A Model of Risk and Mental State Shifts during Social Interaction

Andreas Hula [1], Iris Vilares[2], Peter Dayan[3], and P. Read Montague[4]

[1]Wellcome Trust Centre for Neuroimaging, London, United Kingdom
[2]Wellcome Trust Centre for Neuroimaging, London, United Kingdom
[3]Gatsby Computational Unit, University College London, London, United Kingdom - Wellcome Trust Centre for Neuroimaging, London, United Kingdom
[4]Wellcome Trust Centre for Neuroimaging , London, United Kingdom - Human Neuroimaging Laboratory, Virginia Tech Carilion Research Institute, Roanoke, VA, United States - Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

April 13, 2017

Abstract

Cooperation and competition between human players in repeated microeconomic games offer a powerful window onto social phenomena such as the establishment, breakdown and repair of trust. This offers the prospect of particular insight into populations of subjects suffering from socially-debilitating conditions such as borderline personality disorder. However, although a suitable foundation for the quantitative analysis of such games exists, namely the Interactive Partially Observable Markov Decision Process (I-POMDP), computational considerations have hitherto limited its application. Here, we improve inference in I-POMDPs in a canonical trust game, and thereby highlight and address two previously unmodelled phenomena: a form of social risk-aversion exhibited by the player who is in control of the interaction in the game, and irritation or anger exhibited by both players. Irritation arises when partners apparently defect, and it causes a precipitate breakdown in cooperation. Failing to model one's partner's propensity for it leads to substantial economic inefficiency. We illustrate these behaviours using evidence drawn from the play of large cohorts of healthy volunteers and patients.

1 Introduction

Assessing the characteristics or the internal state of mind of another person is a fundamental requirement for success in human social decision making. Neither people's self-reports, nor any current measurement device provides complete, veridical, information about another person's state. Nevertheless, we are typically quite adept at inferring the preferences and intentions of others and even at manipulating their states, in both cases over the course of multi-round interactions. One way to formalize this capacity is via the so-called interactive Partially Observable Markov Decision Process (I-POMDP; [Gmytrasiewicz and Doshi, 2005]). This is a regular Markov Decision Process (see [Puterman, 2005]) augmented with (a) partial observability (see [Kaelbling et al., 1995]) about the characteristics of a partner; and (b) a notion of cognitive hierarchy (see [Costa-Gomes et al., 2001, Camerer et al., 2004]), associated with the game theoretic interaction between players who model each other.

In recent work, we used approximate inference methods in the I-POMDP to capture the effect of an other-regarding utility preference (namely guilt) in modeling behaviour in a popular multi-round trust task (MRT) [King-Casas et al., 2008, Chiu et al., 2008, Debajyoti et al., 2008, Xiang et al., 2012, Hula et al., 2015].

This model offered powerful accounts of both the behavior of subjects, and also aspects of their neural activity [Debajyoti et al., 2008, Xiang et al., 2012]. However, a detailed inspection of the residual errors revealed two key characteristics that were missing from the model: social risk aversion and irritation. Here, we formalize both, and, by improving the computational characteristics of approximate I-POMDP inference to accommodate them, fit subjects' choices much more closely.

First, investors are dominant in the MRT, in that they can still make substantial sums of money based on initial endowments in each round without investing anything. Perhaps as a result of this, we observed that some investors apparently treat a portion of their endowment as being exclusively theirs; only risking the remainder in the social exchange. This is a form of social preference that is absent in the Fehr-Schmidt model of other-regarding preferences that we adopted as our baseline model [Fehr and Schmidt, 1999]. Here, we treat it explicitly as a form of (social) risk aversion.

A second failure of the existing model is that sample investment profiles are generally too homogeneous. That is, as pointed out in some of the early neuroeconomics studies of the MRT [King-Casas et al., 2008], cooperation between the players can readily break down in the face of apparent defection; with coaxing then being necessary to reestablish it (especially on the part of trustees). Such phenomena appear particularly prevalent in play involving subjects suffering from psychiatric conditions such as borderline personality disorder (BPD) (see for instance [King-Casas et al., 2008]). This condition is frequently characterised by difficulties in maintaining social relationships, sudden ruptures in trust, and social withdrawal or aggression (see [Lieb et al., 2004, Fonagy and Bateman, 2006]).

We therefore augmented the model with a form of irritation. When irritated, subjects can exhibit substantially different rules of behaviour, for instance being unwilling to cooperate at all, and reducing their depth of interpersonal reasoning. This leads to breakdowns in cooperation. To predict what might happen in response to their own choices, and thus, if beneficial to them, to prevent a breakdown, subjects need to model the possibility of such a shift in their partner's state. They can then change their behavior prospectively.

We adapted an existing I-POMDP characterization of the multi-round trust task [Hula et al., 2015] to encompass both social risk aversion and transient irritation. We generated simulated data using this model to show how the inclusion of these dimensions of social manipulation affects the course and understanding of human social exchange. We then demonstrated how the new mechanisms allow us to account for behaviour that appeared anomalous according to our previous model. We also generate confusion matrices to confirm that we can recover the parameters governing these (and other) facets of behaviour in the MRT to a reasonable degree.

1.1 Trust Task

The multi round trust task (see [King-Casas et al., 2005, Xiang et al., 2012, Hula et al., 2015], based on [McCabe et al., 2001]) (see figure 1) is a paradigmatic social exchange game. It involves two people, one playing the role of an "investor" the other that of a "trustee", over 10 sequential rounds. Quantities pertaining to the investor and trustee are denoted by superscripts "I" and "T" respectively. The participants played at the same time but did not know or meet each other at any point.

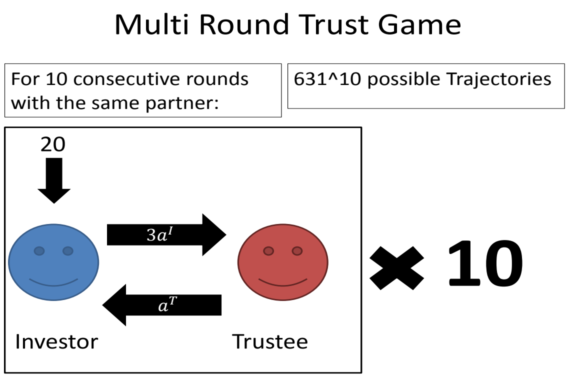

Figure 1: Physical features of the multi round trust game.

Both players know all the rules of the game. In each round, the investor receives an initial endowment of 20 monetary units. The investor can send any a^I units of this amount to the trustee. The experimenter triples this quantity and then the trustee decides how much (an amount a^T) to send back to the investor. This amount must be between 0 points and the whole amount that she receives. The repayment by the trustee is not increased by the experimenter. After the trustee's action, the investor is informed, and the next round starts. On each round the financial payoffs of the two actors can be calculated: for the investor this is:

$$\chi^I(a^I, a^T) = 20 - a^I + a^T \tag{1}$$

and for the trustee:

$$\chi^T(a^I, a^T) = 3a^I - a^T. \tag{2}$$

For computational simplicity, the model treated the possible choices on a coarser grid, allowing for five investor actions and five corresponding trustee reactions. The five investor actions correspond to investing 0, 5, 10, 15 or 20 or $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ of their endowment, while the trustee responses correspond to the return of 0, $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$ or $\frac{2}{3}$ of the received amount. The case in which the investor gives 0 is special, since the trustee has no choice but to return 0. We round real subject actions to the respective nearest grid point.

The Nash equilibrium (based on pure monetary outcomes) for this game mandates a trivial interaction. In the last round, the investor should never invest anything, since the trustee could defect without punishment. Thus the interaction progressively unravels. Real subject behaviour in the game is quite different, and typically leads to substantial investments and returns.

1.2 Generative Model

A generative model of the multi round trust task was introduced in [Hula et al., 2015]; we enrich it here. Those of the parameters that we also assume subjects to infer about each other over the course of interaction are called "intentional"; the other parameters are inferred by the experimenter through the process of fitting the choices (using maximum likelihood), but are merely assumed by the subjects and are constant throughout the experiment. Full details of the model can be found in the supplemental material here and in [Hula et al., 2015].

In the original model, there was a single intentional parameter, namely guilt $\alpha \in \{0, 0.4, 1.0\}$. This denotes the sensitivity of one player to inequality in their favour, as in the Fehr-Schmidt model of inequity aversion [Fehr and Schmidt, 1999]. Subjects were assumed to use Bayesian inference to infer their partners' guilt over the course of the interaction. This is possible since a high guilt ($\alpha = 1$) partner will provide high investments or returns and appear persistently cooperative, while a low guilt ($\alpha = 0$) partner is likely just to maximise their own winnings (and so only cooperate for Machiavellian reasons).

Next, a player could be aware that their partner was also learning about them, a recursive concept formalized as computational theory of mind (ToM) or reasoning level k, and depicted in figure 2. A level 0 investor learns about the trustee, but treats her as being random rather than intentional. A level 1 investor treats the trustee as being level 0, implying that the trustee is assumed to learn about a non-intentional investor. A level 2 investor treats the trustee as being 1, implying that the trustee is assumed to know that the investor is learning about them too. This continues recursively. One consequence of the interplay of I-POMDP modeling and the asymmetric nature of the game is that only even levels yield new insight into investor behaviour, and only odd levels into that of the trustee [Hula et al., 2015]. In the original model, computational considerations restricted the theory of mind to $k^I \in \{0, 2\}$ for the investor and $k^T \in \{0, 1\}$ for the trustee. In the MRT, levels of ToM higher than 4 of ToM do not appear to yield qualitatively new behavioural patterns (see supplementary information), and so we extended consideration to levels $\{0, 2, 4\}$ for investors and $\{0, 1, 3\}$ for trustees.

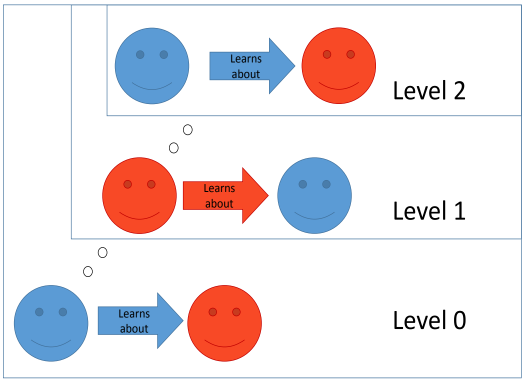

Figure 2: Recursive reasoning about a partner. At level 0 the blue player learns about the partner. At level 1 the blue players knows that the red player learns about them too (that is, that the red player is level 0). At level 2 the blue player knows that the red player knows they are learning about them (i.e. that the red player is level 1 and thinks of the blue player as level 0). This recurses up to higher levels.

Finally, subjects were classified according to their planning capacity P, which quantifies how many steps of the future of the interaction they take into account when assessing the consequences of their actions. In the original model, this could take the values $P \in \{0, 2, 7\}$. However, it turns out that play for $P = 4$ has very similar features to that of $P = 7$, involving exploitation of the partner and inhomogeneous effects caused by the horizon of the game (see supplementary material). Therefore, to liberate the computational capacity to model an additional intentional parameter, we restricted P to $\{1, 2, 3, 4\}$.

Choices were assumed to be made through a stochastic policy π on the basis of expected value action values (calculated in the manner of [Bellman, 1952]) through the medium of a logistic softmax (see [McKelvey and Palfrey, 1992, Gläscher et al., 2010, Huys et al., 2011]). The inverse temperature parameter of the softmax that was fixed at $\beta = \frac{1}{3}$ in the original model, was here fit using values $\beta \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\}$. Note the relatively large numerical values of investment and return, which is why the inverse temperatures may seem relatively small compared with other studies.

To gauge the differences between models with different numbers of parameters, we used the Bayesian Information criterion (BIC), which penalizes the number of parameters n used to fit each subject according to the number m of data points obtained in each exchange. It is defined using the negative loglikelihood (NLL) $-\log \mathbb{P}[x_s|\theta_s^*; M]$ at the best fitting parameters θ_s^* for each subject s under the given model M.

$$\text{BIC}(M) = \sum_{\text{subjects } s} \left(-2\log \mathbb{P}[x_s|\theta_s^*; M] + n(\log(m) - \log(2\pi)\right) \quad (3)$$

In the multi round trustgame $m = 10$, due to the 10 choices per subject. The correction factor is for small m [Draper, 1995].

The parameters of the final model can be seen in table 1.

Table of Parameters

Parameter	Range	Concept
Guilt α	$\{0, 0.4, 1\}$	Measure of tendency to try and reach a fair outcome.
Plan P	$\{1, 2, 3, 4\}$	Number of steps likely planned ahead.
Theory of Mind k	$\{0, 2, 4\}$ or $\{0, 1, 3\}$	number mentalisation steps.
Inverse Temperature β	$\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}\}$	Certainty of own choice preference.
Risk Aversion (Belief) ω ($b(\omega)$)	$\{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8\}$	Value of money kept over (potential) money gained.
Irritability ζ	$\{0, 0.25, 0.5, 0.75, 1.0\}$	Tendency to retaliate on worse than expected partner actions.
Irritation Belief $q(\zeta)$	$\{0, 1, 2, 3, 4\}$	Initial belief on likelihood of the partner being irritable.

Table 1: All Parameters in the full model.

2 Materials and Methods

2.1 Ethics Statement

Informed consent was obtained for all research involving human participants, and all clinical investigation was conducted according to the principles expressed in the Declaration of Helsinki. The procedures were approved by the Institutional Board of Baylor College of Medicine.

2.2 Subject Data

We use the data set shown in [King-Casas et al., 2008], consisting of 93 healthy investors, paired with 93 trustees, of which 55 were BPD diagnosed trustees (BPD Group, "BPD") and 38 were healthy trustees, matched in age, gender, IQ and socio-economic status (SES) with the BPD trustee group (healthy control group, "HC"). The precise demographics can be found in [King-Casas et al., 2008].

2.3 Technical Data

Programs were run at the local Wellcome Trust Center for Neuroimaging (WTCN) cluster using Intel Xeon E312xx (Sandy Bridge) processor cores clocked at 2.2 GHz; no process used more than 1.5 GB of RAM. We used R [R Core Team, 2013] and Matlab [MATLAB, 2010] for data analysis and the boost C++ libraries [Boost-Libraries, 2014] for code generation.

2.4 Algorithmic Change

The approach in [Hula et al., 2015] utilized a sampling based method to explore the decision tree during planning in the trust game, drawing from approximate solution methods for tree search from machine learning (see [Auer et al., 2002a, Auer et al., 2002b] , [Kocsis and Szepesvári, 2006], [Silver and Veness, 2010]). However, if lower levels of calculation are kept in memory and so are immediately available for higher level calculations, then the problem scales linearly in the theory of mind level parameter, rather than exponentially (as for other computational approaches to I-POMDPs; [Gmytrasiewicz and Doshi, 2009], p. 325, 9.2.). This trade off of memory for computation is only practical if the planning horizon is reduced to at most 4 steps into the future.

The various critical scalings are shown in figure 3: Figure 3A;B show the linear growth of computation time and memory respectively with (maximal) theory of mind level (including all lower-level calculations and tree storage). Conversely, figure 3C shows the exponential rise in time for calculating a level $k^I = 4$ investor, level $k^T = 3$ trustee interaction at different planning horizons. Similarly, the exponential growth of memory use in the planning horizon for a level $k^I = 4$ investor, level $k^T = 3$ trustee can be seen in figure 3D.

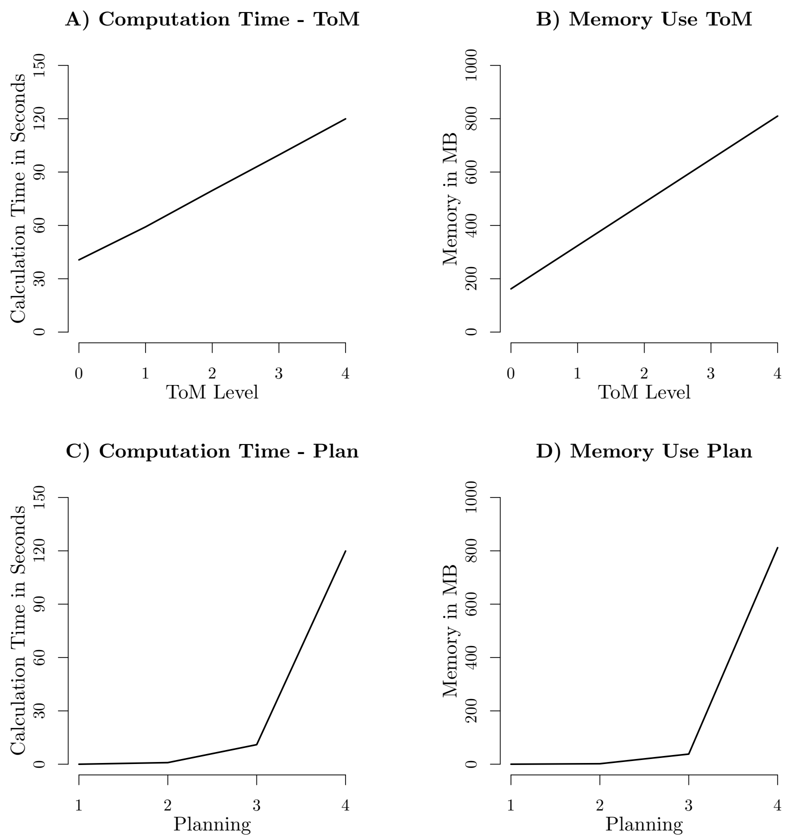

Figure 3: Scaling of computation and memory. A;B) Linear scaling of calculation time (A) and memory (B) with the maximum ToM level used, at planning horizon 4. C;D) Exponential scaling of the computation time (C) and memory (D) for a 10 step simulated interaction at ToM 4 with the planning horizon.

The net result is that it takes less than 2 minutes per generated 10 step interaction, to calculate deterministically (i.e., avoiding approximations from the stochasticity of Monte Carlo-based tree evaluation) a 10 step exchange of a level $k^I = 4$ investor with a level $k^T = 3$ trustee, both having horizons of $P = 4$ steps. This comes at the cost of having to commit 0.8 Gb of RAM to the tree calculation.

3 Results

3.1 Model Failure

Figure 4A shows the average investments and returns in the data from [King-Casas et al., 2008]. The dark blue and dark red lines in figure 4A show the respective average investments and returns for healthy investors playing BPD trustees. The lighter blue and red lines show average investments and returns for healthy investors and healthy trustees who matched the BPD trustees in socio-economic status (SES), IQ, age and gender. Investments averaged about half the initial endowment and evolved over trials. In the second half of the game, investors paired with BPDs invested considerably less than investors paired with healthy trustees. This effect was a central topic in [King-Casas et al., 2008], and was explained by BPD trustees not heeding warning signals from their investor partners indicated investor dissatisfaction with the BPD patients' lack of reciprocation.

The solid bars in figure 4B show the average total investments in the real data for the two groups. These do not differ significantly. The hatched bars show the result of generating data from the model in [Hula et al., 2015] (using the extensions discussed above to higher theory of mind and lower maximal planning). Model data is generated for each dyad, using that dyad's best fitting parameters. The model overestimates the investments of the BPD-paired investors to the tune of about 40%

Figure 4C demonstrates a similar issue for the modelled trustees. The solid bars show the returns of the control and BPD trustees; these again do not differ significantly. However, the simulated HC trustees return significantly less. Although it may seem that the simulated BPD trustees return similar proportions to the actual BPD trustees, this actually flatters the model, since this repayment is based on the over-generous model investment (the hashed bars in part B) rather than the true, more miserly, investment.

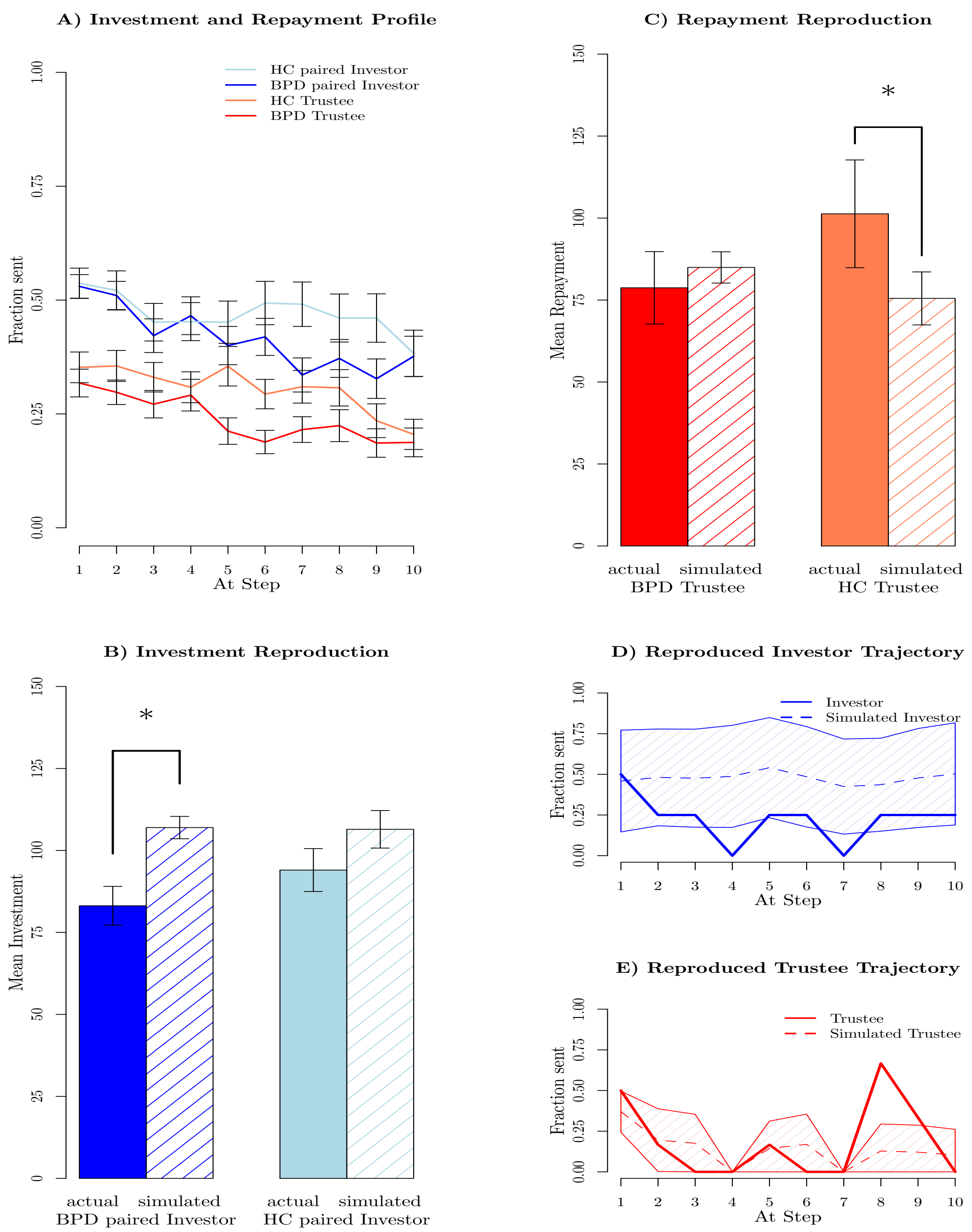

Figure 4: A) Averaged investments and repayments in the data set. Errorbars show standard errors of the mean. B) Average investment in real and in simulated exchanges based on best fit parameters. An asterisk denotes a significant difference ($p < 0.05$, two sided t-test) in means between the original data and the generated exchanges. C) Average repayments in real and in simulated exchanges based on the actual parameters. An asterisk denotes a significant difference ($p < 0.05$, two sided t-test) in means between the original data and the generated exchanges. D) Sample trajectory for an investor vs average of 200 generated exchanges with best fitting parameters, based on the model in [Hula et al., 2015]. Shaded area shows estimated standard deviations. E) Sample trajectory for a trustee vs average of 200 generated exchanges with best fitting parameters, based on the model in [Hula et al., 2015]. Shaded area shows estimated standard deviations.

A second model failure concerns the detailed dynamics of investment across the task. The solid lines in figure 4DE show a selected sample interaction between a healthy investor (figure 4D) and a BPD trustee (figure 4E). The trustee provides a poor return in trial 3, and is met by zero investment in trial 4. The same pattern repeats in trials 6 and 7. The trustee is then far more generous in trial 8; this then coaxes (to adopt a term from [King-Casas et al., 2008]) the investor to continue investing, though after 2 breaks, the investors is unwilling to much increase their investment above a low level. The trustee then defects on trial 10, returning nothing.

[King-Casas et al., 2008]'s conclusion was that a significant portion of the BPD group lacked mechanisms that could consistently repair the faltering interactions that occur when subjects become what we will describe as being irritated. Thus tentative ruptures (in the form of drops in investment level) turned into complete breaks, with the investor using their position of power in the game to punish the trustee.

The dashed lines in figure 4D;E show the result of simulating 200 trajectories using parameters fit to the actual data, and also making predictions at each step based on the actual investments and returns of the dyad prior to each step (explaining why the model return is also 0 on trials 4 and 7). The shaded areas show the empirical standard deviations. The specific reductions are not only absent in the averages; the modelled investment following the trustee's defection on trials 3 and 6 decreased to 0 on only 11% and 13.5% of the sample runs; compared with the collapse to 0 apparent in the actual data.

We addressed these two sources of model failure by introducing two new parameters, associated with risk aversion and irritation.

3.2 Risk Aversion

The investor is in charge in the MRT, since she could simply keep her endowment on each round. It has been noted since the advent of this kind of trust game in [McCabe et al., 2001] that a lack of investment could represent a social form of risk aversion rather than a lack of trust; see [Houser et al., 2009]. This could account for differences in levels of investment regardless of the cooperativity of either partner.

We parameterize such risk aversion as a multiplicative factor ω^I in the payoff functions, increasing or decreasing the evaluation of money that the investor keeps for herself compared to the money returned by the trustee:

$$\chi^I_\omega(a^I, a^T) = \omega^I(20 - a^I) + a^T, \tag{4}$$

with $\omega^I \in [0.4, 1.8]$ (in 7 steps of 0.2). The trustee is subordinate in the task, and so does not have a risk parameter of their own. Instead, the trustee makes an assumption about the investor's degree of risk aversion, at one of the above mentioned 8 values. We capture intentional aspects of trust through guilt, and so treat risk aversion as a non-intentional parameter. However, in keeping with [Harsanyi, 1967], both players are assumed to be consistent, with the investor believing the trustee to know her risk aversion, and to know that she believes this; and the trustee believing that the investor believes this too. We write $b^T(\omega^I)$ for the trustee's belief about the investor's value of ω^I.

Depending on the trustee's belief $b^T(\omega^I)$, there will be either earlier or later attempts at exploitation. If $b^T(\omega^I) < 1$, then the trustee infers the investor will keep investing, even if the trustee has been relatively uncooperative (i.e. the investor will be risk-seeking). Conversely, if $b^T(\omega^I) > 1$, then the trustee will infer that any investment is contingent on their behavior, and there could be negative consequences of poor return. For values $b^T(\omega^I) > 1.4$, the trustee expects the investor to invest so little that building up trust will not be worthwhile in the first place. In this case, the interaction will rupture.

The effect of risk aversion in shifting investment levels can be seen in figures 5A;B. These depict the average investment (A) and repayment (B) trajectories over 200 simulated exchanges in which a trustee with $b^T(\omega^I) = 1$ (i.e., who believes the investor to have $\omega^I = 1$) interacts with investors of varying actual ω^I values (other parameters are given in the caption). Cooperative trustee actions early in the game can make the investor overcome moderate levels of risk aversion (the curve for $\omega^I = 1.2$ merges with the curve for $\omega^I = 0.8$ in the early trials). Higher risk aversion levels ($\omega^I = 1.4$) delay the positive effects of cooperation, such that it increases from a low initial level until step 6, but then drops abruptly due to horizon effects and trustee defection. For the highest risk aversion levels ($\omega^I = 1.6; 1.8$) in figure 5A, inference about other parameters may be hampered. That is, if investments stay low throughout, risk aversion might become nearly the only parameter that can be inferred with certainty. This implies a constraint on any further statistical treatment of behavioural data or derived quantities, such as model-based fMRI analysis.

Figure 5C depicts the effect of the risk aversion belief on the trustee, with the investor now being fixed at $\omega^I = 1.0$ (and not shown). For $b^T(\omega^I) < 1.0$, it can be seen that trustees make early attempts to exploit investors they think are not risk averse, since they assume the investor is still more likely to invest in them than to defect. However, they then defect rather quickly. As the trustee's belief approaches the utilitarian setting $b^T(\omega^I) = 1.0$, she becomes more cooperative and for a longer time period into the game, since this is necessary to keep the investments of the utilitarian investor going. Given a high setting of $b^T(\omega^I) = 1.4$ the trustee returns little, as they consider the likelihood of the investor to keep investing in them to be low and thus see no gain from building up cooperation.

Including risk aversion allows the model to account for the behavioural data much more proficiently, with the average Investor NLL improving from 12.94 to 9.68. The average trustee NLL improves from 11.36 to 9.5. The average BIC for the investors improves from 27.3 to 21.68, and for the trustees, from 23.98 to 21.3.

As figures 5D;E demonstrate, the investment can on average be well reproduced after including risk aversion in the model, although there remains a significant under-reciprocation on the part of the generated HC trustee, compared with the real exchange data. Further, figure 5FG demonstrates that despite an improvement in fit (NLL for the investor goes from 14.04 to 7.47 with $\omega^I = 1.4$; for trustee the NLL goes from 11.93 to 11.53 with $b^T(\omega^I) = 0.6$), this model remains incapable of capturing the transient rupture and - by extension - repair that we examined in figure 4D. Again, the modelled investor decreased to 0 on only 23% of the sample runs on trials 4 and 7, compared with the collapse in the actual investment.

3.3 Irritation

We explained the breakdown in cooperation evident in figure 4DE as arising when the participants become irritated. Formalizing this leads to four considerations: (i) what do subjects do differently when irritated; (ii) what leads a subject to become irritated; (iii) how can irritation be repaired; (iv) and what do subjects know about their own irritability? We offer a highly simplified characterization of all four of these. Individual interactions in the 10 round MRT are too short to license more complex treatments.

Definition 1 (Irritability). *We define the irritated state as associated with planning $P = 0$, guilt $\alpha = 0$, temperature $\beta = \frac{1}{2}$ and complete disregard of beliefs about the other player that have hitherto been established. We model the players' policy π as being a mixture between irritated π_ι and the nonirritated $\pi_{\bar{\iota}}$ choices, with irritation weight v_ι*

$$\pi(a, h) = (1 - v_\iota)\pi_{\bar{\iota}}(a, h) + v_\iota \pi_\iota(a, h).$$

A participant's irritation weight is assumed to start at $v_\iota = 0$, and to increase when their partner's action (investment or return) falls short of the value expected on the basis of the current model they have of the partner (including the partner's potential irritation):

$$v_\iota = \min\{v_\iota + \zeta, 1.0\} \qquad \textit{given unfavorable investment or return} \tag{5}$$

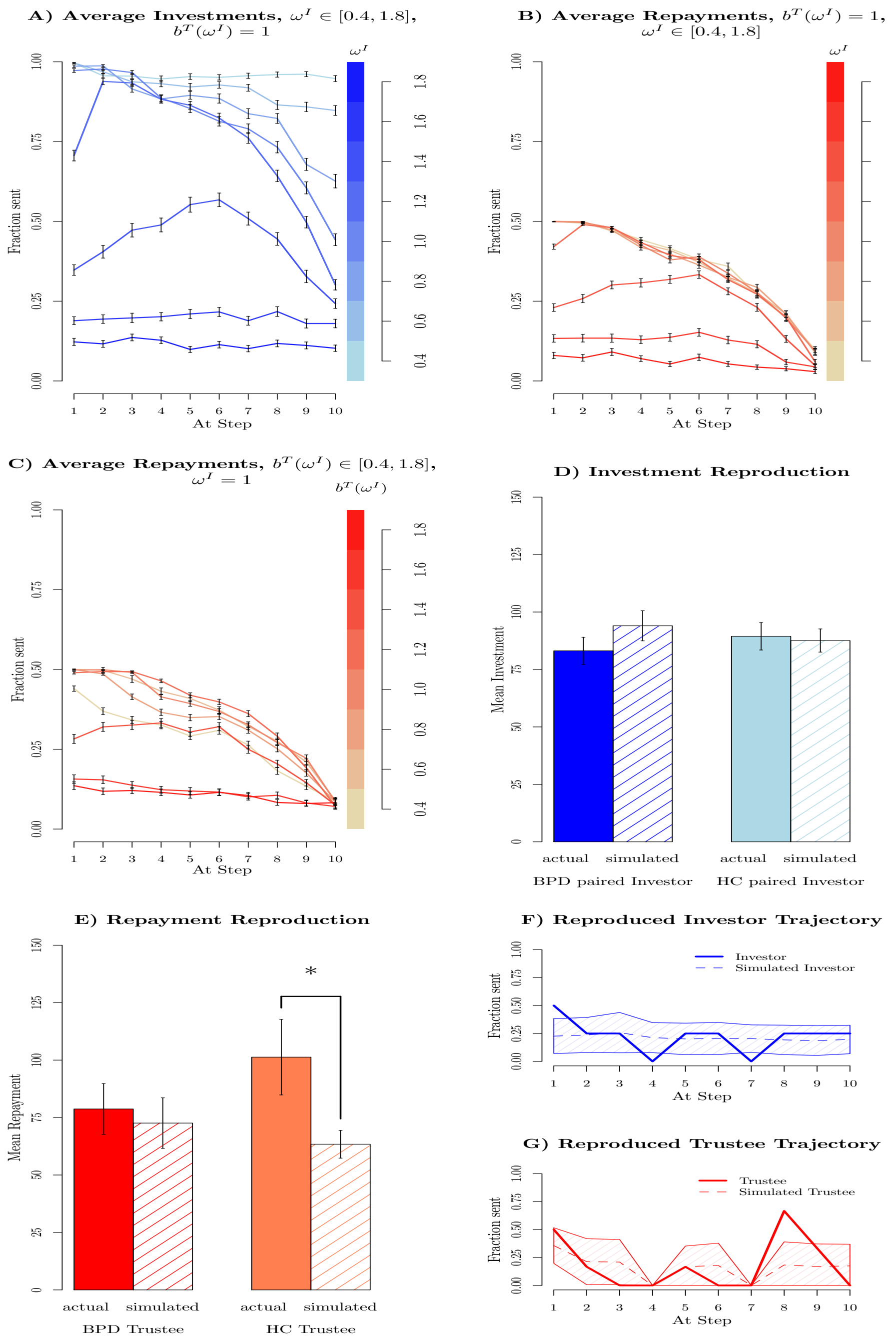

Figure 5: A) Investment profiles for different settings of investor risk aversion. All errorbars are standard errors of the mean. Here, the investor and trustee have guilt $\alpha = 0.4$, inverse temperature $\beta = \frac{1}{2}$, and planning $P = 4$; the investor (respectively trustee) has ToM level $k^I = 2$ ($k^T = 1$). B) Trustee repayment profiles for the interactions depicted in A. C) Repayment profiles for different settings of trustee risk aversion belief. All errorbars are standard errors of the mean. Here, the investor and trustee have guilt $\alpha = 0.4$, inverse temperature $\beta = \frac{1}{2}$, and planning $P = 4$; the investor (respectively trustee) has ToM level $k^I = 2$ ($k^T = 1$). D) Average Investment profiles regenerated from estimated parameters. All errorbars are standard error of the mean. E) Average Repayment profiles regenerated from estimated parameters. All errorbars are standard error of the mean. An asterisk denotes a significant difference ($p < 0.05$, two sided t-test) in means between the original data and the generated exchanges. F) Sample investor trajectory vs average of 200 generated exchanges using the model augmented by risk aversion. Shaded areas are estimated standard deviation. G) Sample trustee trajectory vs average of 200 generated exchanges using the model augmented by risk aversion. Shaded areas are estimated standard deviation.

where ζ is a subject-specific parameter. Irritation decreases through a process of repair when the action exceeds this expected value

$$v_\iota = \max\{v_\iota - \zeta, 0.0\} \qquad \textit{given favorable investment or return} \tag{6}$$

Definition 2 (Intentional Inference about Irritation)**.** *Players maintain and update beliefs about the partner's irritability in exactly the same way as about the partner's guilt: that is, they employ a Dirichlet prior on a multinomial distribution over five possible irritation values $\zeta \in \{0, 0.25, 0.5, 0.75, 1\}$ (dubbed respectively "nonirritable" and four different"irritable" types in the following) and use the same approximate inference rule as is used for guilt.*

However, unlike guilt, which we imagine is a characteristic that varies continuously amongst our participants, we consider a discrete set of possible prior beliefs about irritability. That is, irritability awareness is treated as an additional discrete new parameter ($q^I(\zeta^T); q^T(\zeta^I) \in \{0, 1, 2, 3, 4\}$). The investor's value $q^I(\zeta^T)$ determines prior weights of his belief over the trustee's actual irritability ζ^T. The trustee's value $q^T(\zeta^I)$ determines prior weights of her belief over the investor's actual irritability ζ^I. These priors are intended to cover a suitable range of possibilities; as noted, the MRT involves too few choices to license a richer depiction.

Irritability ignorant subjects ($q(\zeta) = 0$) have prior weights of $(400, 0.1, 0.1, 0.1, 0.1)$, implying they effectively never consider the possibility that their partner could be irritable. At $q(\zeta) = 1$. subjects consider the partner to be rather unlikely to be irritable , corresponding to prior settings of $(4, 0.5, 0.5, 0.5, 0.5)$. At the neutral awareness setting $q(\zeta) = 2$ they consider the partner to equally likely irritable, expressed by prior settings of $(0.4, 0.1, 0.1, 0.1, 0.1)$, meaning they consider a 50 percent chance initially of the partner being irritable. At setting $q(\zeta) = 3$ they consider the partner to be rather likely irritable, corresponding to $(2, 1, 1, 1, 1)$. At setting $q(\zeta) = 4$ subjects are "irritability certain", corresponding to priors of $(0.1, 0.1, 0.1, 0.1, 400)$ and hence will always treat the partner as irritable.

Finally, although we assume that players infer both their partner's inequality aversion and their partner's irritability level during the interaction, we do not allow subjects to consider their *own* future irritation. This follows [Loewenstein, 2005]'s observations of subjects' inability whilst engaging in 'cold' cognition to contemplate the possibility of 'hot' cognition.

A detailed example of the general workings of irritation in the case of a single trajectory with potentially aware participants ($q^I(\zeta^T) = q^T(\zeta^I) = 2$) is shown in figure 6A. At step 2, a subpar repayment by the trustee was introduced by fiat to show the workings of irritation (the expected repayment by the trustee would have been 50%). As a result, the investor's irritation rose to $v^I_\iota = 0.5$. At this point the trustee's belief about the investor's irritability is still at 0.5, as they have not observed the investor's response to their action. At step 3 the investor retaliated against the earlier defection of the trustee. The aware trustee thus updated their irritation beliefs, inferring that the investor was more likely to be irritable (at a marginal probability of $p = 0.58$). Noting the potential cost to the interaction of further irritating the investor, the trustee ensured a better than expected response in the next interaction at step 4. Not only did the trustee repair the interaction, they also ensured that they did not further irritate the investor, at least until the very end of the interaction, as can be seen in the remainder of figure 6A, from step 4. This exactly captures the "coaxing"-type repair mechanism that [King-Casas et al., 2008] suggested to explain differences in investment behaviours elicited by healthy control and BPD trustees.

Figure 6B shows the consequence of a lack of irritation inference in the presence of an irritable investor. The players had the same parameter values as in figure 6A, but were irritability ignorant ($q^I(\zeta^T) = q^T(\zeta^I) = 0$). After the same two initial actions (again introduced by fiat), without a notion of the partner being irritable, the trustee missed the chance to repair the interaction at step 3 and the investor's irritability weight rose to $v^I_\iota = 1$. From this point on the investments stayed low and the trustee did not placate the investor, thus receiving only a paltry total income. Both players failed to extract anything like the full return available from the experimenter.

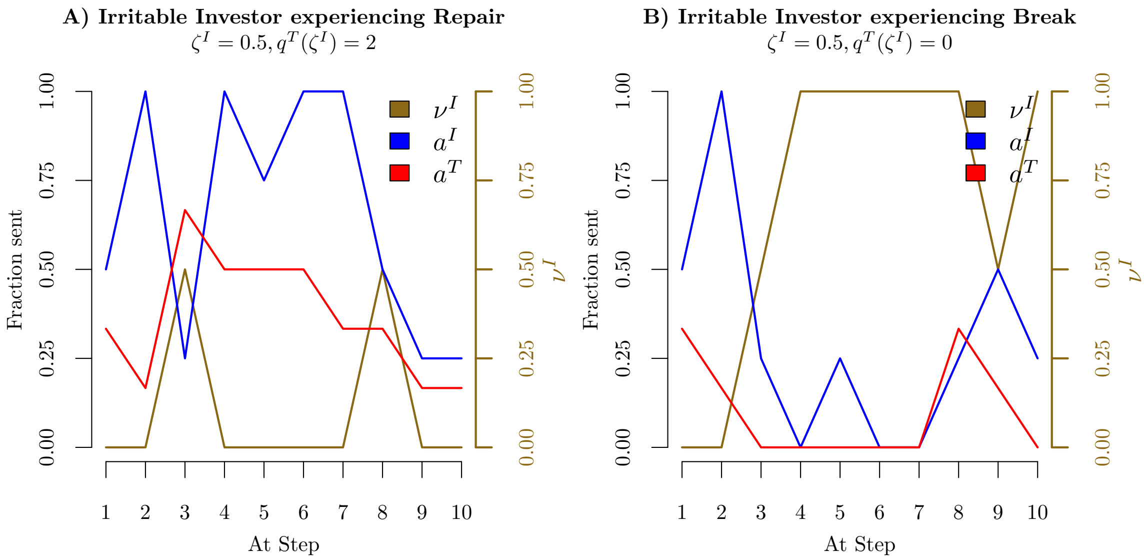

Figure 6: A) Simulated Repair Interaction. Single trajectory of two aware players (blue for investor, red for trustee). The golden line depicts the evolution of the investor irritation weight during the interaction. B) Simulated Break Interaction. Both players were irritability ignorant, thus they do not notice potential irritation. The gold line depicts the evolution of the investor irritation weight during the interaction. For A;B the simulated investor/trustee had $k = 2/1, \zeta = 0.5/0, \alpha = 0.4, P = 4, \beta = \frac{1}{3}$.

Figure 7 shows the effect of irritability and irritability inference on average behaviours over 200 simulated exchanges. As is also evident in figure 4, these averages blur the precise times at which the ruptures happen, but show the consequences in terms of net cooperation. In both cases (A-C; D-F), the trustee is more sophisticated than the investor ($k^I = 0; k^T = 1$) , the investor is either irritable or non-irritable, but unaware; the trustee is nonirritable (other parameters are listed in the caption). The difference between the figures is that the trustee is aware in figure 7A-C, but unaware in figure 7D-F.

Figure 7A show average investment and returns when the trustee is aware for an irritable (dark) and non-irritable (light) investor. The trustee's awareness enables him to keep the investment at almost the same level in both cases. This arises from the excess return of the trustee past step 4. Thus, as in figure 6A, trustees that realize that the investor is irritable will delay exploitation to the late rounds of the game. For the case of the irritable investor, figure 7B shows the average evolution of the inference about partner irritability. The trustee becomes aware that their partner is likely irritable after the first retaliation (as in the particular case in figure 6A). The average internal state of the same investor and trustee can be seen in figure 7C. Overall the irritation weight of the investor is kept low by an aware trustee, who can repair the interaction if needed. The value of irritability between investor and trustee is not symmetric in this example, so that the trustee may reliably repair or fail to repair without being themselves subjected to the effects of irritation.

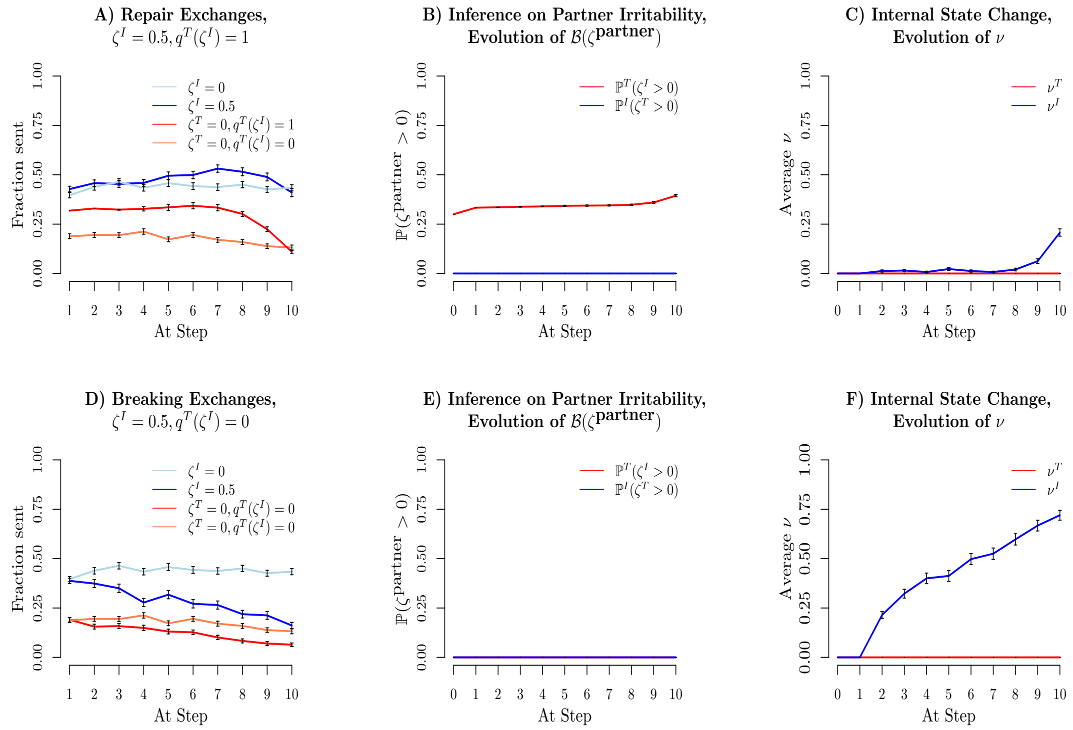

Figure 7: A-C) Simulated repair interaction with an irritability aware trustee ($q^T(\zeta^I) = 1$) and unaware investors who are irritable (dark lines; $\zeta^I = 0.5$) or non-irritable (light lines; $\zeta^I = 0$). A) Average investment profiles in the two cases. B) Average evolution of the irritability beliefs of both partners. The trustee learns correctly that the irritable investor is irritable. C) Average evolution of the irritation weight v_ι for both partners. The awareness keeps the irritation weight low and the trustee can repair the interaction if needed. D-F) Simulated breaking interaction with the same investors, but an unaware trustee ($q^T(\zeta^I) = 0$). The plots show the same quantities as in A-C. All errorbars are standard errors of the mean over 200 simulations. Investor/trustee parameters are $k = 0/1; \alpha = 0.4; \omega = 1.4; P = 4; \beta = \frac{1}{2}$ and the trustee had the fixed belief that the investor's risk aversion was $\omega = 1.4$.

By contrast, figure 7D shows that if the trustee is unaware, then there can be ruptures of cooperation that become apparent even at the group average level. The evolution of the beliefs about irritation is nugatory, as can be seen in figure 7E. Figure 7F shows how the irritation weight reaches high values quickly, only occasionally being reduced by chance repair. We note that the investor is driven up to near equally high investments in the non-irritable case in figure 7A compared to that in the irritable case of figure 7A, despite the trustee actually returning less. This is because the level $k^T = 1, q^T(\zeta^I) = 0$ trustee knows exactly what actions they need to take in order to confuse the inference of the level $k^I = 0$ investor (i.e. what responses the investor will consider unlikely). By contrast, the level $k^T = 1, q^T(\zeta^I) = 1$ trustee in figure 7A accounts for potential irritability right away and thus has to "play along" with the investor $k^I = 0$'s expectations and is less effective in tricking their inference.

Figures 8A;B show an absence of the discrepancies we saw before between data generated from the full model and the subject data. There is no longer a significant difference between generated and original investments or repayments. The complete model predicts 44% of the investor choices (chance is 20%) or equivalently an average NLL of 8.41 on 10 investor choices (form 9.68) and an average NLL of 7.5 or 47% of choice predicted for trustee choices (from 9.5). The final average BIC for the investors is 20.07 and for the trustees is 18.2. Figure 8CD demonstrates that the model qualitatively captures ruptures and repair occurring in real interactions, with the investment decreasing to 0 on 50% of the sample runs on trials 4 and 7. The investor NLL of this interaction improves from 7.47 to 5.42, while the trustee improves from NLL 11.53 to an NLL of 9.5 (with $\zeta^I = 0.5; \zeta^T = 1$).

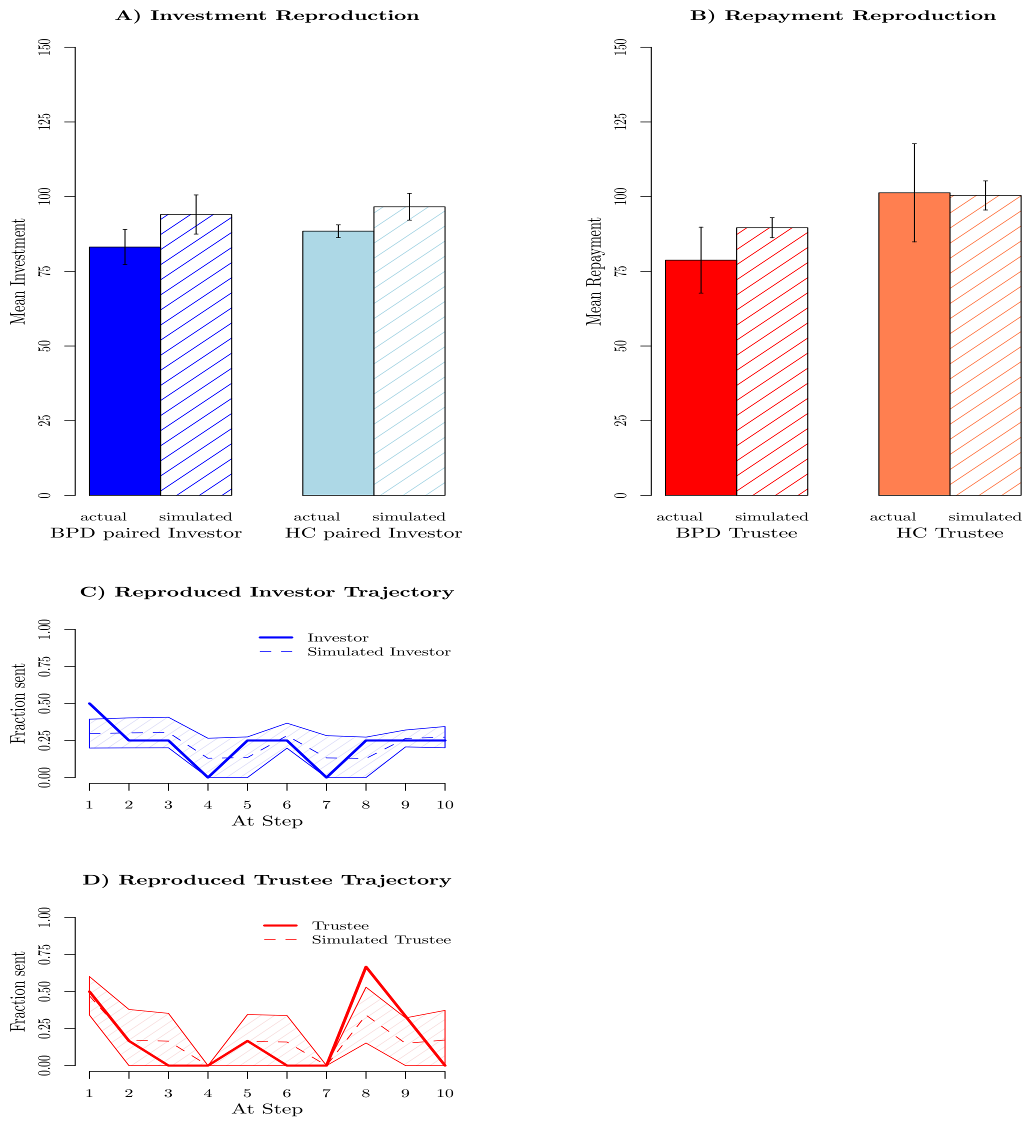

Figure 8: A) Average Investment profiles regenerated from estimated parameters in the full model. All errorbars are standard error of the mean. B) Average Repayment profiles regenerated from estimated parameters in the full model. All errorbars are standard error of the mean. C) Reproduction of sample investor trajectory using 200 simulated interactions with the best fitting parameters. Shaded areas are estimated standard deviation. D) Reproduction of sample trustee trajectory using 200 simulated interactions with the best fitting parameters. Shaded areas are estimated standard deviation.

3.4 Parameter Recoverability

The ultimate model is rather complicated. This raises the concern that the same behaviour might result from radically different settings of the parameters, implying that we would not be able to draw stable or meaningful conclusions from fitting behaviour. Indeed, we have already observed that certain settings will make it impossible to make inferences about some parameters – thus, playing with a highly risk averse investor will give no opportunity for a trustee to express her individual characteristics.

To examine this, we assessed parameter recoverability. That is, we used the parameters obtained from ML estimation on the participants (note that not all values were represented in the population – $\omega^I = 0.6$ is absent, for instance), generated new data *ab initio* from the model, fitted the new data, and quantified any discrepancies between the original and recovered parameters. Figures 9 and 10 show the probability of recovering either the actual or a neighboring parameter value for investor and trustee respectively.

It is apparent that the model has some significant purchase on all the parameters. However, some parameters are much harder to estimate than others. There are perhaps four most egregious forms of confusion. First, irritable subjects can be inferred as being non-irritable (figures 9F; 10F). This occurs if the remaining randomness of the interaction in the model is such that the investor's irritation is not excited. Indeed, the task was not designed with irritation in mind, and so players are not forced or encouraged to irritate each other.

Second, the investor's awareness is not very reliably recovered (figure 9E). It is slightly better recovered for the trustee (figure 10E), who faces a more stringent challenge to keep the investor trusting them.

Thirdly, the inverse temperatures β^I of generated investor trajectories tend to be overestimated (figure 9B). This is not surprising since the preferred actions will be the same for several temperature settings, thus requiring several "unlikely" actions to identify a lower β^I in investors.

Finally, there is a tendency to missestimate the risk aversion belief of the trustee, for high settings of $b^T(\omega^I)$, as the parameter apparently has less influence on trustee choice, than it does for the investor, who is in control of the interaction (see figure 10G). This effect may be driven by the fact that the high risk aversion estimated trustees in this sample, were also estimated to be considerably more irritable and thus the ensuing breaks might have confounded the risk aversion belief inference.

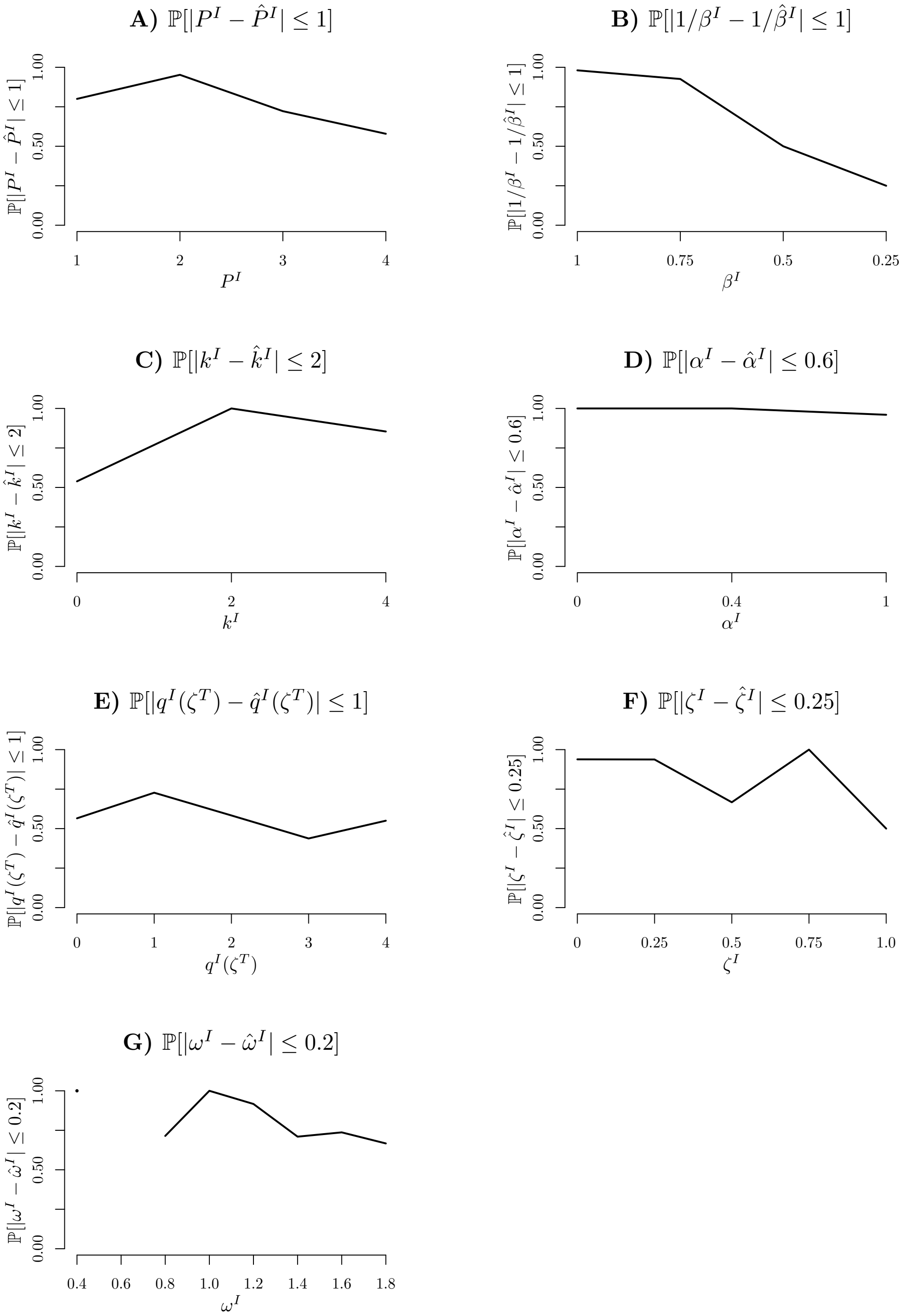

Figure 9: Probability of Parameter recovery or recovery of a neighbouring parameter value from generated exchanges using real subject parameters. A) Probability to recover the planning value P^I or a neighbouring one. B) Probability of recovering β^I or a neighbouring value. C) Probability of recovering ToM k or a neighbouring value. D) Probability of recovering Guilt α^I or a neighbouring value. E) Probability of revovering $q^I(\zeta^T)$ or a neighbouring value. F) Probability of recovering ζ^I or a neighbouring value. G) Probability of recovering risk aversion ω^I or a neighbouring value. The value $\omega^I = 0.6$ did not occur in the original data set.

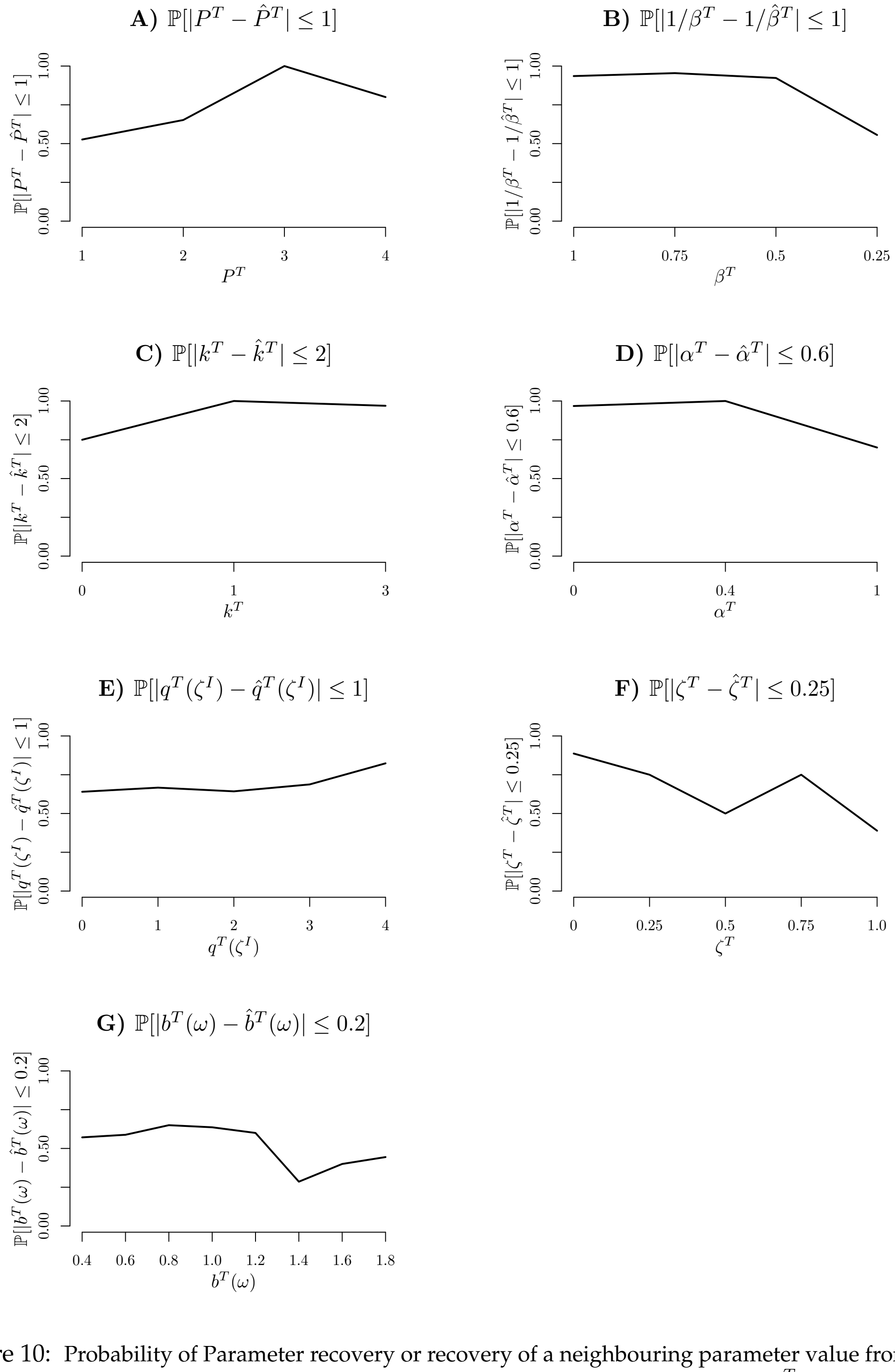

Figure 10: Probability of Parameter recovery or recovery of a neighbouring parameter value from generated exchanges using real subject parameters. A) Probability to recover the planning value P^T or a neighbouring one. B) Probability of recovering β^T or a neighbouring value. C) Probability of recovering ToM k or a neighbouring value. D) Probability of recovering Guilt α^T or a neighbouring value. E) Probability of revovering $q^T(\zeta^I)$ or a neighbouring value. F) Probability of recovering ζ^T or a neighbouring value. G) Probability of recovering risk aversion $b^T(\omega^I)$ or a neighbouring value.

3.5 Behavioural Analysis

The main intent of refining the model is to use it to make inferences about the two investor and two trustee groups that generated the data. The distributions of the new parameters (risk aversion, irritability, awareness) are shown in figure 11A-F. Unexpectedly, using permutation tests, there was no significant different between the BPD and control groups (there is a marginal effect at $p = 0.05$ in investor irritability, with BPD paired investors having a slightly lower average measured irritability compared to HC paired investors). This highlights the fact that in any complex social setting, there are many different components that determine choice.

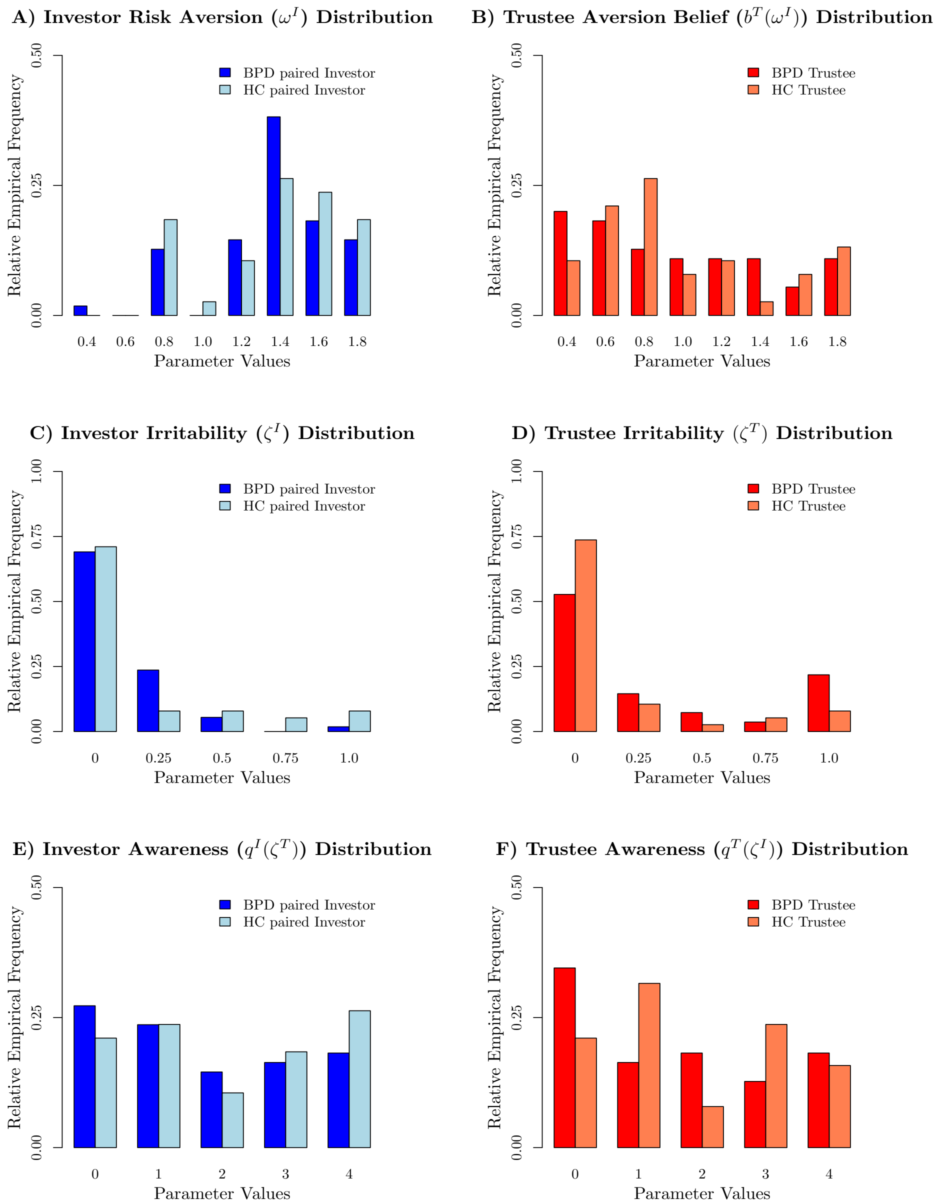

Figure 11: A) Risk Aversion distribution of investors BPD and HC. B) Risk Aversion distribution of trustees BPD and HC. C) Irritability distribution of investors BPD and HC. D) Irritability distribution of trustees BPD and HC. E) Awareness distribution of investors BPD and HC. F) Awareness distribution of trustees BPD and HC.

We therefore considered a slightly model-based characterization of the subjects in which we combined together two key sources of irritation in the model: trustees who are either irritable or are irritation unaware ($\zeta^T > 0$ or $q^T(\zeta) = 0$). These are expected to lead to problems, since the investor is in charge, making for a substantial asymmetry and the above two settings are precisely those under which repair would be hampered (either since the trustee is irritated or since they are ignorant of irritation). We call these trustees as being perilous. This group turns out to be present at a significantly ($p = 0.02$, χ^2-test for equal proportions) higher proportion (72.7%) in the BPD group of [King-Casas et al., 2008], compared with the HC group (47%).

Figure 12 shows investment and repayment profiles for dyads in [King-Casas et al., 2008] including perilous (A) and non-perilous (B) trustees. Not only are these interaction profiles evidently different (in uncorrected two-sided t-tests at $p < 0.05$ at all time points past 1), but also having corrected for this (i.e., sorting healthy controls and BPD trustees according to perilousness), there is no longer a difference between the average investment and return profiles for BPD versus HC dyads ($p > 0.05$ using an uncorrected two-sided t-test).

Figure 12C compares investment and return profiles for investors with little ($\omega^I \leq 1.0$) or substantial risk aversion ($\omega^I \geq 1.0$). Trustee risk aversion profiles do not appear significantly different.

Figure 12D shows the average irritation weight at the time of trustees' choices (averaged across all the subjects in a group, and then compared between groups). The choices of BPD trustees are made under an average 0.24 irritation weight, while the choices of HC trustees are made under an average 0.14. This difference is significant (two sided t-test, $p = 0.03$) between groups. Thus, BPD trustees acted under significantly more irritation.

4 Discussion

Our previous model of the complex collections of choices apparent in the multiround trust task did a generally good job at accounting for many aspects, and generated prediction errors and other parametric regressors that unearthed various key neural processes. However, on closer inspection, it failed to characterize aspects of behaviour at two disparate timescales: a persistent reluctance of the dominant party to submit a portion of their endowment to the potentially fickle trustee in the game; and temporary breakdowns in cooperation and consequent repair. We therefore enriched our model in these two respects, parameterizing risk aversion and irritation.

Note first that, despite its formal appeal, the I-POMDP model has not been extensively used to characterize game theoretic interactions between players. One obvious reason for this is its apparent computational cost. Here we showed that it is perfectly possible to perform approximate I-POMPD inferencee in a relatively complicated model with two intentional dimensions and various other parameters. This augurs well for the future, given the importance and richness of social interactions in both economic decision-making, and as a psychological biomarker in psychiatric conditions.

Risk aversion had previously been suggested as an important factor as part of a discussion as to whether the measurement of trust in games such as the MRT might be corrupted by forms of risk aversion (see [Houser et al., 2009]). It appears that we can measure both in our analysis of the task.

The second, and more extensive change was to capture irritation, along with interactive inference about this made by the two players. This captures the rupture and repair of cooperation, along with the associated threats of these. In the same way that the possibility of punishment or defection maintains cooperative behaviour in tasks such as the public goods game, the possibility of rupture encourages healthy participants to be beneficent.

Our approach to irritability was chosen for its simplicity within the existing model. Further work on a more substantial body of human data will be necessary to fine-tune the dynamics of irritation in

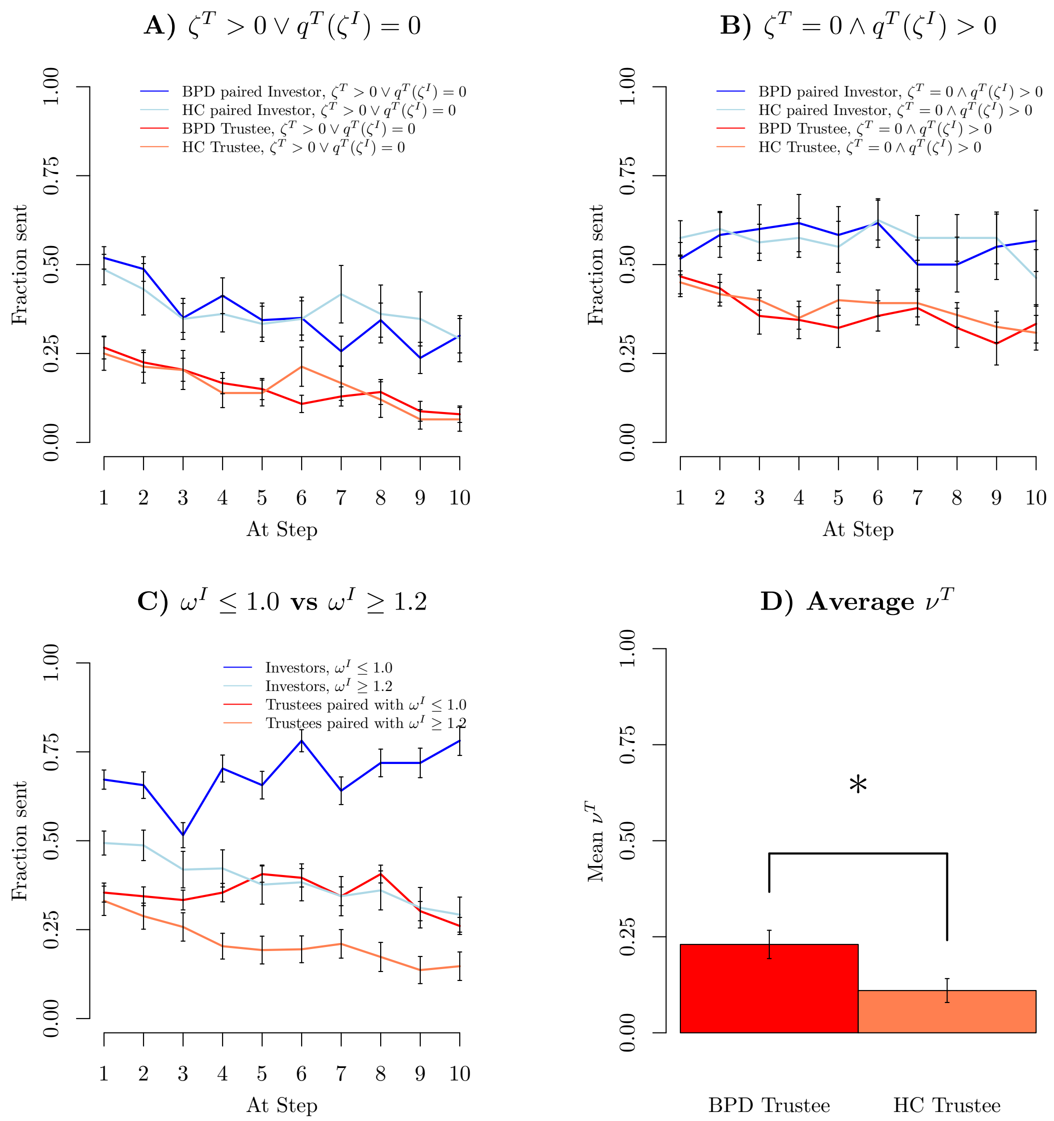

Figure 12: A) Investment and return profile for subgroups of the BPD and HC data sets, defined by $\zeta^T > 0$ or $q^T(\zeta) = 0$. B) Investment and return profile for subgroups of the BPD and HC data sets, defined by $\zeta^T = 0$ and $q^T(\zeta) > 0$. C) Investment and return profile for subgroups defined by $\omega^I \leq 1.0$ (blue, red) or $\omega^I \geq 1.2$ (light blue, coral). D) Average irritation weight of trustee during play by group, BPD (red) and HC (coral). The difference is signficant (two sided t-test) at $p = 0.03$, signified by an asterisk. All errorbars are standard error of the mean.

social exchange. One first step might be to use the model as part of an optimal experimental design framework to realize a computer-based opponent that could extract the most out of each available choice. At present, the relatively small number of actions in our version of the trust task, together with the possibility that the human partners fail to irritate each other even when they are irritable, leaves little room for further sophistication. Given a better understanding of irritation in the model, it would then be possible to refine the concept itself.

The ultimate model has the uncomfortable characteristic of employing 7 parameters to account for the 10 choices of each subject. However, the parameters interact in complex ways in the model, which is why they can generally be reliably inferred, as apparent in our confusion matrices.

The model of irritation departs from conventional models of intentional inference in one important way. In repeated social exchange tasks, it is conventional to model one's partner's *preferences*, which, in Bayes-Nash terms, concerns properties of their *utility functions*. Indeed, this is exactly how earlier studies on the multi round trust game framed the social exchange [Debajyoti et al., 2008, Koshelev et al., 2010, Xiang et al., 2012, Hula et al., 2015]. Here, however, we considered simultaneous intentional inference about both a *utility* and a *policy* (as in [Wunder et al., 2011]) that the player would adopt (indeed, a policy that it would be hard to justify in pure utility terms, given the costs of breaking cooperation). A richer palette of such default behaviors might also prove important in other tasks. Note, though, that it is not yet clear that a suitable notion of equilibrium can be defined (for instance, as the theory of mind level of the players tends to infinity). In Bayes-Nash terms, the present framework considers behavioural strategies incorporating the possibility of irritation shifts. The combination of Kuhn's theorem ([Aumann, 1964]; since our players have perfect recall) and Harsanyi's treatment of mixed strategies ([Harsanyi, 1973]) would be a natural starting point.

The model finally provides a generative model based approach to the characteristica of BPD's play in the multi round trustgame, reported in [King-Casas et al., 2008]. This approach yields a particular type of trustee, the perilous (for the interaction) trustee, which appears to be overrepresented in the BPD sample. As a consequence of this overrepresentation we see considerably more trustee actions likely to have been made under irritation and too little effort to repair the ensuing rupture, especially early on in the game.

Correcting for this subtype, we find equal average behaviours in BPDs and HCs. Thus this subgroup (also present in the HC group, to a lower extend) could be a starting point to define a separate clinical phenotype, currently subsummed under the BPD label with other, less trust rupturing kinds of patients. Such a separation might yield clearer clinical and neurological characterisations of the underlying causes of trust rupturing.

Despite its formal appeal, the I-POMDP model has not been extensively used to characterize game theoretic interactions between players. One obvious reason for this is its apparent computational cost. Here we showed that it is perfectly possible to perform approximate I-POMDP inference in a relatively complicated model with two intentional dimensions and various other parameters. This augurs well for the future, given the importance and richness of social interactions in both economic decision-making, and as a psychological biomarker in psychiatric conditions.

5 Acknowledgements

We thank Terry Lohrenz and Michael Moutoussis for helpful discussions. Special thanks go to the IT support staff at the Wellcome Trust Center for Neuroimaging and Virgina Tech Carilion Research Institute. The authors gratefully acknowledge funding by the Wellcome Trust (Read Montague) under a Principal Research Fellowship, the Kane Foundation (Read Montague) and the Gatsby Charitable Foundation (Peter Dayan). Andreas Hula is supported by the Principal Research Fellowship of Professor Read Montague.

References

[Auer et al., 2002a] Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002a). Finite time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256.

[Auer et al., 2002b] Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. (2002b). The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32:48–77.

[Aumann, 1964] Aumann, R. (1964). Mixed and behavior strategies in infinite extensive games. *Advances in Game Theory, Annals of Mathematics Studies*, 52:627–650.

[Bellman, 1952] Bellman, R. (1952). On the Theory of Dynamic Programming. *Proc. Natl. Acad. Sci. USA*.

[Boost-Libraries, 2014] Boost-Libraries (2014). `http://www.boost.org`.

[Camerer et al., 2004] Camerer, C., Ho, T., and Chong, J. (2004). A cognitive hierarchy model of games. *Q J Econ*, 119:861–898.

[Chiu et al., 2008] Chiu, P., Kayali, M., Kishida, K., Tomlin, D., Klinger, L., and Montague, P. (2008). Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. *Neuron*, 57:463–473.

[Costa-Gomes et al., 2001] Costa-Gomes, M., Crawford, V., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study . *Econometrica*, pages 1193–1235.

[Debajyoti et al., 2008] Debajyoti, R., King-Casas, B., Dayan, P., and Montague, P. (2008). Bayesian Model of Behaviour in Economic Games. *NIPS*, 21:1345–1353.

[Draper, 1995] Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *J.R.Stat.Soc. B*, 57(1):45–97.

[Fehr and Schmidt, 1999] Fehr, E. and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Q J Econ*, 114:817–868.

[Fonagy and Bateman, 2006] Fonagy, P. and Bateman, A. (2006). Mechanisms of change in mentalization-based treatment of BPD. *Journal of Clinical Psychology*, 62(4):411–430.

[Gläscher et al., 2010] Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J. P. (2010). States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. *Neuron*, 66(4):585–595.

[Gmytrasiewicz and Doshi, 2005] Gmytrasiewicz, P. and Doshi, P. (2005). A Framework for Sequential Planning in Multi-Agent Settings. *J Artif Intell Res*, 24:49–79.

[Gmytrasiewicz and Doshi, 2009] Gmytrasiewicz, P. and Doshi, P. (2009). Monte Carlo sampling methods for approximating interactive POMDPs. *J Artif Intell Res*, 31:279–337.

[Harsanyi, 1967] Harsanyi, J. (1967). Games with incomplete information played by "Bayesian" players. *Manage Sci*, 14:159–182.

[Harsanyi, 1973] Harsanyi, J. (1973). Games with randomly disturbed payoffs: a new rationale for mixed-strategy equilibrium points. *Int. J. Game Theory*, 2:1–23, doi:10.1007/BF01737554.

[Houser et al., 2009] Houser, D., Schunk, D., and Winter, J. K. (2009). Distinguishing Trust from Risk: An Anatomy of the Investment Game . *University of Zurich Institute for Empirical Research in Economics Working Paper*.

[Hula et al., 2015] Hula, A., Montague, P., and Dayan, P. (2015). Monte Carlo Planning Method Estimates Planning Horizons during Interactive Social Exchange. *PLoS Comput Biol*, 11(6):e1004254.

[Huys et al., 2011] Huys, Q. J., Cools, R., Gölzer, M., Friedel, E., Heinz, A., and al (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLoS Comput Biol*, 7(4):e1002028.

[Kaelbling et al., 1995] Kaelbling, M., Littman, M., and Cassandra, A. (1995). Planning and acting in partially observable stochastic domains. *Artif Intell*, 101:99–134.

[King-Casas et al., 2008] King-Casas, B., Sharp, C., Lomax, L., Lohrenz, T., Fonagy, P., and Montague, P. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, 321.

[King-Casas et al., 2005] King-Casas, B., Tomlin, D., Anen, C., Camerer, C., Quartz, S., and Montague, P. (2005). Getting to know you: Reputation and Trust in a two-person economic exchange. *Science*, 308:78–83.

[Kocsis and Szepesvári, 2006] Kocsis, L. and Szepesvári, C. (2006). Bandit based Monte-Carlo Planning. *15th European Conference on Machine Learning*, pages 282–293.

[Koshelev et al., 2010] Koshelev, M., Lohrenz, T., Vannucci, M., and Montague, P. (2010). Biosensor Approach to Psychopathology Classification. *PLoS Comput Biol*, 6(10):e1000966.

[Lieb et al., 2004] Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M., and Bohus, M. (2004). Borderline Personality Disorder. *The Lancet*, 364:453–461.

[Loewenstein, 2005] Loewenstein, G. (2005). Hot-Cold Empathy Gaps in Medical Decision Making. *Health Psychology*, 24:49–56.

[MATLAB, 2010] MATLAB (2010). The MathWorks Inc., Natick, Massachusetts.

[McCabe et al., 2001] McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. USA*, 98(20):11832–35.

[McKelvey and Palfrey, 1992] McKelvey, R. and Palfrey, T. (1992). An Experimental Study of the Centipede Game. *Econometrica*, pages 803–836.

[Puterman, 2005] Puterman, M. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.

[R Core Team, 2013] R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Silver and Veness, 2010] Silver, D. and Veness, J. (2010). Monte Carlo Planning in Large POMDPs. *NIPS*, 23:2164–2172.

[Wunder et al., 2011] Wunder, M., Kaisers, M., Yaros, J., and Littman, M. (2011). Using iterated reasoning to predict opponent strategies. *10th International Conference on Autonomous Agents and Multiagent Systems*, 2:593–600.

[Xiang et al., 2012] Xiang, T., Debajyoti, R., Lohrenz, T., Dayan, P., and Montague, P. (2012). Computational Phenotyping of Two-Person Interactions Reveals Differential Neural Response to Depth-of-Thought. *PLoS Comput Biol*, 8(12):e1002841.