



Published in final edited form as:

Science. 2008 August 8; 321(5890): 806–810. doi:10.1126/science.1156902.

## The Rupture and Repair of Cooperation in Borderline Personality Disorder

Brooks King-Casas<sup>1,2</sup>, Carla Sharp<sup>2</sup>, Laura Lomax-Bream<sup>2</sup>, Terry Lohrenz<sup>1</sup>, Peter Fonagy<sup>2,3,4</sup>, and P. Read Montague<sup>1,2,\*</sup>

<sup>1</sup>Computational Psychiatry Unit and Department of Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA

<sup>2</sup>Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA

<sup>3</sup>Research Department of Clinical, Educational, and Health Psychology, University College London, Gower Street, London WC1E 6BT, UK

<sup>4</sup>Anna Freud Centre, London NW3 5SD, UK

### Abstract

To sustain or repair cooperation during a social exchange, adaptive creatures must understand social gestures and the consequences when shared expectations about fair exchange are violated by accident or intent. We recruited 55 individuals afflicted with borderline personality disorder (BPD) to play a multiround economic exchange game with healthy partners. Behaviorally, individuals with BPD showed a profound incapacity to maintain cooperation, and were impaired in their ability to repair broken cooperation on the basis of a quantitative measure of coaxing. Neurally, activity in the anterior insula, a region known to respond to norm violations across affective, interoceptive, economic, and social dimensions, strongly differentiated healthy participants from individuals with BPD. Healthy subjects showed a strong linear relation between anterior insula response and both magnitude of monetary offer received from their partner (input) and the amount of money repaid to their partner (output). In stark contrast, activity in the anterior insula of BPD participants was related only to the magnitude of repayment sent back to their partner (output), not to the magnitude of offers received (input). These neural and behavioral data suggest that norms used in perception of social gestures are pathologically perturbed or missing altogether among individuals with BPD. This game-theoretic approach to psychopathology may open doors to new ways of characterizing and studying a range of mental illnesses.

---

Copyright 2008 by the American Association for the Advancement of Science; all rights reserved.

\*To whom correspondence should be addressed. rmontague@hnl.bcm.edu.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/321/5890/806/DC1](http://www.sciencemag.org/cgi/content/full/321/5890/806/DC1)

Materials and Methods

Figs. S1 and S6

Tables S1 to 5

References

Theoretical and empirical work on cooperation has made tremendous strides by using game theory to provide a normative mathematical setting for understanding social exchange (1–5). This same game-theoretic framework has exposed some of the basic computations underlying social interaction and identified neural correlates of cooperation, reciprocity, and social signaling (6–14). Collectively, this work raises the possibility of a new approach to characterizing and understanding psychopathology from a normative perspective. In conditions ranging from psychosis to developmental and personality disorders, afflicted individuals often display a dramatically perturbed capacity to model others and to sense and respond appropriately to the social signals they emit (15–17). Consequently, using normative game-theoretic probes of social signaling in identified psychopathologies offers the opportunity to understand some of the components of these disorders in terms of malfunctioning computations (18, 19).

We examined two groups of individuals that vary in their capacity to maintain stable interpersonal relationships and used a multiround economic exchange game (Fig. 1) (20) to probe two fundamental components underlying the capacity to sustain a successful social exchange: cooperation and its repair. Specifically, we studied a group of control individuals and a second group of individuals diagnosed with borderline personality disorder (BPD), a psychiatric disorder characterized by unstable relationships, affective dysregulation, and a broad incapacity to trust appropriately the actions and (possible) motives of others (21–23).

Successful cooperation between two agents requires a range of intact computations including the capacity to sense, value, and respond to social gestures exchanged with one's partner (24–26). However, cooperative exchange is fragile and can easily be ruptured by accident (e.g., through impoverished models of social partners) or intent (27). The successful repair of broken cooperation in an iterated exchange requires the capacity to coax one's partner back into cooperation through the medium of generous gestures (28–32). In an economic exchange game, such gestures are encoded as money units, which exposes their immediate cost to each subject. However, for such gestures to be meaningful, cooperating individuals must possess norms that accurately encode what is expected of themselves and their partners, and they must sense when these norms have been significantly violated (33). We pursued the hypothesis that social exchange norms of subjects with BPD are either pathologically perturbed or missing altogether, preventing sustained cooperation or the repair of cooperation after it breaks down.

Before scanning, all participants (healthy subject group and BPD group) underwent diagnostic and symptom assessment and were matched on demographic variables including sex, age, education, and verbal IQ (see table S1) (20). Within each trust exchange, cooperation occurs when an investor and trustee act in a manner that mutually benefits both players. For example, if an investor sends \$20 to a trustee, and the trustee splits the tripled investment (\$60) with the investor, both the investor and trustee profit. The investor earns \$10 more, and the trustee earns \$30 more, than if the investor had sent nothing. However, if a trustee does not repay at least the amount invested, the investor accrues no benefit from the exchange, likely triggering smaller subsequent investments. Thus, increased cooperation is seen with increased money exchanged across the course of the 10-round game. In the current study, subjects diagnosed with BPD played the trustee role against a healthy investor; as

expected, each game began normally with the healthy investor showing a level of trust comparable to that of healthy investors playing healthy trustees (Fig. 1).

### **BPD subjects send signals that break cooperation**

In early rounds of this multiround game, investment levels did not differ between subjects partnered with BPD trustees and subjects partnered with healthy trustees. However, in late rounds of the game, investments were significantly lower for dyads with a BPD trustee relative to control dyads with a healthy trustee (Fig. 1). This definitive downward shift in investment levels for dyads with BPD trustees reflects a break in cooperation [see figs. S2 to S4 and (34) for additional detailed analyses]. We strongly suspected that this breakdown was caused by the social signals emitted by the BPD trustee and were not due to a sampling of healthy investors that happened to be uncooperative. To test this claim empirically, we recruited an additional cohort of 68 healthy investors to play an adaptive computer agent designed to play as either a BPD trustee or a healthy trustee. A significant within-subject effect confirmed that BPD trustee behavior alone elicits a breakdown in cooperation (see fig. S5) (20).

### **Coaxing pays into the future**

Although the BPD subjects showed a consistent tendency to rupture cooperation with their healthy partners, the issue of broken cooperation arises during all interpersonal exchange. Individuals bring to any two-party interaction variability in their expectations of partner behavior, variable sensitivities to deviations from these norms, and variable responses to these deviations. These issues highlight a feature that plagues all social exchange—breakdowns in cooperation are common, and successful social agents must sense (or even anticipate) these failures and select actions to repair them (28–32). Consequently, we specifically sought to identify strategic responses to cooperation breakdowns that allow healthy trustees to maintain high levels of cooperation into the later rounds of the exchange. We focused on rounds in which cooperation was low, that is, rounds in which investments of \$5 or less were made. Low investments act as a viable proxy for broken cooperation because in this economic exchange game and other related games (7, 9, 10, 35), opening investments to a responder (here the trustee) are typically large. For investments to become small, a responder has to “prove” that they are not worthy of a large investment by acting in an untrustworthy manner. Such ruptures of cooperation, nevertheless, provide the trustee an opportunity to repair the broken cooperation. If trustees repay a large fraction of the tripled investment, they signal their trustworthiness in the presence of low offers and may garner greater investments on subsequent rounds (Fig. 2A). We term this behavior “coaxing” and show in Fig. 2B that coaxing pays dividends for four rounds into the future. In this same figure, we summarize the results for low offers when the trustees do not coax—trust and cooperation remain broken. So in the context of this multiround game, coaxing repairs cooperation and pays off generously for the coxer. Remarkably, comparison of healthy trustees to BPD trustees found healthy players to be almost twice as likely as BPD players to coax in the presence of low offers (i.e., repay a third or more of the investment) (Fig. 2C and fig. S3).

## BPD subjects have perturbed norm deviation responses in insular cortex

These behavioral data with humans and computer agents suggest strongly that BPD subjects are likely to emit responses that cause cooperation to rupture, and, once ruptured, they show substantially diminished rates of coaxing (generous gestures) to repair the break (Fig. 2C). A question naturally arises about whether there is a comprehensible and consistent neural correlate of the rupture in cooperation that serves to cue the repair (or not) of cooperation. To investigate the possible neural origins of failures in cooperation and the lack of coaxing, we first sought between-group differences in the sensitivity of trustee brains to the revelation of small relative to large investments. A general linear model (GLM) analysis identified several regions that significantly differentiated the controls from the BPD individuals, including bilateral anterior insula, along with frontal and parietal areas (Fig. 3, top left; also see table S3 and fig. S6). Subsequent within-subject analyses identified only a single region in controls with greater response to small, relative to large investments—bilateral anterior insula. In particular, both a GLM analysis (Fig. 3, top middle; also see table S4) and region-of-interest (ROI) analysis of voxels identified in the between-group comparison show a strong negative relation between bilateral anterior insula activity and investment level (Fig. 3, bottom middle). In stark contrast, in BPD subjects comparable GLM and ROI analyses (Fig. 3, top right and bottom right, respectively; also see table S5) showed no systematic relation of insula response and investment level. Additional ROI analyses of medicated and unmedicated BPD participants (Fig. 4, left and middle), and BPD participants matched on income level (Fig. 4, right) to healthy trustees confirm that the diminished response cannot be attributed to group differences in sex, age, verbal IQ, education, income level, or medication status.

## Repayment-related insula activity intact in BPD subjects

In earlier work using this same multiround trust game (35) and single-shot ultimatum games (12), small offers increase the probability of defection in the partner, and such behavior is preceded by increased activity in the anterior insular cortex [also see (36) for review]. Consequently, we pursued an additional region-of-interest analysis of the anterior insular cortex to examine whether activity in this region predicts small repayments by trustees (Fig. 5). Using the same voxels originally identified in the group by investment-level GLM, we found that both healthy and BPD trustees show a negative relation between insula activity and repayment level during trustee decisions. That is, despite the diminished insula response exhibited by BPD trustees during the receipt of small offers (input), a robust insula response in this group is observed preceding decisions to repay small amounts (output). Thus, healthy subjects show a strong linear relation between anterior insula response and both the monetary offer sent from their partner [input (Fig. 3)] and their monetary response back to their partner [output (Fig. 5)], whereas BPD subjects' insula yielded a linear relationship only to their response back to their partner [output (Fig. 5)], but showed no differential activity upon receipt of offers from their partner (Fig. 3). Moreover, the BPD insula response, on receipt of offers from the investor, showed the same average level of activity across all investment levels, a fact true for both medicated and unmedicated subjects (Fig. 4).

Although the insular cortex has traditionally been associated with pain perception, representation, and interoception (37, 38), responses in the anterior insula have recently been identified in social interactions where norms are violated (12, 14, 39, 40). In an ultimatum game, in which one player offers to split an endowment with a second player, small offers are perceived as unfair and typically refused (41–43). In such instances, anterior insula activity both scales negatively with offer size and predicts whether the offer is subsequently rejected (12). In a related finding, anterior insula activity of observers is greater when a punishment is applied to players perceived as fair relative to players perceived as unfair (13). In nonsocial tasks, activity in the anterior insula has been found to encode representations of risk and uncertainty about decision outcomes (44–50). The association of the insula with a representation of outcome variance suggests that the insula may encode the distribution of likely outcomes in social interactions; that is, responses in the anterior insula may indicate social norm violations within interpersonal contexts. Furthermore, violations of such norms could serve as control signals that update expectations about social partners or at least inform learning about a subset of parameters associated with one's partner. This possibility is supported by the work of Preuschoff and Bossaerts, who have recently reported prediction errors of risk to evoke strong responses in the bilateral anterior insula, consistent with this speculation (49–51). Taken together, these data suggest that the strong negative relation between offer size and activity in the anterior insula seen in healthy trustees reflects the perceived violation of social norms in the two-party trust exchange.

Moreover, in BPD, the robust insula response preceding one's own repayments is discordant with the absent insula relation to the investments of others and highlights the specificity of the observed deficit in BPD. That is, in BPD, the apparent insensitivity of the insula only to offer level size, and not their own repayment, suggests two possibilities: (i) Monetary reward is not reinforcing to individuals with BPD; or, (ii) low offers are not perceived to be a violation of social norms to individuals with BPD. Previous work has shown monetary reward to be a potent reinforcer in natural and laboratory settings in this group (52), which makes the former less likely. Furthermore, studies of interpersonal and emotional processing in BPD suggest that this group holds negative expectations of social partners and exhibits negative evaluative biases (53, 54), consistent with the presence of atypical social norms.

### **Economic probe corresponds with self-report measure of trust**

To compare social norms of healthy trustees to BPD trustees within the current study, we also implemented a self-report measure of trust (55) that gauges expectations across a variety of social situations and with a variety of social agents. Individuals with BPD expressed significantly lower levels of self-reported trust relative to healthy controls [ $P < 0.005$  (Fig. 6)], a finding that agrees with the diminished trust exhibited by this group in the economic exchange. Together, these results suggest that the diminished insula response to low offers does indeed reflect atypical social norms in this group. Put another way, the low offers from social partners do not violate the social expectations of individuals with BPD, which accounts for the diminished insula response in the BPD group.

The atypical social norms of individuals with BPD have considerable implications for both partners in a repeated cooperative exchange. Specifically, the incapacity of individuals with

BPD to recognize norm violations as such and to respond so as to signal trustworthiness to their partner can result in an inability of both individuals to effectively model the intentions of one another. For example, an investor might send a small amount to a trustee to signal lack of trust and may do so with the expectation that the trustee will, in turn, signal a desire to cooperate through a large repayment (coaxing). However, the lack of expected response (failure to coax) may inadvertently signal additional lack of cooperation, regardless of the trustee's intent. In short, the atypical social norms of a single partner can have deleterious effects on the ability of both players to model and predict the actions of social partners and, thus, to inhibit both players' ability to mutually benefit from cooperative exchange.

A number of studies have utilized resting-state and emotional challenge paradigms to investigate personality disorders; however, the current study is the first large-scale investigation of interpersonal behavior among individuals diagnosed with a personality disorder (Axis-II) (22) using a normative economic exchange game. The critical role that interpersonal deficits play across a variety of such disorders, along with the substantive contribution that neuroimaging studies have made in elucidating the etiology of Axis-I clinical disorders, recommend the future use of game-theoretic probes of interpersonal interaction to uncover the neural processes underlying social pathologies (18, 19, 56).

## Acknowledgments

This work was supported by the Child and Family Program at the Menninger Clinic, E. Bleiberg, Director (P.F., L.L., C.S), National Institute of Mental Health (NIMH) grant MH078485 (B.K.); NIMH grant MH52797 (P.R.M.); National Institute of Neurological Disorders and Stroke grant NS045790 (P.R.M.); and National Institute on Drug Abuse grant DA11723 (P.R.M). We thank P. Chiu for comments on this manuscript. We also thank R. Bhimani, C. Bracero, C. Ha, A. Jarman, O. Mosko, and technical staff of the Human Neuroimaging Laboratory at Baylor College of Medicine. We thank referring clinicians affiliated with the Menninger Department of Psychiatry at Baylor College of Medicine for assistance in recruitment, including E. Bleiberg, G. Gabbard, S. Twemlow, and E. Weinberg.

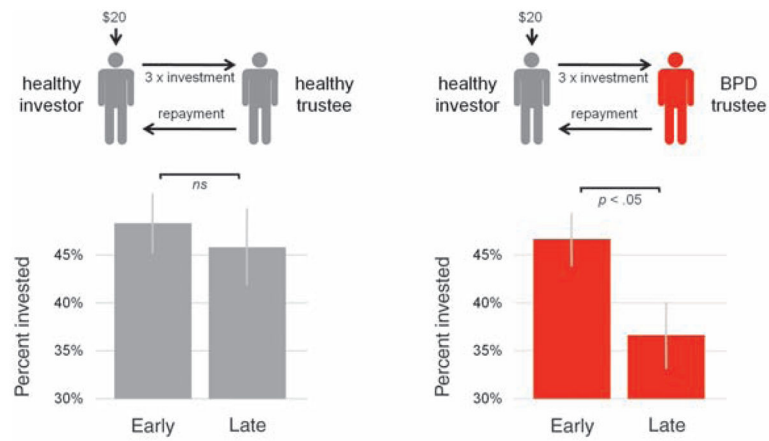
## References and Notes

1. Axelrod R, Hamilton WD. *Science*. 1981; 211:1390. [PubMed: 7466396]
2. Gintis, H. *Game Theory Evolving*. Princeton Univ. Press; Princeton, NJ: 2000.
3. Nowak, MA. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard Univ. Press; Cambridge, MA: 2006.
4. Trivers RL. *Q Rev Biol*. 1971; 46:35.
5. von Neumann, J.; Morgenstern, O. *Theory of Games and Economic Behavior*. 2. Princeton Univ. Press; Princeton, NJ: 1947.
6. de Quervain DJ, et al. *Science*. 2004; 305:1254. [PubMed: 15333831]
7. Delgado MR, Frank RH, Phelps EA. *Nat Neurosci*. 2005; 8:1611. [PubMed: 16222226]
8. Knoch D, Pascual-Leone A, Meyer K, Treyer V, Fehr E. *Science*. 2006; 314:829. [PubMed: 17023614]
9. McCabe K, Houser D, Ryan L, Smith V, Trouard T. *Proc Natl Acad Sci USA*. 2001; 98:11832. [PubMed: 11562505]
10. Krueger F, et al. *Proc Natl Acad Sci USA*. 2007; 104:20084. [PubMed: 18056800]
11. Rilling J, et al. *Neuron*. 2002; 35:395. [PubMed: 12160756]
12. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. *Science*. 2003; 300:1755. [PubMed: 12805551]
13. Singer T, et al. *Nature*. 2006; 439:466. [PubMed: 16421576]

14. Spitzer M, Fischbacher U, Herrnberger B, Grön G, Fehr E. *Neuron*. 2007; 56:185. [PubMed: 17920024]
15. Sprong M, Schothorst P, Vos E, Hox J, Van Engeland H. *Br J Psychiatry*. 2007; 191:5. [PubMed: 17602119]
16. Brüne M, Brüne-Cohrs U. *Neurosci Biobehav Rev*. 2006; 30:437. [PubMed: 16239031]
17. Rogers J, Viding E, Blair RJ, Frith U, Happé F. *Psychol Med*. 2006; 36:1789. [PubMed: 17018169]
18. Chiu PH, et al. *Neuron*. 2008; 57:463. [PubMed: 18255038]
19. Rilling JK, et al. *Biol Psychiatry*. 2007; 61:1260. [PubMed: 17046722]
20. Materials and methods are available as supporting material on *Science Online*.
21. Lieb K, Zanarini MC, Schmahl C, Linehan MM, Bohus M. *Lancet*. 2004; 364:453. [PubMed: 15288745]
22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. 4. American Psychiatric Association; Washington, DC: 2000. text revised
23. Fonagy P, Bateman AW. *J Clin Psychol*. 2006; 62:411. [PubMed: 16470710]
24. Frith CD, Frith U. *Brain Res*. 2006; 1079:36. [PubMed: 16513098]
25. Singer T. *Neurosci Biobehav Rev*. 2006; 30:855. [PubMed: 16904182]
26. Sanfey AG, Loewenstein G, McClure SM, Cohen JD. *Trends Cogn Sci*. 2006; 10:108. [PubMed: 16469524]
27. Axelrod R, Dion D. *Science*. 1988; 242:1385. [PubMed: 17802133]
28. Bendor J, Kramer RM, Stout S. *J Conflict Resolut*. 1991; 35:691.
29. Fudenberg D, Maskin E. *Am Econ Rev*. 1990; 80:274.
30. Wu J, Axelrod R. *J Conflict Resolut*. 1995; 39:183.
31. Nowak MA, Sigmund K. *Nature*. 1992; 355:250.
32. Wedekind C, Milinski M. *Proc Natl Acad Sci USA*. 1996; 93:2686. [PubMed: 11607644]
33. Fehr E, Camerer CF. *Trends Cogn Sci*. 2007; 11:419. [PubMed: 17913566]
34. To explore whether lower levels of tit-for-tat reciprocity may have contributed to such failures of cooperation, linear regressions of behavior between partners were carried out. Consistent with previous work, reciprocity was found to be a strong predictor of subsequent increase or decreases in amount of money sent (table S2). Reciprocity is defined as the fractional change in money sent across rounds by one player in response to the fractional change in money sent across rounds by their partner. Thus, investor reciprocity on round  $j$  was calculated as  $I_j - R_{j-1}$ , where  $I_j$  indicates the fractional change in investment from round  $j - 1$  to round  $j$  and  $R_{j-1}$  indicates the fractional change in repayment from round  $j - 2$  to round  $j - 1$ . Although deviation in perfect tit-for-tat behavior was a significant predictor of change in partner behavior across groups, the strength of this relation did not differ between pairs that included healthy trustees and pairs that included BPD trustees (table S2), which suggests sensitivity and responsiveness to tit-for-tat behavior is intact among individuals with BPD.
35. King-Casas B, et al. *Science*. 2005; 308:78. [PubMed: 15802598]
36. Seymour B, Singer T, Dolan R. *Nat Rev Neurosci*. 2007; 8:300. [PubMed: 17375042]
37. Craig AD. *Curr Opin Neurobiol*. 2003; 13:500. [PubMed: 12965300]
38. Craig AD. *Annu Rev Neurosci*. 2003; 26:1. [PubMed: 12651967]
39. Montague PR, Lohrenz T. *Neuron*. 2007; 56:14. [PubMed: 17920011]
40. Paulus MP. *Science*. 2007; 318:602. [PubMed: 17962553]
41. Camerer, C. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton Univ. Press; Princeton, NJ: 2003.
42. Fehr E, Schmidt KM. *Q J Econ*. 1999; 114:817.
43. Güth W, Schmittberger R, Schwarze B. *J Econ Behav Organ*. 1982; 3:367.
44. Berns GS, et al. *Science*. 2006; 312:754. [PubMed: 16675703]
45. Critchley HD, Mathias CJ, Dolan RJ. *Neuron*. 2001; 29:537. [PubMed: 11239442]
46. Huettel SA, Stowe CJ, Gordon EM, Warner BT, Platt ML. *Neuron*. 2006; 49:765. [PubMed: 16504951]

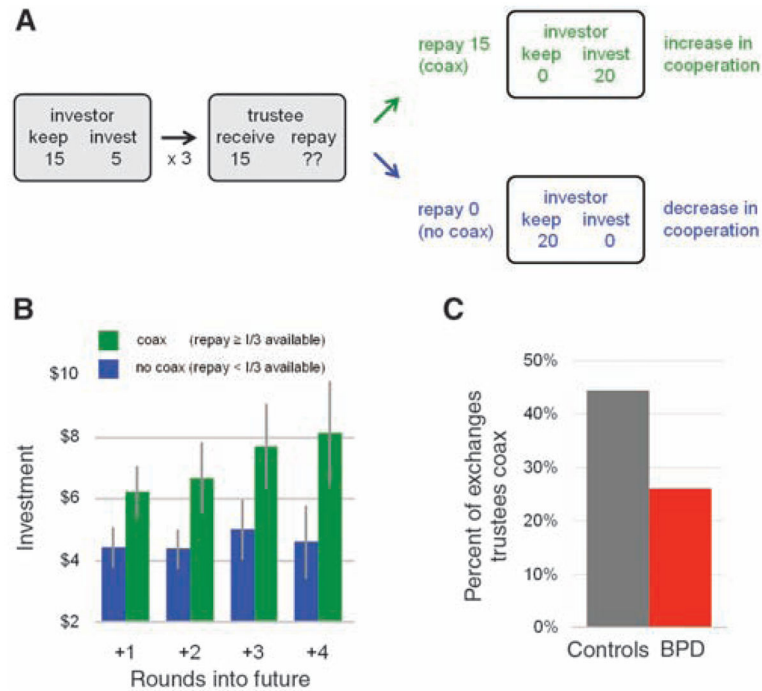
47. Simmons A, Matthews SC, Paulus MP, Stein MB. *Neurosci Lett*. 2008; 430:92. [PubMed: 18079060]
48. Paulus MP, Rogalsky C, Simmons A, Feinstein JS, Stein MB. *Neuroimage*. 2003; 19:1439. [PubMed: 12948701]
49. Preuschoff K, Bossaerts P, Quartz SR. *Neuron*. 2006; 51:381. [PubMed: 16880132]
50. Preuschoff K, Bossaerts P. *Ann N Y Acad Sci*. 2007; 1104:135. [PubMed: 17344526]
51. Preuschoff K, Quartz S, Bossaerts P. *J Neurosci*. 2008; 28:2745. [PubMed: 18337404]
52. Dougherty DM, Bjork JM, Huckabee HCG, Moeller FG, Swann AC. *Psychiatry Res*. 1999; 85:315. [PubMed: 10333383]
53. Putnam KM, Silk KR. *Dev Psychopathol*. 2005; 17:899. [PubMed: 16613424]
54. Sieswerda S, Arntz A, Wolfis M. *J Behav Ther Exp Psychiatry*. 2005; 36:209. [PubMed: 15950931]
55. Rotter JB. *J Pers*. 1967; 35:651. [PubMed: 4865583]
56. Todorov A, Harris LT, Fiske ST. *Brain Res*. 2006; 1079:76. [PubMed: 16480692]





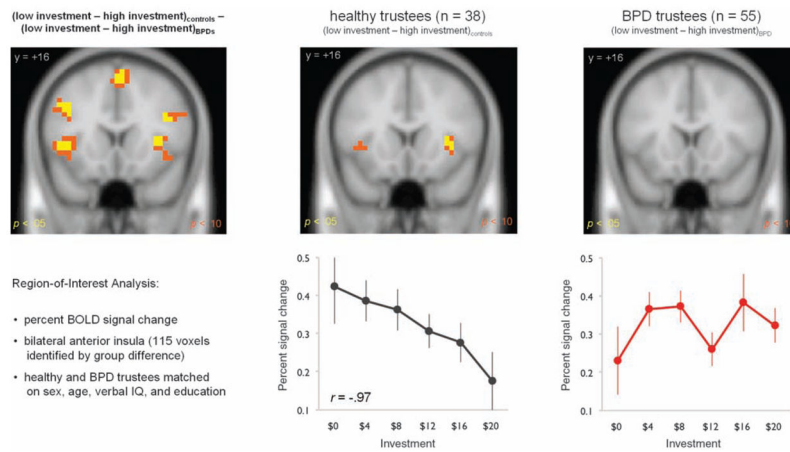
**Fig. 1.**

Cooperation fails across rounds when individuals with BPD engage in a repeated exchange of trust. Among 38 healthy investors (gray) paired with healthy trustees (gray), investments were large and sustained across early (1 to 5) rounds and late (6 to 10) rounds of the game. However, among 55 healthy investors (gray) paired with 55 trustees with BPD (red), a decrease in investment level from early to late rounds of the game indicates a failure in cooperation across the iterated exchange. Mean percent invested and SEM are plotted.

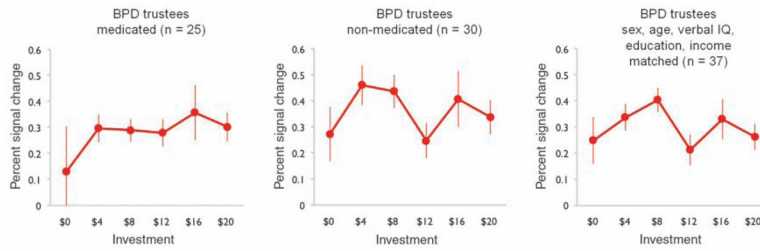
**Fig. 2.**

(A) Coaxing by trustees engenders trust from investors. Cooperation within the iterated trust game is expressed as high levels of investment combined with equitable sharing of investment returns (see fig. S2 for normative behavior). When cooperation falters (i.e., when investments are low), trustees can rebuild trust and cooperation by coaxing back higher and higher investment levels. In this schematic representation, the investor entrusts only a small portion of their endowment (\$5 of \$20 points) to their partner. The \$15 received by the trustee is far less than the \$60 that would have been received if the investor had sent the entire endowment. Thus, in order to maximize their gains, a trustee must induce cooperation in a repeated exchange. The trustee can coax back trust from their partner by repaying a large proportion of the tripled investment—signaling their own trustworthiness and, thus, eliciting larger subsequent investments. Conversely, if the trustee does not coax and instead keeps a large proportion of the tripled investment, the investor is likely to invest less on subsequent rounds and cooperation will continue to devolve. (B) Coaxing pays off. The effect of such coaxing is summarized in the effect of repayment on investments in subsequent rounds among healthy subjects. This analysis is restricted to low investments (all rounds with investments <\$5). Blue bars depict mean  $\pm$  SE investment level following small repayments (repayment less than one-third of the tripled investment; “no coax” condition); green bars depict investment level following large repayments (more than or equal to one-third; “coax” condition). The analysis indicates that coaxing elicits larger investments in rounds following coaxing behavior (green) than for rounds following no coaxing (blue). Note that the coaxing-related increase in investment persists across rounds, such that a large repayment in the current round earns larger investment up to four rounds into the future. (C) BPD subjects coax less. Healthy trustees are twice as likely as BPD trustees to coax when cooperation between players is low. Specifically, healthy trustees are more likely to make a

large repayment ( = investment amount) after having received a small investment ( = \$5). Conversely, BPD trustees are more likely to make a small repayment (< investment amount) after receiving a small investment. The y axis indicates the proportion of exchanges that trustees repay more than or equal to the investment amount after receiving a small investment ( = \$5).

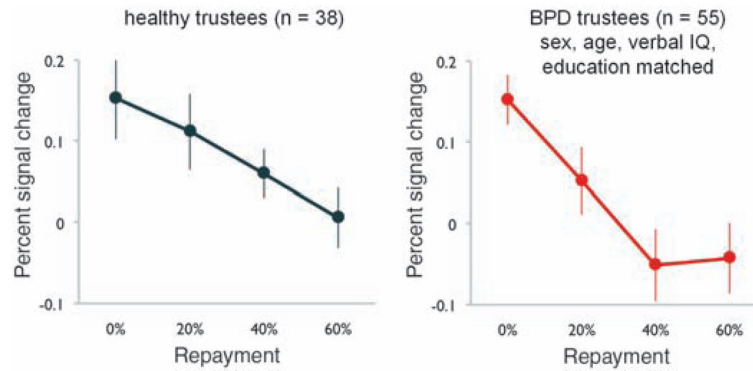
**Fig. 3.**

Response of 38 healthy trustee brains and 55 BPD trustee brains to level of cooperation. **(Top)** Results of both between- and within-group GLM analyses identified cortical regions with greater response to small investments ( $I \leq \$5$ ) relative to large investments ( $I > \$10$ ). All significant voxels are conservatively corrected using false-discovery rate procedures to the  $P < 0.05$  level in yellow;  $P < 0.10$  in orange (see tables S3 to S5). **(Bottom)** Percent change in hemodynamic signal was averaged from the 115 most significant voxels identified in the group-level GLM (top left and table S3) during the 4- to 8-s period following the revelation of investment. The means  $\pm$  SE of the resulting signal are plotted in \$4 bins. Responses in bilateral anterior insula in healthy trustees scale strongly and negatively with the size of investment ( $r = -0.97$ ; bottom middle). In contrast, similar analyses in individuals with BPD showed no such relation.



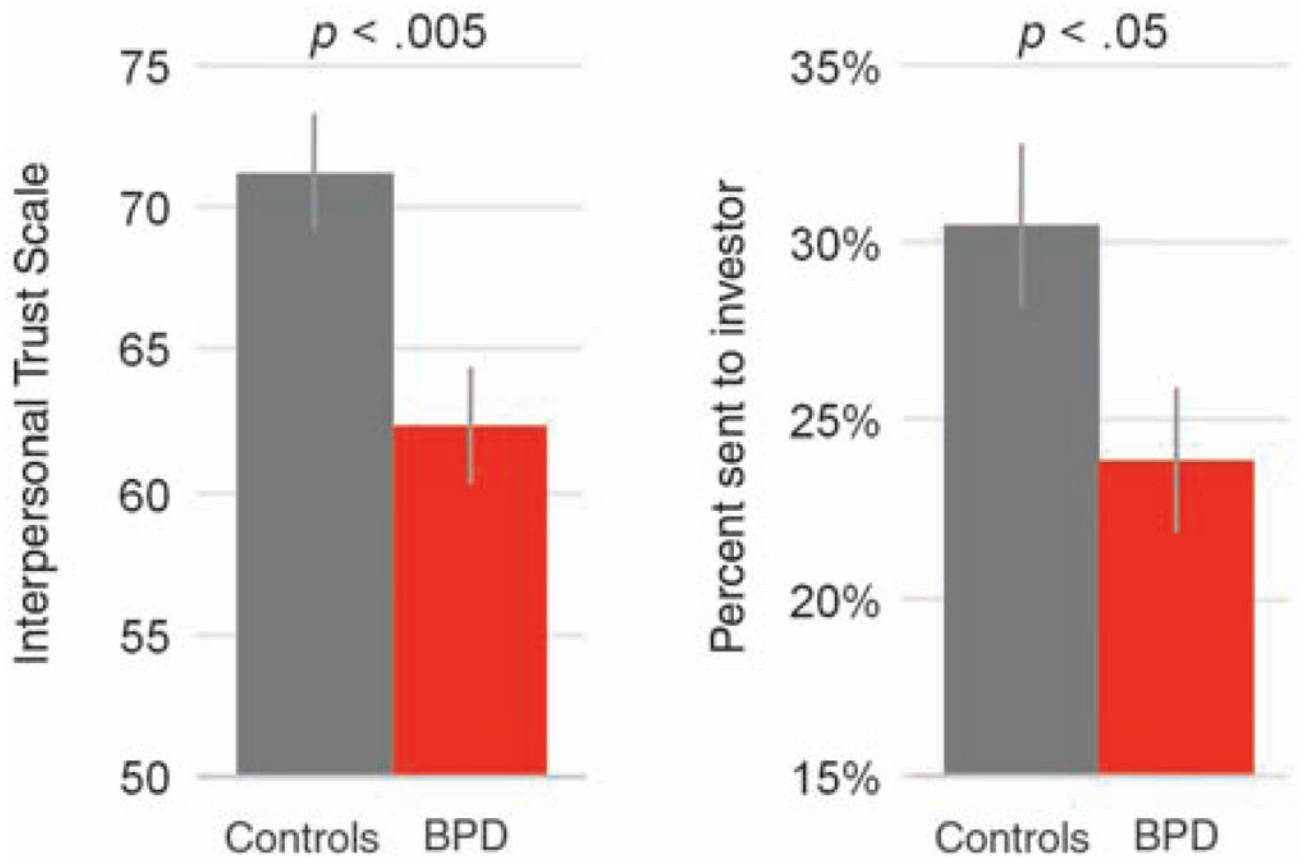
**Fig. 4.**

Diminished relation of insula activity to investment level is not attributable to medication status or income level in BPD trustees. ROI analyses within subsamples of BPD trustees were performed as described in Fig. 3. The systematic negative relation of investment level to bilateral anterior insula activity seen in healthy controls is absent in both medicated and unmedicated BPD trustees, as well as in BPD trustees matched to the income level of the control group.



**Fig. 5.**

Both healthy and BPD trustees exhibit a systematic negative relation of repayment level and bilateral anterior insula activity during their decision to repay. ROI analyses were performed using bilateral anterior insula voxels identified in the between-group analysis illustrated in Fig. 3 and used in the ROI analyses of Figs. 3 and 4. To investigate whether the insula activity also scales with trustee responses to norm violations, insula activity during trustee decisions is plotted by repayment level. Both healthy and BPD trustees show a strong negative relation between activity and repayment level. As trustees repaid more than 60% of the tripled investment in only ~ 7% of rounds, high repayment rounds are excluded from this analysis.



**Fig. 6.** Behavioral and self-report levels of trust indicate perturbed social norms among individuals with BPD. **(Left)** Individuals with BPD report lower trait levels of trust using a self-report measure [Interpersonal Trust Scale (55)]. **(Right)** Individuals with BPD repay less than the healthy group, which indicate lower levels of trust within the exchange game (also see fig. S2).