Abbott, *Theoretical Neuroscience* (MIT Press, Cambridge, MA, 2001).
26. K. C. Berridge, in *The Psychology of Learning and Motivation*, D. L. Medin, Ed. (Academic Press, New York, 2000), pp. 223–278.
27. A. Dickinson, B. W. Balleine, in *Steven's Handbook of Experimental Psychology*, C. R. Gallistel, Ed. (Wiley, New York, 2002), vol. 3, pp. 26–72.
28. G. Pagnoni, C. F. Zink, P. R. Montague, G. S. Berns, *Nat. Neurosci.* **5**, 97 (2002).
29. S. M. McClure, G. S. Berns, P. R. Montague, *Neuron* **38**, 339 (2003).
30. J. P. O'Doherty, P. Dayan, K. Friston, H. Critchley, R. J. Dolan, *Neuron* **38**, 329 (2003).
31. J. O'Doherty *et al.*, *Science* **304**, 452 (2004).
32. B. Seymour *et al.*, *Nature* **429**, 664 (2004).
33. W. Schultz, P. Dayan, P. R. Montague, *Science* **275**, 1593 (1997).
34. P. R. Montague, S. E. Hyman, J. D. Cohen, *Nature* **431**, 760 (2004).
35. R. C. O'Reilly, T. S. Braver, J. D. Cohen, in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, A. Miyake, P. Shah, Eds. (Cambridge Univ. Press, New York, 1999), chap. 11, pp. 375–411.
36. K.-H. Lee, T. F. D. Farrow, S. A. Spence, P. W. R. Woodruff, *Psychol. Med.* **34**, 391 (2004).
37. E. L. Hill, U. Frith, *Philos. Trans. R. Soc. London Ser. B* **358**, 281 (2003).
38. P. A. Johnson, R. A. Hurley, C. Benkelfat, S. C. Herpertz, K. H. Taber, *J. Neuropsychiatry Clin. Neurosci.* **15**, 397 (2003).
39. This work was supported by the Center for Theoretical Neuroscience at Baylor College of Medicine (P.R.M.), National Institute on Drug Abuse (NIDA) grant DA11723 (P.R.M.), National Institute of Neurological Disorders and Stroke grant NS045790 (P.R.M.), National Institute of Mental Health grant MH52797 (P.R.M.), NIDA grant DA14883 (G. Berns), The Kane Family Foundation (P.R.M.), The David and Lucile Packard Foundation (S.R.Q.), and The Gordon and Betty Moore Foundation (S.R.Q.). We thank P. Dayan, J. Li, T. Lohrenz, C. Stetson, and two anonymous referees for comments on this manuscript. We thank the Hyperscan Development Team at Baylor College of Medicine for Network Experiment Management Object (NEMO) software implementation (www.hnl.bcm.tmc.edu/nemo) and G. Berns for early discussions and efforts leading to the development of hyperscanning. We also thank A. Harvey, S. Flaherty, K. Pfeiffer, R. Pruitt, and S. Gleason for technical assistance.

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5718/78/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References

30 November 2004; accepted 7 February 2005
10.1126/science.1108062

Postsynaptic Receptor Trafficking Underlying a Form of Associative Learning

Simon Rumpel,[1] Joseph LeDoux,[2] Anthony Zador,[1] Roberto Malinow[1]*

To elucidate molecular, cellular, and circuit changes that occur in the brain during learning, we investigated the role of a glutamate receptor subtype in fear conditioning. In this form of learning, animals associate two stimuli, such as a tone and a shock. Here we report that fear conditioning drives AMPA-type glutamate receptors into the synapse of a large fraction of postsynaptic neurons in the lateral amygdala, a brain structure essential for this learning process. Furthermore, memory was reduced if AMPA receptor synaptic incorporation was blocked in as few as 10 to 20% of lateral amygdala neurons. Thus, the encoding of memories in the lateral amygdala is mediated by AMPA receptor trafficking, is widely distributed, and displays little redundancy.

Animals continually adapt their behavior in response to changes in the environment. It has long been held that selective modifications in synaptic efficacy represent the physical substrate for this behavioral plasticity (*1*, *2*). Long-term potentiation (LTP), a cellular model of synaptic plasticity, has emerged as a leading candidate mechanism underlying associative forms of learning in the central nervous system (*3*–*12*). Much is now known about the molecular mechanisms during LTP that translate a brief change in electrical activity patterns to a modification in synaptic efficacy (*13*–*23*). Recent studies indicate that synaptic addition of GluR1 subunit–containing AMPA-type glutamate receptors (GluR1-receptors) mediates the synaptic strengthening observed during LTP (*24*, *25*). An attractive

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [2]New York University, New York, NY 10003, USA.

*To whom correspondence should be addressed. E-mail: malinow@cshl.edu

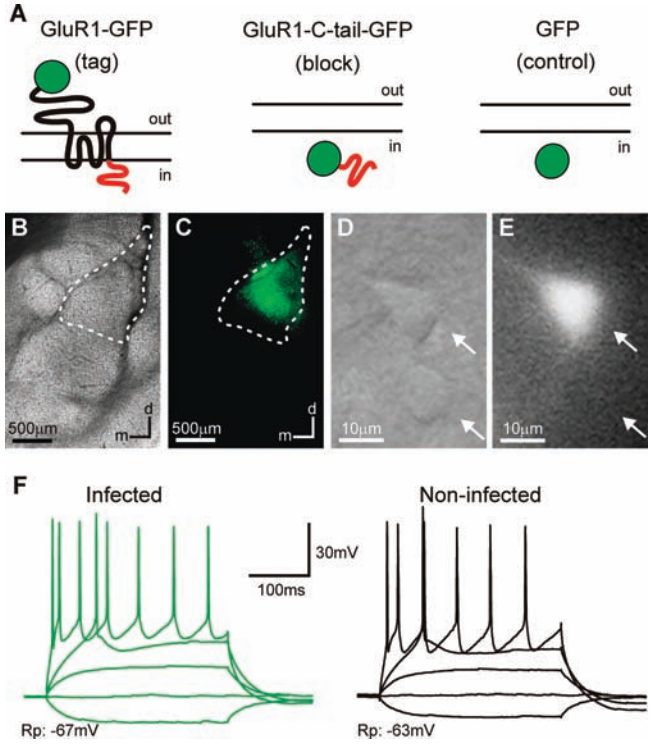

Fig. 1. Viral infection with amplicon vectors does not alter basic electrophysiological properties. (**A**) Schematic of recombinant proteins used in this study: GluR1-GFP, a fusion protein of GFP and the GluR1 subunit; GluR1-C-tail–GFP, a fusion protein of GFP and the last C-terminal 81 amino acids of the GluR1 subunit; and GFP alone. (**B** and **C**) Low magnification transmitted light (B) and epifluorescence (C) images of a coronal section of the right hemisphere including the amygdala. Note the area of GFP-expressing cells within the lateral amygdala (dotted line) 1 day after injection. d, dorsal; m, medial. (**D** and **E**) Highly magnified image of the lateral amygdala by infrared-differential interference contrast microscopy (D) and epifluorescence (E), which contains a neuron expressing (upper arrow) or not expressing (lower arrow) GFP. (**F**) Superimposed current-clamp recordings of an infected (green traces) and noninfected (black traces) neuron during 300-ms current injections of −100, 0, +100, +200, and +550pA. Rp, resting potential of neurons indicated next to traces.