



# Multi-Round Trust Game Quantifies Inter-Individual Differences in Social Exchange from Adolescence to Adulthood

RESEARCH ARTICLE

ANDREAS HULA 

MICHAEL MOUTOUSSIS 

GEERT-JAN WILL 

DANAE KOKORIKOU

ANDREA M. REITER 

GABRIEL ZIEGLER 

NSPN CONSORTIUM

*\*Author affiliations can be found in the back matter of this article*

ED BULLMORE 

PETER B. JONES 

IAN GOODYER 

PETER FONAGY 

P. READ MONTAGUE 

RAYMOND J. DOLAN 

][ubiquity press

## ABSTRACT

Investing in strangers in a socio-economic exchange is risky, as we may be uncertain whether they will reciprocate. Nevertheless, the potential rewards for cooperating can be great. Here, we used a cross sectional sample (n = 784) to study how the challenges of cooperation versus defection are negotiated across an important period of the lifespan: from adolescence to young adulthood (ages 14 to 25). We quantified social behaviour using a multi round investor-trustee task, phenotyping individuals using a validated model whose parameters characterise patterns of real exchange and constitute latent social characteristics. We found highly significant differences in investment behaviour according to age, sex, socio-economic status and IQ. Consistent with the literature, we showed an overall trend towards higher trust from adolescence to young adulthood but, in a novel finding, we characterized key cognitive mechanisms explaining this, especially regarding socio-economic risk aversion. Males showed lower risk-aversion, associated with greater investments. We also found that inequality aversion was higher in females and, in a novel relation, that socio-economic deprivation was associated with more risk averse play.

CORRESPONDING AUTHOR:

**Andreas Hula**

Austrian Institute of Technology,  
Vienna, Austria

[Andreas.Hula@ait.ac.at](mailto:Andreas.Hula@ait.ac.at)

KEYWORDS:

Development; Trust; I-POMDP;  
Gender Difference; IQ Effects;  
Model Based; Age; Adolescent;  
Risk Aversion; Socio-Economic  
Status

TO CITE THIS ARTICLE:

Hula, A., Moutoussis, M., Will, G.-J., Kokorikou, D., Reiter, A. M., Ziegler, G., NSPN Consortium, Bullmore, E., Jones, P. B., Goodyer, I., Fonagy, P., Montague, P. R., & Dolan, R. J. (2021). Multi-Round Trust Game Quantifies Inter-Individual Differences in Social Exchange from Adolescence to Adulthood. *Computational Psychiatry*, 5(1), pp. 102–118. DOI: <https://doi.org/10.5334/cpsy.65>

Socio-economic interactions with strangers are inherently risky. However, despite knowing little about potential partners' intentions or probity, we may nevertheless choose to trust them to achieve greater gains for ourselves and to satisfy our own tastes for equity. Learning how much to trust strangers is particularly important during adolescence, as the community of people with whom we interact expands rapidly beyond the bounds of family and local familiarity, and because exploratory collaboration forms a spring-board for longer-term relationships (see Crone & Dahl, 2012, Nelson et al, 2005). Here, we quantify the development of the willingness to trust strangers across age 14 to 25.

Economic games offer a rigorous tool for examining the financial, risk-benefit trade-off associated with trusting others. Trust games of various kinds have been utilized to study social investment behaviour/trust in diverse samples of adolescents (see van den Bos et al, 2010, van den Bos et al, 2011, van den Bos et al, 2012, Belli, Rogers & Lau, 2012, van de Groep 2018) and even young children (see Rosati et al, 2019), with findings varying considerably. Some of these studies have suggested that trust behaviour (investing in an unknown partner with hope of reciprocity) is modulated by the amounts endowed to the investor before reciprocation (which determines 'socio-economic exchange risk') and by sex (female adolescents tending on average to invest less in unknown others). However, results have not been consistently replicated. Measuring the willingness to invest in anonymous partners during adolescence and young adulthood has yielded conflicting results (for increases see Sutter & Kocher, 2007; van den Bos et al, 2010, for decreases see Derks, Lee & Krabbendam, 2014, with an overview in van de Groep, 2018). Findings likely depend on the concrete choice of paradigm and age bracket.

A standard work-horse in the study of psychiatric disorders is a multi-round variant of the 'investor-trustee', or multi-round trust task (McCabe, Rigdon & Smith, 2003, King-Casas et al, 2005) (called the MRT). In the task, one player acts as 'investor', and the other as a 'trustee'. On each of 10 rounds of the game, the investor receives a regular 'wage' from the experimenter and decides on how much of it to invest into the trustee. The experimenter triples the investment, before the trustee chooses the amount of money to return to the investor. Thus, if the investor takes a risk by investing in the trustee, as the amount of repayment is uncertain, but if the trustee indeed pays back an appropriate amount, both parties benefit.

The mutual dependency between investor and trustee for maximising their respective outcomes permits a fine-grained study of the ability to establish and maintain cooperation. This task has thus been used to understand the neural underpinnings of personality disorders, the social deficits found in autism, and also aspects of other disorders (see King-Casas et al, 2008, Koshelev et al, 2010, Mellick, Sharp, & Ernst, 2019).

In recent research, detailed computational models of this task (see Ray et al, 2008, Xiang et al, 2012, Hula, Montague & Dayan, 2015, Hula et al, 2018) were built. These capture the dynamics of the exchange using a few parameters that quantify key characteristics of each participant. These computational models describe behaviour as arising from an interactive process, wherein participants cannot directly observe the characteristics of others, but can gather information about them from the exchange. Participants are parameterised by their social preference factors (i.e. a socio-economic exchange risk and inequity aversion), their capacity to create and avoid ruptures in cooperation, their prior beliefs about these characteristics of their partners, and the sophistication of their theories of other people's minds, all of which we will describe in detail below. In previous work, risk aversion has turned out to be particularly important in describing investors; as this captures a propensity to prefer money that is not put at risk through investing with the trustee over money that is returned through the interaction. Notably, risk-taking is widely discussed to change markedly from adolescence to adulthood (see Steinberg, 2004, Romer, Reyna, & Satterthwaite, 2017, Defoe et al, 2015), but also to be dependent on other individual factors like sex (see Bach et al, 2020, Reniers et al, 2016, Byrnes, Miller & Schafer, 1999, van Duijvenvoorde et al, 2016), socio-economic status and IQ.

Here, we leverage a large ( $n = 784$ ) cross-sectional sample of adolescents and young adults to investigate the relationship of behaviour in the MRT with individual differences regarding age, IQ, sex and socio-economic status. To gain detailed insight into this question, we use our novel computational model to dissociate latent factors contributing to decision-making in this task, in order to link them to our inter-individual variables of interest (age, IQ, sex and socio-economic status). Following several examples in the literature, which point towards changes in trust and reciprocity during adolescence (see Fett et al, 2014, Eisenberg et al, 1991, Derks, Lee & Krabbendam, 2014, Belli, Rogers & Lau, 2012), we found that differences in risk aversion in the MRT statistically explained a substantial portion of the variability of the investment behaviour observed across participants. Regarding inter-individual differences, we found a marked trend towards higher investments and lower risk aversion with increasing age and IQ.

Regarding sex differences, male participants invested more than females. In addition, we found indications of a difference in the attitude towards outcome inequality between male and female subjects, with females being classified as more inequality averse.

Finally, we discovered a previously unreported effect of socio-economic status (SES) on risk-aversion in the MRT. There was a significant effect towards higher risk aversion for subjects living in more adverse socio-economic conditions. Together, these effects show that individual differences can have an important impact on the propensity to invest, and possibly to profit in real-life economic exchanges with strangers.

## METHODS

### SUBJECTS AND DATA COLLECTION

We administered the MRT game to a large sample of young people from London or Cambridge who participated in the Neuroscience in Psychiatry Network study (NSPN). We also collected basic demographic variables (age and self-reported sex, referred to as sex), Socio-Economic status (Neighbourhood 'households in poverty index' of the Office of National Statistics 2014, a score that is higher for more adverse conditions) and the Wechsler Abbreviated Scale of Intelligence (WASI) Score (see Wechsler, 2008, Weiss et al, 2010 ) for IQ. We used raw IQ scores for our analysis. The sample was equally distributed between females and males and between the ages of 14 to 25. Participants were excluded if they currently received help for a psychiatric problem, if they had moderate or severe learning disability or suffered from a serious neurological disorder. For the main analysis, we included only participants for which the above mentioned 4 demographic variables (age, IQ, sex, SES) were fully available ( $n = 784$  out of 788 participants who completed the multi round trust game,  $n = 403$  female, age range 14–25, mean age 19.05,  $sd = 2.96$ ).

A detailed description of the methodology of the NSPN study can be found in Kiddle et al, 2017.

Data and Code used can be found in a public [github repository](#).

### ETHICS STATEMENT

Participants themselves, and, if they were younger than 16, their legal guardians, provided informed consent. All clinical investigation was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Cambridge Central Research Ethics Committee (12/EE/0250).

### THE MULTIROUND INVESTOR – TRUSTEE TASK

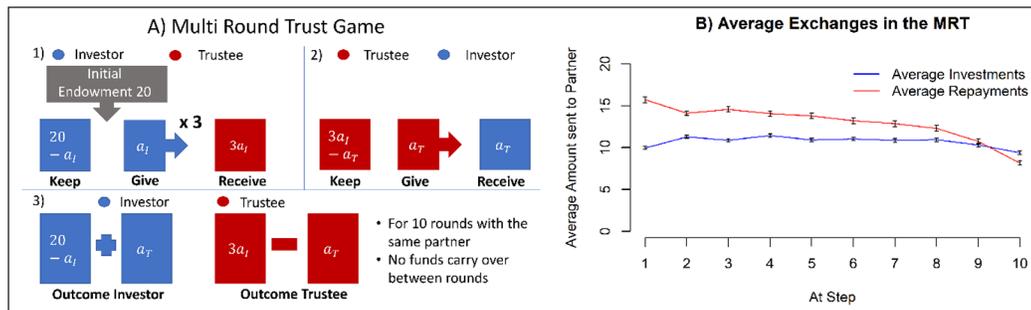
Participants were instructed by trained research assistants about the veridical rules of the game. The task was administered as part of a larger battery of decision-making tasks (see Kiddle et al, 2017).

Participants played the role of the Investor. They were told that they would receive monetary rewards in proportion to their winnings, which was true. As a cover story, they were also led to believe that they were playing with a peer from the same study, playing anonymously from

another site, who would also be paid in proportion to their own winnings. Importantly, they did not know the name, sex or any background of their partner. They were asked to play according to their own goals and preferences, rather than optimize an experimenter-specified goal. At the end of the whole study they were debriefed as to the true nature of their ‘partner’, which was a computer algorithm that emulated the behaviour of healthy adult Trustees (see King-Casas et al, 2005).

Participants were encouraged to play as best they could, but were told that the experimenter was interested in how young people made decisions during interactions according to their own preferences and values. Thus, unlike other tasks in the battery, they were not instructed to achieve any particular goal or outcome – this was up to them.

The rules of the game were as follows (see also **Figure 1A**): In each of the ten rounds, the investor received an initial endowment of 20 monetary units, or play-coins, and could decide the amount (in whole coins) to transfer to the trustee. The experimenter trebled this quantity and then the trustee (in our case, the computer algorithm mentioned above) decided how many coins to return to the investor: this ranged between 0 coins and the total amount received. The repayment by the trustee was not increased by the experimenter. After the trustee’s action, the investor was informed of the outcome, and the next round started. We thoroughly tested the understanding of participants before the game started and encouraged them to ask any questions they had. All participants reported here understood the task well, agreed to play it and provided data of adequate quality for analysis.



**Figure 1 A)** Schematic representation of the multi round trust game (MRT). **B)** Average Investments and Repayments in the MRT in this study. Errorbars are standard errors of the mean (SEM).

## MODEL

We modelled behaviour using an interactive, partially observable Markov decision process (I-POMDP) (see Gmytrasiewicz & Doshi, 2005). We outline the model here, and describe the seven parameters  $\theta = (\alpha, \omega, k, P, \zeta, q(\zeta), \beta)$  that determine the behaviour of our modelled investors, with details on the employed internal state model, inference and statistical properties of the model to be found in Hula et al, 2018. In short, the I-POMDP is predicated on three structural characteristics: the rules of the task (which define an MDP); the assumed characteristics of the subjects playing the task; and the ignorance of the subjects about their partners (leading to partial observability), ‘partner’ in this context always referring to the trustee. For computational reasons, we exploited prior studies to restrict the values of the parameters.

According to the I-POMDP characterization, the investor assumes that the trustee processes the information received through the task and makes choices in a structurally similar manner to the investor themselves. In our actual task, this is not formally true, since the trustee is simulated by a computer program that matches the current situation to a database of recorded interactions. Nevertheless, this assumption allows us to build a faithful model of the investors, and thereby interpret their individual characteristics.

We note that the parameters used here were discretized partly for computational reasons, because the interactive inference in the I-POMDP is quite costly (for  $\alpha$  and  $\zeta$ ). Furthermore, some of the parameters are discrete by nature ( $k, P$ ). The remaining ones ( $\omega, q(\zeta), \beta$ ) were discretized on a grid to allow for exhaustive search over the overall parameter space.

## Inequality Aversion

We start with positive inequality aversion or guilt,  $\alpha \in \{0, 0.4, 1\}$ , the settings corresponding to no inequality aversion, slightly averse and fully averse (the grid for  $\alpha$  having been determined in Hula, Montague & Dayan, 2015). This quantifies how sensitive a subject is to advantageous unequal outcomes (see Fehr & Schmidt, 1999), i.e. how much they prefer an equal outcome. If  $\chi_I$  denotes the current round outcome of the investor, and  $\chi_T$  denotes the current round outcome of the trustee, then the subjective utility  $u_I$  of the investor with inequality aversion  $\alpha$  is

$$u_I = \chi_I - \alpha \max\{\chi_I - \chi_T, 0\}. \quad (1)$$

where the outcome is

$$\chi_I = (20 - a_I) + a_T. \quad (2)$$

and

$$\chi_T = 3a_I - a_T. \quad (3)$$

That is, the utility for an investor with  $\alpha > 0$  is reduced if their associated trustee earns less than themselves. The equivalent is true for the trustee. For the investor, a critical contribution to the trustworthiness of the trustee can be framed as beliefs and learning of the trustee's  $\alpha$ .

If the trustee actually has a high value of  $\alpha$ , then it is safe for the investor to make a substantial investment. However, in our anonymous setting, the investor does not know the trustee's inequality aversion. Instead, the investor has to learn this from the trustee's behaviour. In principle, the trustee would be in the same position relative to the investor's inequality aversion. This makes the problem partially observable – at least some information is not known to the players, who therefore are assumed to perform approximate Bayesian inference to learn about these unknown factors.

## Risk Aversion

The parameter  $\omega \in \{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8\}$  quantifies the risk aversion of an investor, i.e. the relative value they accord to the (certain) amount that they keep and do not invest. This is quantified by modifying the notional outcome for the investor,  $\chi_I$ : If the amount the investor gives is denoted by  $a_I$  and the trustee return is denoted by  $a_T$ , then

$$\chi_I^{true} = \omega(20 - a_I) + a_T. \quad (4)$$

And

$$u_I^{true} = \chi_I^{true} - \alpha \max\{\chi_I^{true} - \chi_T, 0\}. \quad (5)$$

That is, the investor weighs the kept amount  $(20 - a_I)$  more ( $\omega > 1$ ) or less ( $\omega < 1$ ) than the amount  $a_I$  received through the interaction. This automatically weighs the variable (and hence uncertain) repayments differently from the certain amount that is kept. We assume that the investor believes her risk aversion to be “natural” in the sense that her model of the trustee will operate as if her actual degree of risk aversion was known. The trustee on the other hand is modelled as having a fixed belief of the investor's risk aversion, dubbed  $b^T(\omega)$ , which is fit in the same manner as the other parameters, i.e. through maximum likelihood estimation. Since all our subjects are investors, this parameter will not play a further role in the analysis.

## Theory of Mind

The most conceptually complex parameter is the level of thinking (theory of mind (ToM) level) (see Costa-Gomes, Crawford & Broseta, 2001, and related Camerer, Ho & Chong, 2004)  $k \in \{0, 1, 2, 3, 4\}$ . Level  $k$ -thinking classifies beliefs about the nature of other players by the number of mentalizing steps they employ to model the interaction. Take inequality aversion. A level 0 investor (“she”) would build a model of the trustee (“he”) based on her beliefs about his degree of “inequality aversion”

(see Ray et al, 2008, Xiang et al, 2012, Hula, Montague & Dayan, 2015, Hula et al, 2018)]. A level 1 investor would also model the trustee's beliefs about her own degree of inequality aversion. Thus, the investor has to put herself into the trustee's shoes to mentalize what he believes about her. This recurses – thus a level 2 investor would model the trustee's beliefs about her own beliefs about his inequality aversion. In general if the investor is aware that “my trustee partner might think on a (maximum) level  $k-1$ ” then the investor's decision making will be at level  $k$  (since they take the level  $k-1$  acting into account), however we make the further simplifying assumption that a level  $k$  investor considers her partner to operate *exclusively* at level  $k - 1$ . The alternative would be a level  $k$  investor potentially considering the trustee to operate at all levels below  $k$ , but at least at level  $k-1$  with positive probability, which would however diffuse the choice preferences too much within the 10 round interaction of the MRT.

The same is true for the trustee modelling the investor. In our models, this potentially unbounded iterative structure of beliefs was limited to four steps, since higher levels do not appear to lead to notably different behaviours (see Hula et al, 2018).

## Planning

Next comes parameter  $P \in \{1, 2, 3, 4\}$ , which quantifies the number of future exchanges to be taken into account in thinking forward. This is called the subject's planning horizon and for the investor it has two main effects (detailed in Hula, Montague & Dayan, 2015): longer planning horizons allow for the execution of consistent gameplay strategies and allow to pre-empt exploitation by the partner by realizing initial cooperation might be succeeded by exploitation of the obtained reputation of trust worthiness (in particular taking the form of the trustee no longer repaying fair splits).

## Irritability

The parameter  $\zeta \in \{0, 0.25, 0.5, 0.75, 1\}$ , called irritability, provides a model of punishment actions, i.e. sudden investments or repayment close to 0, for instance after an action by the trustee that did not satisfy the expectations of an investor. We provide some more detail on this parameter, as it (and the following irritation awareness below) might be the most idiosyncratic of the model parameters, since it arose out of the study of Borderline Personality Disorder (BPD) data (see also King-Casas et al, 2008). This parameter governs a potential state shift to and from an irritated state, characterised by setting the planning horizon,  $P$ , to 0, removing any positive inequality aversion/guilt,  $\alpha$ , and failing to model the mental state of the other person (i.e., when either person is acting irritated, their theory of mind level  $k$  is locally set to  $-1$ , which implies that they fail to mentalize about the other player at all). The values of all other parameters are kept equal to those in the “nonirritated” state. The size of the irritation parameter controls how quickly a subject switches from the non-irritated state to the fully irritated state i.e. after an irritation a subject with  $\zeta = 0.25$  will follow a mixture policy of 25% irritated action preference and 75% non-irritated action preference, while a subject with  $\zeta = 0.75$  will follow a 75% irritated action preference and 25% non-irritated action preference. Further irritation events increase the shift to the irritated policy i.e. a subject with  $\zeta = 0.25$  following a 25% irritated policy getting irritated immediately again will shift to a 50% irritated policy, while a subject with  $\zeta = 0.75$  following a 75% irritated policy will shift to a 100% irritated policy. Irritation can be undone by reciprocating or raising investment despite the irritated action and this shifts the policy making back from the irritated state at an analogue pace to the shift towards irritation. In this case the irritated state is left again and the “normal” method for making choices governs choice again (additional details in Hula et al, 2018).

## Irritation Awareness

Additionally, subjects may be able to infer, whether their partner might be irritable. This is realized via approximate Bayesian inference on the partner's irritability parameter based on 5 possible initial belief distributions represented in the parameter  $q(\zeta) \in \{0, 1, 2, 3, 4\}$ , called “irritation awareness”. This parameter governs (in a monotonically increasing manner) how sensitive choice making is to the possibility of partner irritability (see Hula et al, 2018 for further details), i.e. a subject starts out

with different prior weights on the 5 possible values of the partner's  $\zeta$ , which correspond to “never irritable”, “unlikely to be irritable”, “possibly irritable”, “likely irritable” and “certainly irritable”.

## Interactive Inference

Two of the parameters (positive inequality aversion/guilt and irritability) are assumed to be inferred by the investors during the interaction, while the investor assumes all other parameters, (risk aversion, initial irritation belief and planning), to be the same in the partner as in themselves. We followed this strategy because computational constraints and data constraints (i.e., the number of rounds played per subject) limit the number of variables that we can model to be inferred jointly. Other candidates besides inequality aversion and irritability to be inferred interactively might be planning, theory of mind or risk aversion. However, for theory of mind and planning the amount of evidence gathered interactively during the ten rounds of the game (rather than from the concluded full interaction) would likely be little and considerable assumptions on the most likely level and planning horizon would have to be in place for the actor to benefit from this, which in turn is likely to yield again a situation as with the currently fixed parameters. For risk aversion, we did not consider this since there is currently no mechanism in the model that would shift risk aversion in one way or another and hence we did not expect it to yield interesting new forms of interaction. Future models will investigate this further.

We note that interactive inference by the investor was limited in two cases: If the investor invested 0 (and hence the trustee could not act) and if the investor chose the “invest 5” action: In this case the fair split is achieved by the trustee returning nothing and hence the trustee action does not carry information about inequality aversion preferences or irritability of the trustee.

## Inverse Temperature and Choice Making

The investor model can calculate so-called action values  $Q(a|h)$ . These quantify the expected value over the future planning horizon of executing action  $a$  (an investment or a return) given that the past history of investments and returns is  $h$ .

The action values determine the probability of choosing an action  $a$  through a logistic softmax function (Eq. 6) with a temperature parameter  $\beta \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{1}{1}\}$ :

$$\mathbb{P}[a|h] = \frac{e^{\beta Q(a|h)}}{\sum_c e^{\beta Q(c|h)}} \quad (6)$$

Again, for computational convenience, we constrain  $a$  for the investor to take one of five possible values, corresponding to  $\{[0,2], [3,7], [8,12], [13,17], [18,20]\}$ , being treated as investments of 0, 5, 10, 15 or 20 respectively. Similarly, the trustee returns are discretized to rounded fractions  $\{0, 1/6, 1/3, 1/2 \text{ or } 2/3\}$  of the tripled investment, i.e., the amount they receive. Returns above  $2/3$  of the received amount were very rare.

## Classification and Parameters

Each subject is classified according to the parameter vector which generated the lowest negative log likelihood for the observed interaction. That value was found by search over all possible parameter values.

A summary of all parameters and their ranges can be found in [Table 1](#) below.

## Investment Structure

The average investment and repayment values per round are shown in [Figure 1B](#). If the investor gives 10, then any return from the trustee of more than 10 nets the investor with as much as, or more than the trustee, and so represents reliable or over-reciprocating cooperation. [Figure 1B](#) shows that the trustee starts out in this manner (automatically reflecting the characteristics of the human participants on which its choices are based); this might coax the investor into

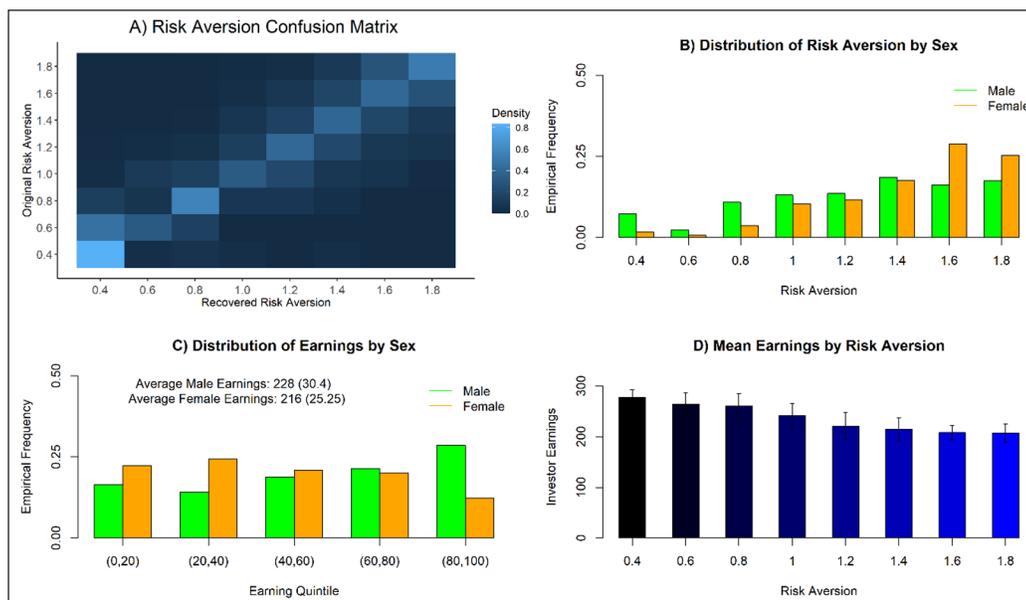
PARAMETER SYMBOL	PARAMETER NAME	RANGE	MEANING
$\alpha$	Inequality Aversion/Guilt	{0, 0.4, 1}	Degree of sensitivity to an unfair outcome against the other player.
$\omega, \mathbf{b}^T (\omega)$	Risk Aversion	$\left\{ 0.4, 0.6, 0.8, 1, \right. \\ \left. 1.2, 1.4, 1.6, 1.8 \right\}$	Multiplier for value of money kept over money returned by the partner.
$k$	ToM Level	{0, 1, 2, 3, 4}	Number of recursive reasoning steps in representing beliefs of the other player.
$P$	Planning	{1, 2, 3, 4}	Number of steps ahead planned into the interaction.
$\zeta$	Irritability	{0, 0.25, 0.5, 0.75, 1}	Measure of shift towards punishment behaviour, when experiencing below expectation partner actions.
$\mathbf{q}(\zeta)$	Irritation Awareness	{0, 1, 2, 3, 4}	Awareness of partner irritability. 0 = unaware, 4 = partner for sure irritable.
$\beta$	Inverse Temperature	$\left\{ \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{1}{1} \right\}$	Measure of stochasticity in choices given their expected utilities.

investing more. It also reveals that the average investment remained remarkably stable. The actual individual investment trajectories are related to the model mechanisms in more complex ways, such as decreasing investments towards the end of the interaction (an effect of planning, see Hula, Montague & Dayan, 2015) and substantial drop offs of investments during the interaction (for irritable investors, see Hula et al, 2018).

### Risk Aversion Confusion Analysis

To confirm that our generative model pin-pointed key parameters for individual participants, a prerequisite for analysis of individual variation, we used the model to create sample trajectories (1 for each dyad) based on the parameter values that we inferred from the human subjects. Using these generated dyads, we estimated a full set of new parameters. **Figure 2A** shows the confusion matrix for the parameter that turns out to be most critical: risk aversion. This matrix compares the actual value of the risk aversion on the basis of which a trajectory was created to the value of risk aversion that we inferred from the trajectory. This quantifies the quality of inference about the crucial risk aversion parameter. We found that parameter recovery of risk aversion was very stable (with misestimations consisting primarily of a shift to a directly adjacent parameter value, which would have limited impact on the kind of monotone relationships we are investigating below) and thus inferences based on the risk aversion parameter are justified.

**Table 1** Model Parameters, Parameter Ranges and Interpretation of the parameters.



**Figure 2** From top left to bottom right. **A)** Confusion matrix: trajectories were sampled from the generative model using the full collection of parameters inferred from each subject (y-axis); then new parameter values were re-estimated from these trajectories (x-axis). The matrix shows the conditional probability of the re-estimated value of risk aversion as a function of the generating risk aversion. Lighter values indicate higher probabilities. Rows sum to 1. **B)** Frequency of occurrence of each risk aversion value, separated by sex. **C)** Average earnings separated by sex and classified according to membership in total earnings quintiles in this study. **D)** Average investment levels separated by subjects' Risk Aversion parameter. Error bars are standard deviations (SD).

## Analysis Set-Up

The prime expression of trust in the MRT is the willingness to invest in a partner. Our first hypothesis thus was, that the parameter estimates would correlate with the average/sum investment levels. We thus proceeded to estimate the proportion of variance in the total investment  $y$  across all rounds that could be accounted for linearly (using the function “lm” implemented in R, see R Core Team, 2017) by the fitted parameters treated as discrete ordinal quantities (In Eq. 7 below,  $\epsilon$  denotes an error term):

$$y = b_0^\theta + b_k^\theta k + b_p^\theta P + b_\alpha^\theta \alpha + b_\omega^\theta \omega + b_\beta^\theta \beta + b_\zeta^\theta \zeta + b_{q(\zeta)}^\theta q(\zeta) + \epsilon. \quad (7)$$

This explained a full 69% of the variance (adjusted  $R^2$ ) of the total investment, while the socio-economic risk aversion parameter  $\omega$  as a sole regressor can account for 49% (adjusted  $R^2$ ) of the variance. This is reasonable, as the sum investments would be most affected by a parameter governing the preference of money kept over money gained directly.

We then examined how demographic variables may account for social-cognitive characteristics, as captured by model parameters. To assess this, we used linear and quadratic regression for the metric parameters (risk aversion, temperature), and ordinal regression for the others (ToM level, planning, inequality aversion, irritability and the irritation belief). To make the effects of the demographic variables in the linear models more easily comparable, we standardized them by Z-Scoring (subtracting the respective mean and dividing by the standard deviation) before squaring them and before using them in the regressions.

Using  $\phi$  as a placeholder for variable of interest for the metric variables, we considered models of the form

$$\phi = b_0^\phi + b_{Age}^\phi Age + b_{IQ}^\phi IQ + b_{SES}^\phi SES + b_{sex}^\phi sex + b_{Age^2}^\phi Age^2 + b_{IQ^2}^\phi IQ^2 + b_{SES^2}^\phi SES^2 + \epsilon \quad (8)$$

Here, for sex, we conventionally coded male as 0 and female as 1.

Using  $\phi$  again as a placeholder for the variable of interest, for ordinal regressions (implemented in R (R Core Team, 2017) via the “polr” function), we studied

$$l(\mathbb{P}[\phi < i]) = b_{0,i}^\phi + b_{Age}^\phi Age + b_{IQ}^\phi IQ + b_{SES}^\phi SES + b_{sex}^\phi sex + b_{Age^2}^\phi Age^2 + b_{IQ^2}^\phi IQ^2 + b_{SES^2}^\phi SES^2 + \epsilon \quad (9)$$

With a logit link function  $l$  and a given level  $i$ . We only report significance levels for the variables and not for the intercepts in the ordinal regression.

We started with these full models and employed stepwise model selection, based on the Akaike Information Criterion (AIC), in order to arrive at the most parsimonious model to describe our data. Model selection via the AIC was implemented using via the “stepAIC” function in the statistics language R (see (R Core Team, 2017)). At each step, this considers whether AIC improves upon removing just one term, so tests 7 models in the first step, 6 models in the second, etc., until the AIC can no longer be improved. Below we give the winning model for each parameter and discuss the significant and marginally significant (at  $p < 0.05$  and  $p < 0.1$  respectively) surviving variables in the winning model.

We illustrate the correlations between the demographic variables below in [Table 2](#) (Kendall’s tau coefficients) and [Table 3](#) (p-values).

	AGE	IQ	SES	SEX
Age	1			
IQ	0.088	1		
SES	0.1	-0.046	1	
Sex	-0.004	-0.09	-0.047	1

**Table 2** Kendall’s tau correlations (below diagonal) among the demographic variables employed in this study.

	AGE	IQ	SES	SEX
Age	0			
IQ	0.00026*	0		
SES	0.000021*	0.056	0	
Sex	0.88	0.002*	0.11	0

**Table 3** p-values (below diagonal) for the correlations among the demographic variables employed in this study. Asterisk (\*) signifies correlations with a p-Value below 0.05.

## RESULTS

The model captured trial-by-trial behaviour well, the negative loglikelihood (NLL) being 8.13, corresponding to an average 0.43 likelihood in the model for the actually made subject choice, based on the current game history and the model parameters (uniform chance = 0.2). In the following we present the outcomes of model selection on the regression models between demographic factors and model parameters.

### RELATIONS BETWEEN SOCIOECONOMIC RISK AVERSION AND DEMOGRAPHIC VARIABLES

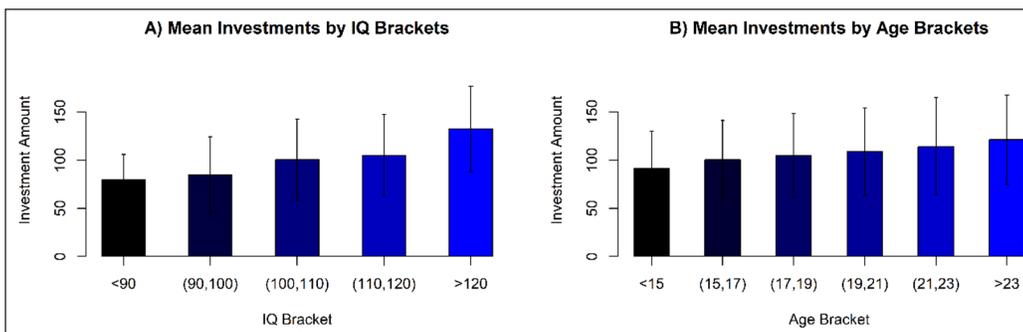
For risk aversion ( $\phi = \omega$ ), the ultimate model was:

$$\omega = b_0^{\omega} + b_{Age}^{\omega} Age + b_{IQ}^{\omega} IQ + b_{SES}^{\omega} SES + b_{sex}^{\omega} sex + b_{SES^2}^{\omega} SES^2 + \epsilon, \quad (10)$$

We found highly significant negative relationship between Risk Aversion and Age and raw (non-age-standardized) IQ, and a significant relationship with sex (respectively:  $b_{Age}^{\omega} = -0.3, p = 10^{-5}$  for Age,  $b_{IQ}^{\omega} = -0.43, p = -10$  for IQ and  $b_{sex}^{\omega} = 0.82, p = 10^{-10}$  for sex ) as well as a positive relation to the SES score  $b_{SES}^{\omega} = 0.19, p = 0.04$ ) and a negative relation to the squared SES score  $b_{SES^2}^{\omega} = -0.12, p = 0.01$ ).

We investigated the effects on risk aversion in more depth, as risk aversion explained a high proportion of the variance in total investments and was strongly correlated with the socio-demographic variables. We plot the distribution of the risk aversion parameter split by sex in **Figure 2B**. Every risk aversion setting occurred in each sex, but in differing proportions. The key difference was an enhanced presence of high risk aversion settings in females compared to males. This influences excess earnings on average of males (significance and 95% confidence interval for the mean difference  $\Delta\mu: p = 10^{-7}, \Delta\mu \in [8.1, 16.9]$ , male mean = 228, sd = 30.4, female mean = 216, sd = 25.25). This represents a small to medium effect size (Cohens  $d = 0.44$ ), however we note that for this type of study, which is heavily affected by interindividual differences, current literature suggests, that small effect sizes are common and the associated effects potentially relevant (see Funder & Ozer, 2019). We note that the relevant extremes of the risk aversion distribution are very well reproduced in model-derived simulations, as evidenced by the confusion matrix in **Figure 2A**. Thus, we rule out the possible explanation of this being due to systematic skewing caused through model-fitting. The distribution of membership in one of the 5 earnings quintiles by sex can be seen in **Figure 2C**. This closely matches the respective risk aversion distribution in our generative model. The effects of risk aversion on the obtained earnings of an investor can be seen in **Figure 2D**.

To illustrate the demographic effects found here, in an empirical correlate of risk aversion (namely the total investments), **Figure 3A** and **3B** show the relationships between raw (non-age-standardized) IQ vs. Investment and Age vs. Investment. The monotonic effect of each of these variables on total investments is readily observed.



**Figure 3 A)** Average investments separated by raw (non-age-standardized) IQ groups (brackets of width 10 IQ points). Error bars are standard deviations (SD). **B)** Average total Investments ordered by participant age, in 2 year brackets. Error bars are standard deviations (SD).

For inequality aversion ( $\phi = \alpha$ ), the ultimate model was:

$$l(\mathbb{P}[\alpha < i]) = b_{0,i}^\alpha + b_{Age}^\alpha Age + b_{SES}^\alpha SES + b_{sex}^\alpha sex + b_{Age^2}^\alpha Age^2 + \epsilon, \tag{11}$$

We found a significant ( $b_{sex}^\alpha = 0.47, p = 0.001$ ) effect of sex on inequality aversion, females having a higher empirical frequency of inequality aversion values above 0. One empirical correlate of this is the percentage of male or female actors making 0-investments in the data set (total number 247 of 0-investments, 159 by male, 88 by female participants) which is 64.4% of 0-investments for male and 35.6% for female participants. Furthermore, we found a marginally significant effect of age on inequality aversion ( $b_{Age}^\alpha = -0.14, p = 0.06$ ). The effects of age squared ( $b_{Age^2}^\alpha = 0.11, p = 0.16$ ) and SES ( $b_{SES}^\alpha = -0.11, p = 0.13$ ) were not significant despite surviving model selection.

**RELATION BETWEEN IRRITATION AWARENESS, IRRITABILITY, PLANNING, THEORY OF MIND AND DEMOGRAPHIC VARIABLES**

For several parameters ( $\phi = q(\zeta), \zeta, p, k$ ), we obtained no noteworthy (i.e. surviving model selection non-constant) relation with the basic demographic variables.

**RELATION BETWEEN INVERSE TEMPERATURE AND DEMOGRAPHIC VARIABLES**

For the inverse temperature ( $\phi = \beta$ ), the ultimate model was:

$$\beta = b_0^\beta + b_{IQ}^\beta IQ + b_{IQ^2}^\beta IQ^2 + b_{SES^2}^\beta SES^2 + \epsilon, \tag{12}$$

We found a significant though small effect of raw (non-age-standardized) IQ ( $b_{IQ}^\beta = 0.024, p = 0.03$ ), a (marginally) significant effect of IQ squared ( $b_{IQ^2}^\beta = 0.015, p = 0.07$ ), and no significant effect of  $SES^2$  ( $b_{SES^2}^\beta = -0.009, p = 0.14$ ), despite the latter factor surviving model selection. This suggests that our current model fits slightly better for higher IQ subjects.

**CORRELATIONS BETWEEN MODEL PARAMETERS**

We then investigated the correlations between model parameters in this human data sample. The correlation coefficients and their p-values can be seen in [Tables 4](#) and [5](#) respectively below:

	<i>K</i>	<i>P</i>	$\omega$	$\beta$	$\alpha$	$\zeta$	<i>Q</i> ( $\zeta$ )
<b><i>k</i>, ToM Level</b>	1						
<b><i>P</i>, Planning</b>	0.0003	1					
<b><math>\omega</math>, Risk Aversion</b>	0.097	0.13	1				
<b><math>\beta</math>, Inverse Temperature</b>	-0.015	0.01	-0.1	1			
<b><math>\alpha</math>, Inequality Aversion</b>	0.06	0.12	0.34	-0.13	1		
<b><math>\zeta</math>, Irritability</b>	0.038	-0.0002	0.05	-0.089	0.04	1	
<b><i>q</i>(<math>\zeta</math>), Irritation Awareness</b>	-0.046	-0.12	0.22	-0.29	0.06	0.007	1

**Table 4** Kendall's tau correlations (below diagonal) of the subject parameters derived from the minimum NLL fit.

	<i>K</i>	<i>P</i>	$\omega$	$\beta$	$\alpha$	$\zeta$	<i>Q</i> ( $\zeta$ )
<b><i>k</i>, ToM Level</b>	0						
<b><i>P</i>, Planning</b>	0.99	0					
<b><math>\omega</math>, Risk Aversion</b>	0.001*	0.000005*	0				
<b><math>\beta</math>, Inverse Temperature</b>	0.65	0.7	0.0005*	0			
<b><math>\alpha</math>, Inequality Aversion</b>	0.07	0.00015*	0*	0.00004*	0		
<b><math>\zeta</math>, Irritability</b>	0.26	0.99	0.1	0.005*	0.2	0	
<b><i>q</i>(<math>\zeta</math>), Irritation Awareness</b>	0.14	0.000039*	0*	0*	0.049*	0.8	0

**Table 5** p-values (below diagonal) of the subject parameters derived from the minimum NLL fit. Asterisk (\*) signifies correlations with a p-Value below 0.05. P-values below  $10^{-13}$  denoted as 0.

Risk Aversion correlates positively with inequality aversion, planning, theory of mind level and irritation awareness, as well as negatively with the inverse temperature parameter. Hence subjects with high risk aversion would tend to prefer equitable outcomes and demonstrate longer planning, while subjects with lower risk aversion would tend to have higher choice certainty. Inequality aversion correlates negatively with inverse temperature and positively with risk aversion, planning and irritation awareness. This suggests that more inequality averse subjects tended to be more cautious when engaging in the exchange. The inverse temperature parameter correlates negatively with risk aversion, irritability, irritation awareness and inequality aversion, suggesting that subjects optimizing their outcome followed more consistent strategies. Irritation awareness correlates negatively with planning and the inverse temperature and positive with risk aversion and inequality aversion, suggesting that subjects more aware of partner retaliation were also more risk averse, yet had a tendency to be slightly more inequality averse.

## DISCUSSION

We analysed the decision-making behaviour of a large cross-sectional sample ( $n = 784$ ) of 14 to 25 year-olds performing as investors in a social exchange task, in which they chose their preferred level of investment to optimise their own preferences about earnings of themselves and their partners. We employed an interactive-agent-based model that quantified investor behaviour according to seven key characteristics, or parameters. In total, these key characteristics explained almost 70% of the variance of the summed investments, hence giving a close account of an important statistic of the success of the social interaction. We found that the novel model-based measure of risk aversion in the MRT (introduced in Hula et al, 2018) alone accounted for 44% of the variance in the total earnings of the subjects.

The socio-economic risk aversion parameter consistently decreased with age, resulting in on average higher investment amounts and earnings in young adults compared to adolescents. This may be an important mechanistic factor in the observed (but not undisputed) relationship in the literature between trust and age during this developmental period (see van den Bos et al, 2010). Risk aversion was lower on average in participants with higher IQ and males compared to females, controlling for age and socio-economic status. We also found a trend towards inequity averse play to be stronger in females compared to males.

The various dependencies of risk aversion relate in interesting but complex ways to the existing literature of how reward sensitivity and strategies of play develops in adolescence in a multitude of economic games (see an overview in Sutter et al, 2019). This may partly be because our definition of risk in this task is slightly different from that in other framings, for instance including an endowment effect. We observed that socio-economic risk aversion decreased with age between 14 and 25 years of age. Previous findings are that risk aversion outside the context of social exchange increases with age much later in the life span (see Rutledge et al, 2016), or is independent of age (for instance van de Groep et al, 2018). In contrast, and consistent with our findings, trust and reciprocity in trust games has been found to increase over adolescence (in particular see van den Bos et al, 2010, Reniers et al, 2016).

We note that the reported effect sizes could be seen as small, however this is quite typical for the kind of study of interindividual differences, that we performed. A recent discussion of this aspect can be found in Funder & Ozer, 2019.

Our finding that risk aversion decreased with IQ is consistent with observations that IQ is correlated with higher investments and earnings, and trust and reciprocity (see Proto, Rustichini & Sofianos, 2017). Here, we caution that IQ is subject to factors such as motivation (see Breuning & Zella, 1978, Duckworth et al, 2011, Borghans, Meijers & Weel, 2013), education, socio-economic factors (see Baker et al, 2015) and even the life context in which people from different countries and ethnicities grow up. This is consistent with socioeconomic risk-aversion in our task being a motivational parameter, rather than one reflecting how costly a more complex or accurate model of the task might be. The latter might be expected under a simplistic assumption that IQ reflects essential cognitive capacity.

The differences in risk aversion that we found between females and males is consistent with others in the behavioural literature (see van de Groep et al, 2018, van den Akker et al, 2020, Gneezy, Leonard & List, 2009, Charness & Gneezy, 2012) and in recent fMRI studies (see Lemmers-Jansen et al, 2017). However not every study found this effect (see van den Bos et al, 2010).

Our findings on the effects of sex on inequality aversion are also consistent with the literature on adults (see for instance Fehr, Naef & Schmidt, 2006). It might be related to differences in empathy and perspective taking, that were found in studies of adolescent behaviours too (see Van der Graaff, 2014). There may be a sex specific difference in the distribution of response strategies to taking risks and social threat. However, we note that the effect on inequality aversion though significant, is relatively small and would need further investigation.

Finally, it is notable that in our model we observed a novel correlation of risk aversion with socioeconomic status. Risk aversion was higher (and thus trust measured as investments lower) in subjects in more adverse socio-economic conditions (although a quadratic effect in SES counteracts this observation for the high end of the SES spectrum i.e. the most adverse status subjects, in our coding). This is in line with studies finding positive effects of economic support during adolescence having effects on the behaviour in economic games (see Amir, Jordan & Rand, 2018, Luo et al, 2018). This finding is new (and will require replication), yet studies using other measures of trust to measure the effect of SES do exist (see Navarro-Carrillo, Valor-Segura & Moya, 2018, Stamos, Altsitsiadis & Dewitte, 2019). Notably, the study by Amir, Jordan & Rand, 2018 did also find an association between increases in risk aversion and more adverse conditions during childhood, but in that study the subjects with more adverse backgrounds also were more likely to offer higher splits on average in an ultimatum game. The simultaneous presence of a linear and a quadratic effect (with opposing signs) of SES on risk aversion in our analysis suggests a more complicated relationship, on which further research is needed.

Concerning the relation between IQ and the inverse temperature parameter, it appears that higher IQ subjects played less stochastically given how they evaluated the options available, at least according to our model. This is consistent with recent work from our group showing that a measure of decision-making which loads heavily on inverse-temperature across many, mostly non-social, tasks has a similar relation to IQ (see Moutoussis et al, 2021).

Some limitations apply to the model-based results of this study. Overall, the reliability, predictive and construct validity of MRT tasks remain to be better established, although the present study does provide evidence for the external validity of our task. In our case, the model parameters needed to be discrete because of computational limitations. While current work on easing this restriction is ongoing, we are not yet able to let parameters like inequality aversion and risk aversion or inverse temperature vary freely. Secondly, parameter recoverability is always subject to limitations, in particular in a task where partner actions strongly influence the dynamics of the interaction (i.e. responses to cooperative or uncooperative partner reveal different aspects of investment behaviour respectively, but only 1 dyad is available per subject). Changes in task design that may address some of these issues in the future include providing trustees with a base income as well, ensuring that the trustee can always act non-trivially or obtaining multiple games per subject. The validity of 'risk aversion' in socio-economic exchanges needs to be characterised both in terms of test-retest but also construct validity, including its relation to other measures of risk preference. Multiple different games and multiple instances of the same game played per subject could potentially yield more robust classifications.

Limitations of the overall study design can also be addressed in future research. The cross-sectional nature of the present sample limited us to statements about population distributions rather than within-subject, developmental effects. Finally, studies that focus specifically on sex or gender and socio-economic status may be necessary to further investigate the mechanisms through which they shape decision making and risk aversion. Specifically, we speculate that the lived experience of young people influences their confidence to take risks in economic exchanges like the present one.

Neurobiologically, changes in trust behaviour have been related to neural results on cortical thickness, which can provide a venue for further research using structural MRI and resting state data (see Bellucci et al, 2018, Tamnes et al, 2018) in combination with multi round trust game data.

Furthermore, a previous study in fact relates a lack of increase in neural reward response in adolescent males with depressive symptoms in these subjects (see Morgan et al, 2013), so the relation between risk aversion in the trust game and depressive symptoms warrants further investigation.

## ACKNOWLEDGEMENTS

RJD is supported by a Wellcome Investigator Award (ref 098362/Z/12/Z). MM and all NSPN consortium authors are supported by Wellcome Strategic Award (ref 095844/7/11/Z); The Max Planck – UCL Centre for Computational Psychiatry and Ageing is a joint initiative of the Max Planck Society and UCL. MM receives support from the UCLH Biomedical Research Centre. PF is in receipt of a National Institute for Health Research (NIHR) Senior Investigator Award (NF-SI-0514-10157). PF was, in part, supported by the NIHR Collaboration for Leadership in Applied Health Research and Care (CLAHRC) North Thames at Barts Health NHS Trust, and PBJ by the NIHR ARC East of England. PRM gratefully acknowledges funding of the Wellcome Trust under a Principal Research Fellowship (ref 091188/Z/10/Z). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Andrea Reiter acknowledges support by the German Research Foundation (Deutsche Forschungsgesellschaft, DFG RE 4449/1-1).

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR AFFILIATIONS

**Andreas Hula**  [orcid.org/0000-0003-4253-5604](https://orcid.org/0000-0003-4253-5604)

Austrian Institute of Technology, Vienna, Austria

**Michael Moutoussis**  [orcid.org/0000-0002-4751-0425](https://orcid.org/0000-0002-4751-0425)

Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom

**Geert-Jan Will**  [orcid.org/0000-0003-1887-9829](https://orcid.org/0000-0003-1887-9829)

Institute of Psychology, Leiden University, Leiden, the Netherlands

**Danae Kokorikou**

Anna Freud Centre, London, United Kingdom

**Andrea M. Reiter**  [orcid.org/0000-0002-5209-3996](https://orcid.org/0000-0002-5209-3996)

Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom; Lifespan Developmental Neuroscience, Faculty of Psychology, Technische Universität Dresden, Germany; Department of Neurology, Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

**Gabriel Ziegler**  [orcid.org/0000-0001-6589-6416](https://orcid.org/0000-0001-6589-6416)

Centre for Cognitive Neurology and Dementia Research, Magdeburg, Germany; German Center for Neurodegenerative Diseases, Magdeburg, Germany

**NSPN Consortium**

NSPN Consortium, UK

**Ed Bullmore**  [orcid.org/0000-0002-8955-8283](https://orcid.org/0000-0002-8955-8283)

Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom; Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge, United Kingdom; Medical Research Council/ Wellcome Trust Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, United Kingdom; Max Planck University College London Centre for Computational Psychiatry, London, United Kingdom

**Peter B. Jones**  [orcid.org/0000-0002-0387-880X](https://orcid.org/0000-0002-0387-880X)

Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom; Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge, United Kingdom

**Ian Goodyer**  [orcid.org/0000-0001-9183-0373](https://orcid.org/0000-0001-9183-0373)

Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom; Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge, United Kingdom

**Peter Fonagy**  [orcid.org/0000-0003-0229-0091](https://orcid.org/0000-0003-0229-0091)

Anna Freud Centre, London, United Kingdom; Research Department of Clinical, Educational and Health Psychology, University College London, United Kingdom

## REFERENCES

- Amir, D., Jordan, M. R., & Rand, D. G. (2018). An uncertainty management perspective on long-run impacts of adversity: The influence of childhood socioeconomic status on risk, time, and social preferences. *Journal of Experimental Social Psychology*, 79, 217–226. DOI: <https://doi.org/10.1016/j.jesp.2018.07.014>
- Bach, D. R., Moutoussis, M., Bowler, A., & Dolan, R. J. (2020). Predictors of risky foraging behaviour in healthy young people. *Nature Human Behaviour*, 1–12. DOI: <https://doi.org/10.1038/s41562-020-0867-0>
- Baker, D. P., Eslinger, P. J., Benavides, M., Peters, E., Dieckmann, N. F., & Leon, J. (2015). The cognitive impact of the education revolution: A possible cause of the Flynn Effect on population IQ. *Intelligence*, 49, 144–158, ISSN 0160-2896. DOI: <https://doi.org/10.1016/j.intell.2015.01.003>
- Belli, S. R., Rogers, R. D., & Lau, J. Y. F. (2012). Adult and adolescent social reciprocity: Experimental data from the Trust Game. *Journal of Adolescence*, 35(5), 1341–1349. DOI: <https://doi.org/10.1016/j.adolescence.2012.05.004>
- Bellucci, G., Hahn, T., Deshpande, G., & Krueger, F. (2018). Functional connectivity of specific resting-state networks predicts trust and reciprocity in the trust game. *Cognitive, Affective, & Behavioral Neuroscience*. DOI: <https://doi.org/10.3758/s13415-018-00654-3>
- Borghans, L., Meijers, H., & Weel, B. (2013). The importance of intrinsic and extrinsic motivation for measuring IQ. *Economics of Education Review*, 34, 17–28, ISSN 0272-7757. DOI: <https://doi.org/10.1016/j.econedurev.2013.01.008>
- Breuning, S. E., & Zella, W. F. (1978). Effects of individualized incentives on norm-referenced IQ test performance of high school students in special education classes. *Journal of School Psychology*, 16(3), 220–226, ISSN 0022-4405. DOI: [https://doi.org/10.1016/0022-4405\(78\)90004-3](https://doi.org/10.1016/0022-4405(78)90004-3)
- Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. Washington, DC: American Psychological Association. DOI: <https://doi.org/10.1037/0033-2909.125.3.367>
- Camerer, C. F., Ho, T. H., & Chong, J. S. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3), 861–898. DOI: <https://doi.org/10.1162/0033553041502225>
- Charness, G., & Gneezy, U. (2012). Strong Evidence for Sex Differences in Risk Taking. *Journal of Economic Behavior & Organization*, 83(1), 50–58. DOI: <https://doi.org/10.1016/j.jebo.2011.06.007>
- Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5), 1193–1235. DOI: <https://doi.org/10.1111/1468-0262.00239>
- Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience*, 13, 636–650. DOI: <https://doi.org/10.1038/nrn3313>
- Defoe, I. N., Dubas, J. S., Figner, B., & van Aken, M. A. (2015). A meta-analysis on age differences in risky decision making: Adolescents versus children. *Psychological Bulletin*, 141(1), 48–84. DOI: <https://doi.org/10.1037/a0038088>
- Derks, J., Lee, N. C., & Krabbendam, L. (2014). Adolescent trust and trustworthiness: role of gender and social value orientation. *Journal of adolescence*, 37(8), 1379–1386. DOI: <https://doi.org/10.1016/j.adolescence.2014.09.014>
- Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Loeber, M. S. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences*, 108(19), 7716–7720. DOI: <https://doi.org/10.1073/pnas.1018601108>
- Eisenberg, N., Miller, P. A., Shell, R., McNalley, S., & Shea, C. (1991). Prosocial development in adolescence: A longitudinal study. *Developmental Psychology*, 27(5), 849–857. DOI: <https://doi.org/10.1037/0012-1649.27.5.849>
- Fehr, E., Naef, M., & Schmidt, K. M. (2006). Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments: Comment. *American Economic Review*, 96(5), 1912–1917. DOI: <https://doi.org/10.1257/aer.96.5.1912>
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3), 817–868. DOI: <https://doi.org/10.1162/003355399556151>
- Fett, A. K. J., Shergill, S. S., Gromann, P. M., Dumontheil, I., Blakemore, S. J., Yakub, F., & Krabbendam, L. (2014). Trust and social reciprocity in adolescence—a matter of perspective-taking. *Journal of adolescence*, 37(2), 175–184. DOI: <https://doi.org/10.1016/j.adolescence.2013.11.011>

- Funder, D. C., & Ozer, D. J.** (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2, 156–168. DOI: <https://doi.org/10.1177/2515245919847202>
- Gmytrasiewicz, P. J., & Doshi, P.** (2005). A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, 24, 49–79. DOI: <https://doi.org/10.1613/jair.1579>
- Gneezy, U., Leonard, K. L., & List, J. A.** (2009). Sex Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica*, 77(5), 1637–1664. DOI: <https://doi.org/10.3982/ECTA6690>
- Hula, A., Montague, P. R., & Dayan, P.** (2015). Monte Carlo Planning Method Estimates Planning Horizons during Interactive Social Exchange. *PLoS Comput Biol*, 11(6), e1004254. DOI: <https://doi.org/10.1371/journal.pcbi.1004254>
- Hula, A., Vilares, I., Lohrenz, T., Dayan, P., & Montague, P. R.** (2018). A Model of Risk and Mental State Shifts during Social Interaction. *PLoS Comput Biol*, 14(2), e1005935. DOI: <https://doi.org/10.1371/journal.pcbi.1005935>
- Kiddle, B., Inkster, B., Prabhu, G., Moutoussis, M., & Whitaker, K.** (2017). NSPN Consortium N, Jones P. The NSPN 2400 Cohort: a developmental sample supporting the Wellcome Trust Neuroscience in Psychiatry Network. DOI: <https://doi.org/10.1093/ije/dyx117>
- 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. DOI: <https://doi.org/10.1126/science.1156902>
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R.** (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308(5718), 78–83. DOI: <https://doi.org/10.1126/science.1108062>
- Koshelev, M., Lohrenz, T., Vannucci, M., & Montague, P. R.** (2010). Biosensor Approach to Psychopathology Classification. *Plos Computational Biology*, 6(10). DOI: <https://doi.org/10.1371/journal.pcbi.1000966>
- Lemmers-Jansen, I. L. J., Krabbendam, L., Veltman, D. J., & Fett, A. K. J.** (2017). Boys vs. girls: Gender differences in the neural development of trust and reciprocity depend on social context. *Developmental Cognitive Neuroscience*, 25, 235–245, ISSN 1878-9293. DOI: <https://doi.org/10.1016/j.dcn.2017.02.001>
- Luo, Y., Héту, S., Lohrenz, T., Hula, A., Dayan, P., Ramey, S. L., Sonnier-Netto, L., Lisinski, J., LaConte, S., Nolte, T., Fonagy, P., Rahmani, E., & Montague, P. R.** (2018). Early childhood investment impacts social decision-making four decades later. *Nat Commun*, 9, 4705. DOI: <https://doi.org/10.1038/s41467-018-07138-5>
- McCabe, K., Rigdon, M. L., & Smith, V.** (2003). Positive Reciprocity and Intentions in Trust Games. *J Econ Behav Organ*, 52(2), 267–275. DOI: [https://doi.org/10.1016/S0167-2681\(03\)00003-9](https://doi.org/10.1016/S0167-2681(03)00003-9)
- Mellick, W., Sharp, C., & Ernst, M.** (2019). Depressive Adolescent Girls Exhibit Atypical Social Decision-Making In An Iterative Trust Game. *Journal of Social and Clinical Psychology*, 38(3), 224–244. DOI: <https://doi.org/10.1521/jscp.2019.38.3.224>
- Morgan, J. K., Olino, T. M., McMakin, D. L., Ryan, N. D., & Forbes, E. E.** (2013). Neural response to reward as a predictor of increases in depressive symptoms in adolescence. *Neurobiology of Disease*, 52, 66–74, ISSN 0969-9961. DOI: <https://doi.org/10.1016/j.nbd.2012.03.039>
- Moutoussis, M., Garzón, B., Neufeld, S., Bach, D. R., Rigoli, F., Goodyer, I., Bullmore, E., Fonagy, P., Jones, P., Hauser, T., Romero-Garcia, R., St Clair, M., Vértes, P., Whitaker, K., Inkster, B., Prabhu, G., Ooi, C., Toseeb, U., Widmer, B., Bhatti, J., Villis, L., Alrumaithi, A., Birt, S., Bowler, A., Cleridou, K., Dadabhoy, H., Davies, E., Firkins, A., Granville, S., Harding, E., Hopkins, A., Isaacs, D., King, J., Kokorikou, D., Maurice, C., McIntosh, C., Memarzia, J., Mills, H., O'Donnell, C., Pantaleone, S., Scott, J., Fearon, P., Suckling, J., van Harmelen, A. L., Kievit, R., Guitart-Masip, M., & Dolan, R. J.** (2021) Decision-making ability, psychopathology, and brain connectivity. *Neuron*, ISSN 0896-6273. DOI: <https://doi.org/10.1016/j.neuron.2021.04.019>
- Navarro-Carrillo, G., Valor-Segura, I., & Moya, M.** (2018). Do you Trust Strangers, Close Acquaintances, and Members of Your Ingroup? Differences in Trust Based on Social Class in Spain. *Soc Indic Res*, 135, 585–597. DOI: <https://doi.org/10.1007/s11205-016-1527-7>
- Nelson, E. E., Leibenluft, E., McClure, E. B., & Pine, D. S.** (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine*, 35, 163–174. DOI: <https://doi.org/10.1017/S0033291704003915>
- Proto, E., Rustichini, A., & Sofianos, A.** (2017). Intelligence, Personality and Gains from Cooperation in Repeated Interactions. *CESifo Working Paper Series*, 6121.
- R Core Team.** (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Ray, D., King-Casas, B., Dayan, P., & Montague, P. R.** (2008). Bayesian Model of Behaviour in Economic Games. *NIPS*, 21, 1345–1353.

- Reniers, R. L., Murphy, L., Lin, A., Bartolomé, S. P., & Wood, S. J.** (2016). Risk perception and risk-taking behaviour during adolescence: the influence of personality and gender. *PLoS one*, 11(4), e0153842. DOI: <https://doi.org/10.1371/journal.pone.0153842>
- Romer, D., Reyna, V. F., & Satterthwaite, T. D.** (2017). Beyond stereotypes of adolescent risk taking: Placing the adolescent brain in developmental context. *Developmental cognitive neuroscience*, 27, 19–34. DOI: <https://doi.org/10.1016/j.dcn.2017.07.007>
- Rosati, A. G., Benjamin, N., Pieloch, K., & Warneken, F.** (2019). Economic trust in young children. *Proc. R. Soc. B*, 286. DOI: <https://doi.org/10.1098/rspb.2019.0822>
- Rutledge, R. B., Smittenaar, P., Zeidman, P., Brown, H. R., Adams, R. A., Lindenberg, U., Dayan, P., & Dolan, R. J.** (2016). Risk Taking for Potential Reward Decreases across the Lifespan. *Curr Biol*, 26(12), 1634–1639. DOI: <https://doi.org/10.1016/j.cub.2016.05.017>
- Stamos, A., Altsitsiadis, E., & Dewitte, S.** (2019). Investigating the effect of childhood socioeconomic background on interpersonal trust: Lower childhood socioeconomic status predicts lower levels of trust. *Personality and Individual Differences*, 145, 19–25. DOI: <https://doi.org/10.1016/j.paid.2019.03.011>
- Steinberg, L.** (2004). Risk taking in adolescence: what changes, and why? In *Adolescent Brain Development: Vulnerabilities and Opportunities*. Sep, 2003, New York, NY, US.
- Sutter, M., & Kocher, M. G.** (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior*, 59, 364–382. DOI: <https://doi.org/10.1016/j.geb.2006.07.006>
- Sutter, M., Zoller, C., & Glätzle-Rützler, D.** (2019). Economic behavior of children and adolescents – A first survey of experimental economics results. *European Economic Review*, 111, 98–121, ISSN 0014-2921. DOI: <https://doi.org/10.1016/j.euroecorev.2018.09.004>
- Tannes, C. K., Overbye, K., Ferschmann, L., Fjell, A. M., Walhovd, K. B., Blakemore, S. J., & Dumontheil, I.** (2018). Social Perspective Taking Is Associated With Self-Reported Prosocial Behavior and Regional Cortical Thickness Across Adolescence. *Developmental Psychology*, 54(9), 1745–1757 0012-1649/18/\$12.00. DOI: <https://doi.org/10.1037/dev0000541>
- van de Groep, S., Meuwese, R., Zanolie, K., Güroglu, B., & Crone, E. A.** (2018). Developmental Changes and Individual Differences in Trust and Reciprocity in Adolescence. *Journal Of Research On Adolescence*, 30(S1), 192–208. DOI: <https://doi.org/10.1111/jora.12459>
- van den Akker, O. R., van Assen, M. A. L. M., van Vugt, M., & Wicherts, J. M.** (2020). Sex Differences in Trust and Trustworthiness: A Meta-Analysis of the Trust Game and the Gift-Exchange Game. *psyarxiv*. DOI: <https://doi.org/10.1016/j.joep.2020.102329>
- van den Bos, W., van Dijk, E., & Crone, E. A.** (2012). Learning whom to trust in repeated social interactions: A developmental perspective. *Group Processes & Intergroup Relations*, 15(2), 243–256. DOI: <https://doi.org/10.1177/1368430211418698>
- van den Bos, W., van Dijk, E., Westenberg, M., Rombouts S. A., & Crone, E. A.** (2011). Changing brains, changing perspectives: the neurocognitive development of reciprocity. *Psychological Science*, 22(1), 60–70. DOI: <https://doi.org/10.1177/0956797610391102>
- van den Bos, W., Westenberg, M., van Dijk, E., & Crone, E. A.** (2010). Development of trust and reciprocity in adolescence. *Cognitive Development*, 25(1), 90–102. DOI: <https://doi.org/10.1016/j.cogdev.2009.07.004>
- Van der Graaff, J., Branje, S., De Wied, M., Hawk, S., Van Lier, P., & Meeus, W.** (2014). Perspective taking and empathic concern in adolescence: Gender differences in developmental changes. *Developmental Psychology*, 50(3), 881–888. DOI: <https://doi.org/10.1037/a0034325>
- van Duijvenvoorde, A. C., Peters, S., Braams, B. R., & Crone, E. A.** (2016). What motivates adolescents? Neural responses to rewards and their influence on adolescents' risk taking, learning, and cognitive control. *Neuroscience & Biobehavioral Reviews*, 70, 135–147. DOI: <https://doi.org/10.1016/j.neubiorev.2016.06.037>
- Wechsler, D.** (2008). Wechsler Abbreviated Scale of Intelligence: WASI-IV (Fourth). *NCS Pearson Inc.*
- Weiss, L. G., Saklofske, D. H., Coalson, D., & Raiford, S. E.** (2010). WASI-IV clinical use and interpretation: Scientist-practitioner perspectives. Academic Press.
- Xiang, T., Ray, D., Lohrenz, T., Dayan, P., & Montague, P. R.** (2012). Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. *PLoS Comput Biol*, 8(12), e1002841. DOI: <https://doi.org/10.1371/journal.pcbi.1002841>

**TO CITE THIS ARTICLE:**

Hula, A., Moutoussis, M., Will, G.-J., Kokorikou, D., Reiter, A. M., Ziegler, G., NSPN Consortium, Bullmore, E., Jones, P. B., Goodyer, I., Fonagy, P., Montague, P. R., & Dolan, R. J. (2021). Multi-Round Trust Game Quantifies Inter-Individual Differences in Social Exchange from Adolescence to Adulthood. *Computational Psychiatry*, 5(1), pp. 102–118. DOI: <https://doi.org/10.5334/cpsy.65>

Submitted: 24 March 2021  
 Accepted: 16 August 2021  
 Published: 14 October 2021

**COPYRIGHT:**

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Computational Psychiatry* is a peer-reviewed open access journal published by Ubiquity Press.

