RUNNING HEAD: Cooperation in Mood Disorders

Increased Cooperative Behavior across Remitted Bipolar I Disorder and Major Depression: Insights Utilizing a Behavioral Economic Trust Game

Desmond C. Ong[1], Jamil Zaki[1], & June Gruber[2*]

[1]Department of Psychology, Stanford University
[2]Department of Psychology and Neuroscience, University of Colorado at Boulder

IN PRESS at *Journal of Abnormal Psychology*

**Corresponding Author*
June Gruber, Ph.D.
University of Colorado Boulder
Department of Psychology & Neuroscience
1905 Colorado Avenue
345 UCB, Muenzinger D321C
Boulder, CO 80309
Email: june.gruber@colorado.edu

Author Note: D.C.O., J. Z. and J. G. designed the research, analyzed the data, and wrote the paper. J.G. oversaw data collection. We thank John Purcell and Benjamin Billingsley for assistance in data collection. The authors declare no conflict of interest.

Abstract

Mood disorders impact social functioning, but might contribute to experiences—like affective distress—that might result in increased cooperative behavior under certain circumstances. We recruited participants with a history of bipolar I disorder (n = 28), major depressive disorder (n = 30), and healthy controls (n = 27) -- to play a well-validated behavioral economic Trust Game, a task that provides a well-controlled experimental scenario to measure cooperative behavior for the first time across both groups. Both mood-disordered groups cooperated significantly more than the control group, but did not differ from one another. These results suggest that, in some contexts, a history of mood disturbance can produce enhanced cooperation, even in the absence of current mood symptoms. We discuss the clinical significance of enhanced cooperation in mood disorders and point to key directions for future research.

Keywords: Cooperation, Trust, Bipolar Disorder, Major Depressive Disorder

General Scientific Summary

Individuals with mood disorders suffer disruptions in social functioning, but under the right circumstances, might also display *more* prosocial behaviors. We find that individuals in remission from bipolar I disorder and major depressive disorder cooperated more in an economic game than individuals with no such history. The results suggest that in some contexts, a history of clinical mood disturbance generates increases in some forms of prosocial behavior, even in the absence of current mood symptoms.

Increased Cooperative Behavior across Remitted Bipolar I Disorder and Major Depression: Insights Utilizing a Behavioral Economic Trust Game

Mood disorders like bipolar I disorder (BD) and major depressive disorder (MDD) have a debilitating impact on psychosocial functioning (e.g., MacQueen, Young, & Joffe, 2001). Less understood are the mechanisms by which mood disturbance translates to psychosocial dysfunction. Recently, researchers have begun utilizing rigorous behavioral economic approaches to unpack observed social impairments by first examining simple behavioral markers in disorders, through carefully controlled experimental scenarios with clearly defined incentives (e.g., King-Casas & Chiu, 2012). Cooperation—acting in ways that benefits others—is one such behavioral marker crucial to maintaining relationships and coordination in groups (e.g., Rand & Nowak, 2013; Zaki & Mitchell, 2013). Hence, abnormalities in cooperative behavior could be one mechanism that mediates the impact of mood disorders on broader social dysfunction, and it is critical to study whether and how such abnormalities are expressed in mood disorders. Here we use a game theoretic paradigm to provide an "intermediate" level of analysis between disorder and social impairment, by connecting BD and MDD to abnormalities in specific social behaviors that contribute to broader social functioning. Specifically, we assess cooperation for the first time across both individuals with BD and MDD currently in remission, as compared to healthy non-psychiatric controls, using a well-validated index of cooperative behavior.

Interestingly, past research suggests that individuals with a history of mood disturbances may exhibit *increased* cooperative behavior when compared with healthy adults. This idea is consistent with classic work in psychology demonstrating that

distress—a feature of BD and MDD (Shamay-Tsoory, Harari, Szepsenwol, & Levkovitz, 2009; O'Connor, Berry, Weiss, & Gilbert, 2002)—is a strong motivation of prosocial behavior (Batson, 2011; Cialdini et al., 1997). Indeed, two studies using a behavioral economic game (the Ultimatum Game) found that individuals diagnosed with major depression cooperated more than healthy controls, by offering their partner a larger proportion of a pot of money (DeStoop et al., 2012), and accepting more unfair divisions from their partner (Harlé, Allen, & Sanfey, 2010). A study of individuals in remission from MDD similarly found that they cooperated more in a Prisoner's Dilemma game than healthy controls (Pulcu et al., 2015, Table 4: Welch's $t(37)=2.16$; $p=.038$; see Supplemental Materials). It is worth noting other studies that find no differences in cooperation between healthy controls and currently diagnosed MDD individuals (Pulcu et al, 2015; Unoka et al., 2009), or BD individuals (Saunders, Goodwin, & Rogers, 2015) on other economic tasks.

The present investigation translated a well-validated behavioral economic game paradigm in healthy adults to provide insights into social processes impacted by a history of psychiatric mood disturbance. This work provides broader significance in several key ways. First, as described above, economic games provide tight quantification of social behaviors in psychiatric samples. This literature is small but growing, and more data is required across more contexts and symptomatic states, especially in the remitted mood phase. Second, previous studies have never previously measured both BD and MDD participants together in the same task. Here, we do so, allowing us to examine diagnostically specific, versus transdiagnostic, effects of mood disorders on specific social behaviors like cooperation.

Methods

Participants

Participants were recruited as part of a broader study on emotion and mood (Yale University IRB HIC #1309012679; University of Colorado Boulder IRB #14-0390) using posted flyers, online advertisements, and referrals from outpatient mental health centers and psychiatric hospitals. Of the final 85 study participants, 28 were diagnosed with BD type I in remission (i.e., not currently manic, depressed, or mixed; rBD), 30 with MDD in remission (i.e., not currently depressed; rMDD), and 27 were non-psychiatric controls (CTL) who did not meet current or past criteria for any DSM-IV-TR Axis I disorder. We chose to examine both BD and MDD groups in a remitted mood phase to first examine potential group differences in prosocial behavior impacted even without mood symptoms (e.g., Gruber, Harvey, & Purcell, 2011). rBD and rMDD participants were not excluded on the basis of comorbid disorders (aside from current substance or alcohol use disorders) given that mood disorders are commonly comorbid with other Axis I disorders (e.g., Kessler, Chiu, Demler, & Walters, 2005), though we verified that BD was the primary, or most severe, diagnosis for the rBD group, and MDD for the rMDD group (Di Nardo, O'Brien, Barlow, Waddell, & Blanchard, 1983). Exclusion criteria for all groups included history of severe head trauma, stroke, neurological disease, severe medical illness (e.g., autoimmune disorder, cardiovascular disease, HIV/AIDS), or current alcohol or substance abuse in the past six months.

Measures of Clinical Functioning

Diagnostic evaluation. A trained post-baccalaureate researcher confirmed all diagnoses using the Structured Clinical Interview for DSM-IV (SCID-IV; First et al., 2007). A second reviewer from the research team independently rated a subset (n=75; 90%) of videotaped interviews from the broader study protocol. Ratings across all Axis I disorders matched 100% (κ=1.00) of primary diagnoses. During the SCID-IV, the interviewer collected information concerning illness duration and lifetime number of mood episodes.

Mood symptoms. We measured current symptoms of mania using the Young Mania Rating Scale (YMRS; Young et al., 1978), an 11-item, clinician-rated measure with scores ranging from 0 to 60, and current symptoms of depression using the Inventory of Depressive Symptomatology (IDS-C; Trivedi et al., 2004), a 30-item, clinician-rated measure with scores from 0 to 84. We verified current remitted mood status (i.e., neither manic, depressed, nor mixed mood state) for all groups according to both current SCID-IV criteria and cutoff scores on the YMRS (≤7), and IDS-C (≤11). Intra-class correlations (ICCs) for the same subset of participants were strong for both the YMRS (=0.90) and IDS-C (=0.99).

Global functioning. We used the Global Assessment of Functioning (GAF; *DSM-IV Axis V)* Scale to assess general functioning in the past week on a scale from 1 (lowest level of functioning) to 100 (highest level). The ICC for the same subset of participants (n=75; 90%) was acceptable (=0.65).

Cognitive Functioning. We assessed cognitive functioning using the Mini-Mental Status Examination, a brief objective measure of cognitive status and impairment (MMSE; Folstein, Folstein, & McHugh, 1975). Raw scores (range: 0-30) were calculated

as the total number of trials correct. All participants exceeded the eligibility cutoff score (≥24). We also assessed executive functioning using the letter-number sequencing subtest of the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997). Participants are read aloud a series of increasingly long strings of randomly ordered numbers and letters, and have to verbally repeat back all numbers in ascending numerical order, followed by all letters in alphabetical order. Raw scores (range: 0-21) were calculated as the total number of trials correct, from which WAIS-III age-normed scaled scores were computed.

Behavioral Economic Cooperation Game

To objectively measure cooperative behavior, we used a standardized version of the Trust Game (TG; **Figure 1)**. The TG is widely used in behavioral economics as an index of both trusting and trustworthy behaviors, and as a proxy of cooperative tendencies (Berg, Dickhaut, & McCabe, 1995). The task involves two players: the Investor and the Trustee. On each round, the Investor receives an endowment of $10, and decides to invests $X. This investment is then tripled and given to the Trustee, who decides how much ($Y) of $3X to return to the Investor. If the Investor chooses to invest $X and the Trustee chooses to return $Y, the Investor receives $(10 – X + Y), while the Trustee receives $(3X–Y). This procedure is repeated for multiple rounds.

In the present study, participants played the role of the Trustee, and believed the investor with whom they played was another participant. In fact, we designed a computer algorithm to simulate investor decisions. Previous studies using the TG and other games have similarly used computer programs to standardize conditions that participants were exposed to (e.g., McCabe et al., 2001; Sripada et al., 2009). In iterated paradigms with repeated rounds of interactions, humans tend to react to the outcomes of previous rounds,

and reinforcement-learning (RL) models serve as a good approximation to such behavior (Dayan & Niv, 2008). Intuitively, a human investor would increase their investment if the trustee returned a decent amount the previous round, and would decrease their investment if they felt cheated. The average human trustee, from a meta-analysis of over 20,000 participants, returns an average of 37% of the total sum they were entrusted with (Johnson & Mislin, 2011). Hence, we designed a simple RL investor that "expected" participants to behave like the average trustee and return 40% of the total amount that they were entrusted with, which corresponds to 120% of the investment. If a participant returned more/less than this expectation, the investor will increase/decrease its investment on the subsequent round. Additional details are included in the Supplemental Materials, and code can be found at https://github.com/desmond-ong/CooperationInMoodDisorders.

Procedure

Study design and procedures were reviewed and approved by the University of Colorado Boulder and Yale University Institutional Review Boards. Participants arrived at the laboratory, provided written and verbal informed consent and completed diagnostic and cognitive assessment measures. Participants then completed the TG task along with other unrelated laboratory tasks as part of the larger study protocol.

Participants participated in the TG task using a computer in an individual testing room. The experimenter remained in an adjacent room and via microphone. Participants learnt that they were about to play a two-player Investment Game, and saw two stick figures representing players A and B. On each round, A receives $10 and chooses an amount $X to invest, which is automatically tripled and given to B. B decides how much

of $3X to return to return to A. Participants saw an example round where A invests $7 out of $10 (which becomes $21), and B returns $12 out of $21: A ends the round with $10-$7+$12 = $15, and B ends the round with $21-$12 = $9. Participants could repeat the instructions until they understood. Next, participants were assigned to the role of Player B, and learnt that they would be playing 10 rounds with another participant in a separate room. Participants saw a screen that claimed to be connecting their computer to their "partner". In reality, they played with a computer program (described above). After 10 rounds, participants were debriefed, and compensated for their participation at a standard hourly rate. We did not include additional financial compensation as a function of the amount earned in the game.

Results

Demographic and Clinical Characteristics

We report demographic and clinical characteristics in **Table 1**. The groups did not differ with respect to age, gender, and ethnicity (p's>0.47). Although participants differed with respect to income ($\chi^2(10)=28.9$, $p=0.001$), income did not affect any of the behavioral results (p's>0.50), and our main results regarding group differences still hold after excluding those with more than $100,000 annual income. All groups also scored well below standardized cutoffs on the YMRS (≤7) and IDS-C (≤11). There was a marginally significant group main effect suggesting differences in YMRS scores ($p=0.073$). The rMDD group scored higher than the rBD ($p=0.047$) and CTL ($p<0.001$) groups on subsyndromal depressive symptoms (IDS-C), and rBD and CTL groups did not differ from each other ($p=0.10$). The rBD and rMDD groups scored lower on global

functioning (GAF) than the CTL group ($p<0.001$ for both), and the rBD group also scored lower than the rMDD group ($p<0.001$) on GAF. Participants did not differ on either MMSE scores ($p=0.22$), or WAIS-III scores ($p=0.70$). Additionally, the rBD and rMDD groups did not differ in illness duration ($p=0.45$), though the rMDD group had an earlier average age of onset ($p=0.002$).

Cooperative Behavior on the Trust Game

To compare returns across rounds with different investments, we calculated a *repayment rate*, the proportion of the investment the participant returns (i.e., $Y/$X). There were no significant trends of repayment rates over rounds ($\chi^2(3)=2.24$; $p=.52$). To account for the repeated nature of the data, we ran linear mixed-effects models with condition as fixed effects, and with random intercepts and random slopes (on rounds) by participant (**Table 2**). Across all rounds, rBD participants displayed a significantly greater repayment rate, repaying 30.3% more of the investment, than control participants ($b=0.303$, 95% confidence interval [0.059, 0.548]; $t(82)=2.43$; $p=0.017$). rMDD participants similarly repaid more than control participants, with an effect almost reaching significance ($b=0.241$ [0.0002, 0.481]; $t(82)=1.96$; $p=0.053$). There was no difference between the repayment rates of rBD and rMDD participants ($p=0.61$). Overall, clinical history explained a significant proportion of variance (marginal $R^2=.049$ for fixed effects variance only, and conditional $R^2=.677$ for fixed and random effects variances; after Johnson, 2014).

Next, we calculated several key measures that represent meaningful summaries of behavior: (1) the total amount invested in each participant after all ten rounds, (2) the total amount repaid by each participant, (3) the total amount kept by each participant and

(4) the total amount kept by the investor. Note that these quantities are all partly correlated with the participant's repayment rate, as the simulated investor followed a deterministic algorithm, up to random variations in learning rates. We regressed these summary measures on condition using simple linear models (**Table 2**). Recall that the investor can invest up to $10 per round for 10 rounds for a potential maximum of $100. On average, rBD participants encouraged the investor to invest $12.41 more as compared to control participants (b=12.41 [0.64, 24.19], $t(82)$=2.07; p=0.04), and repaid $34.72 more to the investor (b=34.72 [2.87, 66.57]; $t(82)$=2.14; p=0.04). The investor earned marginally more if they were playing with a rBD participant than a control participant (b=22.31 [0.24, 44.38]; $t(82)$=1.98; p=0.051); however, rBD participants did not earn more than control participants (p=0.76). Thus, the investor that played against the average rBD participant would have invested more, and "earned" slightly more after ten rounds, than if they had played against the average CTL participant. rMDD participants did not differ significantly from controls along any of these four measures (p's>0.15). In summary, not only did participants with a history of BD differ in their own behavior as compared to controls, they also encouraged more trusting behavior for "people" who play with them, resulting in better outcomes for their partners.

Discussion

Utilizing a behavioral economic trust game, we found that a history of mood disturbance is associated with *increased* cooperative behavior when individuals are trusted with an investment. Specifically, both rBD and rMDD groups in our study repaid more to a simulated investor in the Trust Game than a healthy control group. These

results are consistent with the idea that although mood disturbance, such as affective distress, generates hardship for individuals, they might, in some contexts, also lend themselves to other-oriented, cooperative behavior (Batson, 2011; Batson et al., 1983; Cialdini et al., 1997; O'Connor et al., 2002). Our findings are consistent with previous work showing increased cooperation in current and remitted MDD on other economic tasks (DeStoop et al., 2012; Harlé et al., 2010; Pulcu et al., 2015).

These data generate an interesting dilemma: if mood disorders are associated with increased cooperation in the lab, why do they also track strained social functioning outside the lab? One possibility is that economic games simplify and emphasize social responsibilities more than real-world interpersonal encounters, and that individuals with MDD and BD might respond to this clear social structure. In our study, another player trusted participants with an investment, and so our participants might feel more compelled to repay that trust by cooperating. This is consistent with another examination of compliance in cooperation in MDD. Using a modified Trust Game in which the investor added an explicit request for a specific amount of repayment, Zhang, Sun, and Lee (2012) found that currently depressed patients repaid the exact requested amount more frequently than healthy controls, even when they had the freedom to repay more or less than the requested amount. By contrast, real-world contexts are much more nuanced than economic games. People have to rely on complex cues about others' emotions when deciding how to direct their cooperative behavior (Zaki & Ochsner, 2012). To the extent that mood disorders hamper these individuals' ability to understand complex social signals (Devlin, Zaki, Ong, & Gruber 2016; Samamé, Martino, & Strejilevich, 2012), such individuals may be unable to express their cooperative motives.

More broadly, increased cooperative behavior observed in bipolar and major depressive disorders may have downstream maladaptive consequences. For example, individuals may cooperate with good intentions but produce negative consequences for themselves or those they intended to help (e.g., starving oneself to reduce one's burden on caregivers; Oakley, Knafo, Madhavan, & Wilson, 2011). Being excessively cooperative and/or trusting may also leave such individuals more vulnerable to exploitation by others; this could, in turn, generate life stressors that are known contributors to exacerbated depressive and manic symptom levels. Notwithstanding these clinical risks, it is possible that increased cooperative behavior might afford important opportunities. For example, therapists could work with patients to scaffold more structure in patients' social relationships to channel this cooperative motivation into adaptive group relevant settings (e.g., at work or with family). Much work remains to unpack the potentially deleterious aspects of excessive cooperation in mood disorders, as well as ways to capitalize upon its strengths.

The present investigation should be carefully interpreted within the confines of several caveats. First, although the Trust Game task is a well-validated task, it represents only one form of cooperation, and does not capture all the nuances of cooperation in daily life. Future work should obtain parallel assessments of real-world social functioning to provide direct links between experimental outcomes and social functioning in daily life. Second, it is worth noting that game theoretic tasks operationalize social phenomena as *behaviors*, irrespective of the underlying psychological mechanisms (Camerer, 2003). Cooperative behaviors can reflect many sources, including distress (Batson et al., 1983) but also reputation management (Harbaugh, 1998), reciprocity (Trivers, 1971), and

compliance with social norms (Nook et al., 2016). Future work should more closely examine these mechanisms and their effects on prosociality in mood disorders. Third, to obtain ecologically valid samples, we did not exclude participants from the BD or MDD clinical groups on the basis of comorbidities. Future studies with larger sample sizes should examine whether the effects of disorders on cooperation varies as a function of such comorbidities to rule out the possibility that the observed findings are not better attributed to comorbidity versus a history of mood disturbance.

Fourth, both BD and MDD participants were remitted at the time of testing. The use of a remitted sample enables the identification of areas of potential preservation during remission, but it also leaves open critical questions about how cooperative behavior covaries as a function of current mood symptom severity. Future work should carefully examine the relationship between current symptom status and cooperation. Fifth, the present sample sizes are respectable given the severe nature of the psychiatric groups recruited and complexity of the measured variables. However, there may still have been insufficient statistical power to assess within-group affective correlates of cooperative behaviors, between-group differences between BD and MDD groups, and other more complex interaction effects. Future studies (or meta-analyses) should also examine cooperative behavior in larger sample sizes and across psychopathology more broadly, such as in borderline personality disorder (King-Casas et al, 2008; Unoka et al, 2009), or anxiety (Rodebaugh et al, 2015; Sripada et al, 2009). Sixth, given the challenges of accessing an unmedicated clinical sample, we were unable to investigate the influence of medication status on cooperation. Future studies should explore larger sample sizes to assess both functional correlates of cooperation within and across

diagnoses and the effects of medication status. Finally, although our study was cross-sectional, it is important for future prospective studies to examine the potential protective effects of increased cooperative behavior on symptom remission and mood relapse.

In conclusion, although BD and MDD are characterized by disruptions in social behavior and relationships, here, we document increased cooperation in an economic game among individuals in remission from these disorders. We hope that this work, and future studies, will provide a deeper understanding of the social consequences of affective disorders, and inform treatment options that leverage cooperation to improve psychosocial functioning.

References

Batson, C.D. (2011). *Altruism in Humans*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195341065.001.0001

Batson, C.D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A.M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology, 45*(3), 706-718. http://dx.doi.org/10.1037/0022-3514.45.3.706

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122-142. http://dx.doi.org/10.1006/game.1995.1027

Camerer, C. (2003). *Behavioral Game Theory*. Princeton, NJ: Princeton University Press.

Cialdini, R.B., Brown, S.L., Lewis, B.P., Luce, C., & Neuberg, S.L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, *73*(3), 481-494. http://dx.doi.org/10.1037//0022-3514.73.3.481

Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current Opinion in Neurobiology*, *18*(2), 185-196. http://dx.doi.org/10.1016/j.conb.2008.08.003

DeStoop, M., Schrijvers, D., De Grave, C., Sabbe, B., & De Bruijn, E.R. (2012). Better to give than to take? Interactive social decision-making in severe major depressive disorder. *Journal of Affective Disorders*, *137*(1), 98-105. http://dx.doi.org/10.1016/j.jad.2011.12.010

Devlin, H.C., Zaki, J., Ong, D.C., & Gruber, J. (2016). Tracking the emotional highs, but

missing the lows: Hypomania risk is associated with positively biased empathic accuracy inference. *Cognitive Therapy and Research, 40(1), 72-79*. http://dx.doi.org/10.1007/s10608-015-9720-6

Di Nardo, P.A., O'Brien, G.T., Barlow, D.H., Waddell, M.T., & Blanchard, E.B. (1983). Reliability of DSM-III anxiety disorder categories using a new structured interview. *Archives of General Psychiatry*, *40*(10), 1070-1074. http://dx.doi.org/10.1001/archpsyc.1983.01790090032005

First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J. (2007). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-VP).* New York, NY: Biometrics Research, New York State Psychiatric Institute.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189-198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Gruber, J., Harvey, A.G., & Purcell, A. (2011). What goes up can come down? A preliminary investigation of emotion reactivity and emotion recovery in bipolar disorder. *Journal of Affective Disorders*, *133*(3), 457-466. http://dx.doi.org/10.1016/j.jad.2011.05.009

Harbaugh, W.T. (1998). The prestige motive for making charitable transfers. *The American Economic Review, 88*(2), 277-282

Harlé, K.M., Allen, J.J., & Sanfey, A.G. (2010). The impact of depression on social economic decision making. *Journal of Abnormal Psychology*, *119*(2), 440-446.

http://dx.doi.org/10.1037/a0018612

Johnson, P.C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944-946. http://dx.doi.org/10.1111/2041-210X.12225

Johnson, N.D., & Mislin, A.A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865-889. http://dx.doi.org/10.1016/j.joep.2011.05.007

Kessler, R.C., Chiu, W.T., Demler, O., & Walters, E.E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617-627. http://dx.doi.org/10.1001/archpsyc.62.6.617

King-Casas, B., & Chiu, P.H. (2012). Understanding interpersonal function in psychiatric illness through multiplayer economic games. *Biological Psychiatry*, *72*(2), 119-125. http://dx.doi.org/10.1016/j.biopsych.2012.03.033

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, *321*(5890), 806-810. http://dx.doi.org/10.1126/science.1156902

MacQueen, G.M., Young, L.T., & Joffe, R.T. (2001). A review of psychosocial outcome in patients with bipolar disorder. *Acta Psychiatrica Scandinavica*, *103*(3), 163-170. http://dx.doi.org/10.1034/j.1600-0447.2001.00059.x

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of*

the National Academy of Sciences, *98*(20), 11832-11835. http://dx.doi.org/10.1073/pnas.211415698

Nook, E.C., Ong, D.C., Morelli, S.A., Mitchell, J.P., & Zaki, J. (2016). Prosocial conformity: Generalization across behavior and affect. *Personality and Social Psychology Bulletin, 42*(8), 1045-1062. http://dx.doi.org/10.1177/0146167216649932

O'Connor, L.E., Berry, J.W., Weiss, J., & Gilbert, P. (2002). Guilt, fear, submission, and empathy in depression. *Journal of Affective Disorders*, *71*(1), 19-27. http://dx.doi.org/10.1016/s0165-0327(01)00408-6

Oakley, B., Knafo, A., Madhavan, G., & Wilson, D.S. (Eds.). (2011). *Pathological Altruism*. Oxford University Press.

Pulcu, E., Thomas, E.J., Trotter, P.D., McFarquhar, M., Juhasz, G., Sahakian, B.J., Deakin, J.F.W., Anderson, I.M., Zahn, R., & Elliott, R. (2015). Social-economical decision making in current and remitted major depression. *Psychological Medicine*, 45(06), 1301-1313. http://dx.doi.org/10.1017/s0033291714002414

Rand, D.G., & Nowak, M.A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413-425. http://dx.doi.org/10.1016/j.tics.2013.06.003

Rodebaugh, T.L., Heimberg, R.G., Taylor, K.P., & Lenze, E.J. (2015). Clarifying the Behavioral Economics of Social Anxiety Disorder Effects of Interpersonal Problems and Symptom Severity on Generosity. *Clinical Psychological Science*, 4(1), 107-121 http://dx.doi.org/10.1177/2167702615578128

Samamé, C., Martino, D.J., & Strejilevich, S.A. (2012). Social cognition in euthymic bipolar disorder: Systematic review and meta-analytic approach. *Acta*

Psychiatrica Scandinavica, *125*(4), 266-280. http://dx.doi.org/10.1111/j.1600-0447.2011.01808.x

Saunders, K.E.A., Goodwin, G.M., & Rogers, R.D. (2015). Borderline personality disorder, but not euthymic bipolar disorder, is associated with a failure to sustain reciprocal cooperative behaviour: implications for spectrum models of mood disorders. *Psychological Medicine*, *45*(08), 1591-1600. http://dx.doi.org/10.1017/S0033291714002475

Shamay-Tsoory, S., Harari, H., Szepsenwol, O., & Levkovitz, Y. (2009). Neuropsychological evidence of impaired cognitive empathy in euthymic bipolar disorder. *Journal of Neuropsychiatry and Clinical Neurosciences, 21*(1), 59-67. http://dx.doi.org/10.1176/appi.neuropsych.21.1.59

Sripada, C.S., Angstadt, M., Banks, S., Nathan, P.J., Liberzon, I., & Phan, K.L. (2009). Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *Neuroreport*, *20*(11), 984-989. http://dx.doi.org/10.1097/wnr.0b013e32832d0a67

Trivedi, M.H., Rush, A.J., Ibrahim, H.M., Carmody, T.J., Biggs, M.M., Suppes, T., ... & Kashner, T.M. (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: A psychometric evaluation. *Psychological Medicine*, *34*(01), 73-82. http://dx.doi.org/10.1017/s0033291703001107

Unoka, Z., Seres, I., Áspán, N., Bódi, N., & Kéri, S. (2009). Trust game reveals restricted

interpersonal transactions in patients with borderline personality disorder. *Journal of Personality Disorders*, *23*(4), 399-409. http://dx.doi.org/10.1521/pedi.2009.23.4.399

Wechsler, D. (1997) *WAIS-III administration and scoring manual*, The Psychological Corporation, San Antonio, TX.

Young, R.C., Biggs, J.T., Ziegler, V.E., & Meyer, D.A. (1978). A rating scale for mania: reliability, validity and sensitivity. *The British Journal of Psychiatry*, *133*(5), 429-435. http://dx.doi.org/10.1192/bjp.133.5.429

Zaki, J., & Mitchell, J.P. (2013). Intuitive prosociality. *Current Directions in Psychological Science*, *22*(6), 466-470. http://dx.doi.org/10.1177/0963721413492764

Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience*, *15*(5), 675-680. http://dx.doi.org/10.1038/nn.3085

Zhang, H.J., Sun, D., & Lee, T. (2012). Impaired social decision making in patients with major depressive disorder. *Brain and Behavior*, *2*(4), 415-423. http://dx.doi.org/10.1002/brb3.62

Table 1
Demographic and Clinical Characteristics

	rBD (*n*=28)	rMDD (*n*=30)	CTL (*n*=27)	Statistic
Demographic				
Age (Yrs)	29.9 (6.8)	27.9 (6.4)	28.4 (6.5)	*F*= 0.765
Female (%)	64%	73%	63%	χ^2= 0.837
Caucasian (%)	79%	77%	74%	χ^2= 0.156
Education (Yrs)	14.23 (2.02)	15.95 (1.79)	15.96 (2.20)	*F*=6.93**
Employed (%)	53.6%	70.0%	70.4%	χ^2=2.26
Partnered (%)	42.9%	63.3%	63.0%	χ^2=1.81
Number Children	1.5 (0.7)	1.1 (0.4)	1.3 (0.7)	*F*=2.57
Annual Income				χ^2=28.9**
<$10K	17.9%	13.3%	25.9%	
$10K-$25K	46.4%	30.0%	3.7%	
$26K-$50K	28.6%	23.3%	33.3%	
$51K-$75K	0%	6.7%	11.1%	
$76K-$100K	3.6%	20.0%	0 %	
>$100K	3.6%	6.7%	25.9%	
Cognitive				
MMSE	29.0 (1.7)	28.4 (1.8)	29.1 (1.4)	*F*=1.53
WAIS-III	10.8 (3.1)	11.1 (2.9)	10.4 (2.8)	*F*=0.35
Clinical				
YMRS	1.32 (1.49)	1.07 (1.20)	0.56 (0.97)	*F*= 2.70 [a]
IDS-C	3.43 (2.33)	4.77 (3.04)	2.30 (2.05)	*F*= 6.83 [bc]
GAF	70.46 (7.22)	76.97 (7.11)	85.85 (6.41)	*F*= 34.08 [abc]
Remission Duration (Mos.)	36.95 (36.03)	31.40 (33.36)	--	--
Age at Onset (Yrs)	18.10 (5.34)	14.50 (3.04)	--	*F*= 10.2 [b]
Illness Duration (Yrs)	12.16 (5.85)	13.3 (5.17)	--	*F*= 0.50
# Comorbid Disorders	0.14 (0.36)	0.43 (0.63)	--	*F*= 4.63 [b]
% Comorbid Disorders	14%	36.7%	--	χ^2=4.44
# Depressive Episodes	13.48 (17.53)	17.23 (24.10)	--	*F*= 0.45
# Manic Episodes	13.63 (21.53)	--	--	--
# Medications	1.46 (1.26)	0.60 (0.77)	--	F= 10.1 [b]
anticonvulsants	39.3%	3.33%	--	
lithium	10.7%	0%	--	
neuroleptics	35.7%	0%	--	
stimulants	0%	3.33%	--	
antidepressants	35.7%	43.3%	--	
benzodiazepines	10.7%	3.33%	--	
sedative-hypnotics and other anxiolytics	14.3%	6.67%	--	

Note: rBD=remitted Bipolar Disorder group; rMDD=remitted Major Depressive Disorder group; CTL=Healthy control group; Employed=Employed full-time or part-time; Partnered=Married or in a relationship; YMRS=Young Mania Rating Scale; IDS-C=Inventory of Depression Symptomatology, Clinician Rating; GAF=Global Assessment of Functioning; Age at Onset=Age of first depressive or manic episode; # Comorbid Disorders=the number of current DSM-IV-TR Axis I comorbidities; % Comorbid Disorders=at least one DSM-IV-TR Axis I comorbidity; # Medications=the number of psychotropic medications currently taken (including anticonvulsants, lithium, neuroleptics, stimulants, antidepressants, benzodiazepines, and sedative-hypnotics and other anxiolytics); Mean values are given with standard deviations in parentheses.

[a] $p<0.05$ for rBD and CTL

[b] $p<0.05$ for rBD and rMDD

[c] $p<0.05$ for rMDD and rCTL

*$p < .05$, **$p < .01$, ***$p < .001$

Table 2. *Multiple Regression Analyses of Behavior in Trust Game*

	Repayment Rate (Amount repaid / Amount invested)				
Model (1)	Coefficient	Standard Error	95% CI	*t*	*p*
rBD (> CTL)	0.303	0.125	[0.059, 0.548]	2.43	0.017 *
rMDD (> CTL)	0.241	0.123	[0.0002, 0.481]	1.96	0.053
rBD (> rMDD)	*0.063*	*0.122*	*[-0.176, 0.301]*	*0.52*	*0.608*
Marginal R^2	0.049				
Conditional R^2	0.677				

Summary Measures	Total Amount Invested in Participant (2)	Total Amount Repaid by Participant (3)	Total Amount Kept by Participant (4)	Total Amount Kept by Investor (5)
rBD (> CTL)	12.41 (6.01) *	34.7 (16.3) *	2.52 (8.28)	22.3 (11.3)
rMDD (> CTL)	5.94 (5.91)	22.2 (16.0)	-4.37 (8.14)	16.3 (11.1)
R^2	0.050	0.054	0.009	0.049

Model (1) is a linear mixed-effects model predicting repayment rate with clinical history as fixed effects, and random intercepts and slopes (on round number) by participant. The rBD>rMDD contrast was calculated by re-running the model with rMDD as the base group. Marginal and Conditional R^2 are reported, after Johnson (2014). Models (2-5) predict summary measures calculated after the end of all rounds; they are fixed effects only linear models, as there are no repeated measurements. The units on the unstandardized coefficients for Models (2-5) are in terms of hypothetical dollars, with standard error in parentheses. Note: these summary measures are partly correlated with the repayment rate, but they also provide an estimate of the behavioral effect size across the whole task.

*$p < .05$, **$p < .01$, ***$p < .001$

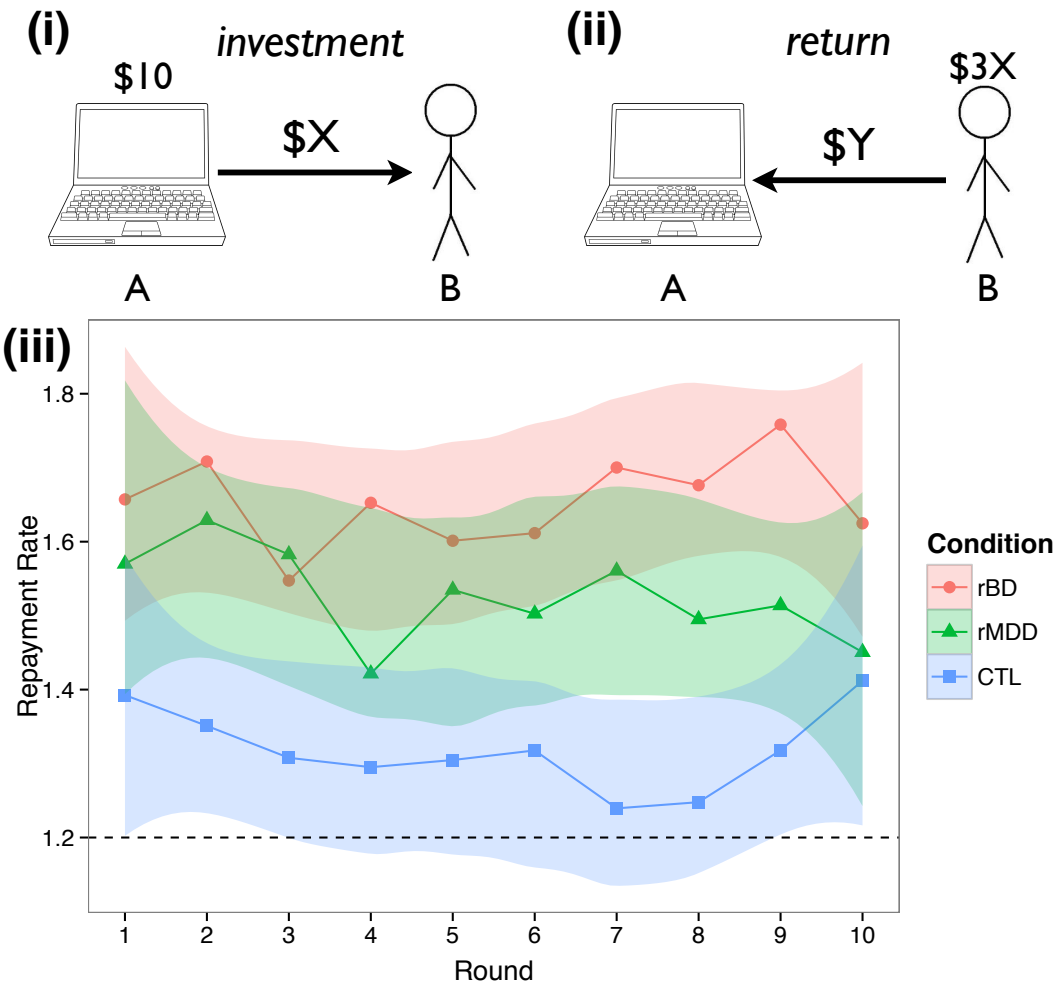

Figure 1: (i) Illustration of the investment portion of a round, where the computer chooses to invest $X. **(ii)** Return portion: participant chooses to return $Y out of $3X. **(iii)** Mean repayment rates ($Y/$X) over rounds, colored by condition. Shaded regions indicate LOWESS-smoothed 95% confidence intervals. The dashed line at 1.2 indicates the expected repayment rate (details in text).

Supplemental Materials

All code and data can be found at our public Github repository at: https://github.com/desmond-ong/CooperationInMoodDisorders

Note Regarding Pulcu et al (2015) in the Introduction

In the Introduction in the main text, we described a calculation based on data from Pulcu et al (2015). Pulcu et al (2015) studied a sample of currently diagnosed MDD, remitted MDD, and healthy controls, and used a variety of tasks. They summarized their outcome measures in their Table 4, and they offer one-way ANOVA statistics looking at differences across the three groups. In particular, we were interested in the reported results of their Prisoner's Dilemma (PD) game. In Table 4, they reported the means and SDs of the number of defections in the PD, where lower numbers indicate more cooperative behavior ($M_{Control}$ ($SD_{Control}$) = 2.0 (4.1), $n_{Control}$ = 33; M_{rMDD} (SD_{rMDD}) = 0.4 (1.2), n_{rMDD} = 35; M_{cMDD} (SD_{cMDD}) = 3.5 (6.1), n_{cMDD} = 24). They additionally report and interpret the results of a one-way ANOVA: "Patients with cMDD defected significantly more in the PD ($F_{2,73}$ = 3.113, p = 0.05) than the other groups." (Pulcu et al, 2015, pp. 1306). This is the only mention of the analyses done on the PD data, although the PD result is one of the paper's main results. First, note that the inference (that cMDD defected more than other groups) cannot be made just from the one-way ANOVA alone. That inference requires post-hoc t-tests, and thus, their statement is technically incorrect. Second, if we perform the post-hoc t-tests using the provided means, standard deviations, and sample sizes, we find that remitted MDDs cooperated significantly *more* than controls (Welch's $t(37)=2.16$; $p=0.038$), and there is no difference between the behavior

of current MDDs and controls ($t(38)=1.05$; $p=0.30$). The calculations to reproduce these tests can be found in our analysis code. Thus, the quoted statement in Pulcu et al (2015) is misleading, and the actual results are contrary to the main message of their paper (although there are other results in the paper that we did not re-analyze). Upon our reinterpretation, their PD results actually suggest that remitted MDD cooperate *more* than healthy controls in the PD, and there are no differences between current MDD and healthy controls.

Additional Details of Trust Game Paradigm

As mentioned in the text, we used a standardized version of the Trust Game (TG; Berg, Dickhaut, & McCabe, 1995), and we designed a computer program to simulate the decisions of the investor. Previous studies using the TG and other economic games have also used computer programs in order to standardize conditions that participants were exposed to (e.g., McCabe, Houser, Ryan, Smith, & Trouard, 2001; Sripada et al., 2009). We implemented the TG task using Psychtoolbox in Matlab.

In this section, we describe the motivation behind the computer program we designed. In iterated paradigms with repeated rounds of interactions, human players change their behavior on subsequent rounds based upon the outcomes of previous rounds. Intuitively, a human investor would increase his investment if the trustee returned a decent amount the previous round, and would likely decrease his investment if he felt he was cheated. The average human trustee, from a meta-analysis of over 20,000 participants, returns an average of 37% of the total sum they initially receive (Johnson & Mislin, 2011).

Previous work has found that reinforcement-learning (RL) models serve as a good approximation to human behavior in such paradigms (e.g., Dayan & Niv, 2008). Hence, we designed a simple reinforcement-learning (RL) agent that expected trustees to return 40% of the trustees' total amount. In the current study, investments are tripled, so 40% expectation of the total amount corresponds to 1.2 times the amount the program invested. To reiterate, our RL program "expects" participants to return the same amount as the average trustee. Any deviations from this expectation would constitute a *prediction error*: If a trustee in this study returned more than the expected amount (i.e., 1.2x what the algorithm invested), this constitutes a *positive prediction error*, and would lead our algorithm to increase investment on subsequent rounds. By contrast, if the trustee returned less than this expectation, this constitutes a negative prediction error, and will decrease our algorithm's subsequent investments. The amount the RL agent changes its behavior on the subsequent round depends on the prediction error and its learning rate, α: If the program invested \$X and the trustee returned \$Y, where $Y > 1.2X$, then the program would *increase* its investment by $\alpha(Y-1.2X)$ on the following round. Like humans, our program exhibited *loss aversion*, where a negative prediction error weighs more than a positive prediction error, by doubling α on negative prediction errors. If the participant returned $Y < 1.2X$, the program's investment on the following round would *decrease* by $2\alpha(1.2X-Y)$.

We made four additional design choices. First, on the first round, the program always started by investing half of their \$10 investment. Second, we used an α that pseudo-randomly decreased over time, to approximate the human tendency of converging to equilibrium behavior. Third, the program's investments were capped between \$1 and

$10. We chose to set the floor at $1 because if the program invested $0, the participant would have no other option except to return $0; this would continue until the end of the ten rounds, with the participant being unable to respond and influence the behavior of the program. Fourth, on round 6, the program would additionally increase its investment by half of the difference between the maximum ($10) and the amount previously invested. If the program invested $3 on round 5, it would increase its investment by $3.50 (half of $10 - $3) on round 6, in addition to its reaction to the prediction error from round 5 (which could be positive, or negative). We added this positive "boost" to encourage cooperation from the participant in cases of poor cooperation. In other words, the program gives a "second chance" by boosting its trust of the participant at the midway point in cases where trustees initially fail to cooperate, as was previously documented in Borderline Personality Disorder (e.g., King-Casas et al., 2008).