Journal of Economic Psychology 49 (2015) 95–107

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

# Theory of mind, perceived intentions and reciprocal behaviour: Evidence from individuals with Autism Spectrum Disorder

Vittorio Pelligra [a,1], Andrea Isoni [b,*], Roberta Fadda [c,2], Giuseppe Doneddu [d,3]

[a] Department of Economics and Business, University of Cagliari, Italy
[b] Behavioural Science Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, UK
[c] Department of Pedagogy, Psychology, Philosophy, University of Cagliari, Italy
[d] Centre for Pervasive Developmental Disorders, Azienda Ospedaliera Brotzu, Cagliari, Italy

ARTICLE INFO

*Article history:*
Received 19 March 2013
Received in revised form 24 April 2015
Accepted 5 May 2015
Available online 14 May 2015

*JEL classification:*
C72
C91

*PsycINFO classification:*
2340
3040
3250

*Keywords:*
Social preferences
Theory of mind
Intention detection
Autism

ABSTRACT

Evidence suggests that departures from pure self-interest are due, at least partly, to individuals conditioning their behaviour on the perceived intentions of others. We present a new experiment that refines the study of intention-based other-regarding motives. Using a series of mini-ultimatum games that have been extensively studied in the literature, we compare the behaviour of normally-developing (ND) children to that of children with Autism Spectrum Disorder (ASD), who typically lack the ability to attribute intentions to the observed actions of others. We find that ND children's rejection behaviour responds systematically to changes in the set of available options, in line with previous findings. ASD children's rejections are virtually unaffected by the intentions that could be inferred from the games' strategy space. These differences are mainly driven by ASD children with low mentalising abilities.



## 1. Introduction

In the last few decades, there has been a surge of interest in the study of *social preferences*, the highly systematic departures from the assumption of *self-interest* that is the cornerstone of standard theoretical analyses of strategic interactions. This investigation has revealed that individuals are often concerned about other individual's payoffs as well as their perceived motivations and beliefs (e.g. Camerer, 2003; Cooper & Kagel, 2013).

* Corresponding author. Tel.: +44 (0)2476 150370.
*E-mail addresses:* pelligra@unica.it (V. Pelligra), a.isoni@warwick.ac.uk (A. Isoni), robfadda@unica.it (R. Fadda), iosettodoneddu@aob.it (G. Doneddu).
[1] Tel.: +39 070 6753319.
[2] Tel.: +39 070 6757502.
[3] Tel.: +39 070 539382.

http://dx.doi.org/10.1016/j.joep.2015.05.001
0167-4870/

In this article, we focus on the role of perceived intentions on *reciprocal* behaviour – many individuals' tendency to be kind in response to someone else's kindness and unkind in response to their unkindness (e.g. Dufwenberg & Kirchsteiger, 2004; Rabin, 1993).

In many situations, whether an act counts as kind or unkind ultimately depends on the actor's intentions, and very often the same outcome can generate a different reaction if it is perceived as intentional rather than unintentional. In the context of strategic games, as well as in many real world situations, intentions cannot typically be observed, but can often be inferred by means of counterfactual reasoning: what a player could have done but did not may lead to different interpretations of the same behaviour, and therefore make a big difference to the final outcome of an interaction.

Reciprocity is thought to be one of the reasons why positive but unequal offers are often rejected in the ultimatum game (UG) (Güth, Schmittberger, & Schwarze, 1982; Güth & Tietz, 1990; Oosterbeek, Sloof, & Van de Kuilen, 2004). In the UG, a *proposer* (P) is endowed with a number of points or an amount of money – usually 10 points or $10 – and has to decide how many to keep for himself and how many to give to a *responder* (R), in the understanding that his proposed distribution is implemented if R accepts, but both P and R get nothing if R rejects. A typical finding is that offers of 40% or less of the available surplus are rejected with high frequencies (e.g. Cooper & Kagel, 2013; Roth, 1995). Reciprocity can explain this behaviour if R considers unkind, and therefore rejects, any offer below a certain *fair* threshold (e.g. Dufwenberg & Kirchsteiger, 2004).[4]

In the context of the UG, the role of intentions has been studied using two main approaches. The first approach adopts Ps that cannot act intentionally. Blount (1995) found that, when Ps are computerised, Rs' minimum acceptable offers in a $10 UG are significantly lower ($1.20 on average) than when the Ps are human ($2.91 on average). The second approach consists in restricting P's strategy space by either constraining his maximum offer (e.g. Nelson, 2002) or devising a multiplicity of *mini*-ultimatum games (Bolton & Zwick, 1995) in which only two splits are available, with one kept constant and the other varied (e.g. Brandts & Solà, 2001; Falk, Fehr, & Fischbacher, 2003). Nelson (2002) finds that an offer of $4 in a $20 ultimatum game is often rejected when the proposer could offer any amount up to $20, but very likely to be accepted when offers are capped at $4. Falk et al. (2003) – henceforth FFF – study a series of four mini-UGs and find that the same relatively unequal split is rejected significantly more frequently when the only available alternative is equal than when it is even more unequal. Brandts and Solà (2001) report similar findings. More recently, Güney and Newell (2013) combine these two approaches using FFF's design with (human) Rs facing either human or computerised Ps, and find differences in line with intention-driven reciprocal behaviour.

In this paper we propose a complementary approach. Instead of manipulating features of the game pertaining to proposers, we focus on the responder's side. Using FFF's experimental paradigm, we investigate whether the pattern of rejections for the same, and relatively unequal, split varies between Rs with potentially different abilities to attribute intentions to Ps' behaviour. We achieve this by studying a sample including children with Autism Spectrum Disorder (ASD), who are known to be impaired in their *theory of mind* – henceforth ToM – the ability to attribute intentions and mental states to others (Baron-Cohen, Leslie, & Frith, 1985; Baron-Cohen, Tager-Flusberg, & Cohen, 1993, 2000). We compare the behaviour of ASD Rs with the behaviour of normally-developing (ND) Rs. Our conjecture is that, if the patterns of rejections in FFF's mini-UG games are intention-driven, ASD Rs should not display the same responsiveness to the inequality of other available alternative offers as ND Rs. We also measure our participants' ToM abilities using standard tests to explore to what extent any behavioural differences between ASD and ND participants is mediated by differences in ToM.

Our study is closely related to earlier work by Sally and Hill (2006), who compared the behaviour of ASD and ND children of various age groups in a series of repeated Prisoners' Dilemma, Dictator and Ultimatum games. In relation to the UG, they report that Ps affected by ASD were likely to follow one of two strategies, either offering half of the points or keeping all points for themselves, while ASD Rs were much more likely to accept low offers, as well as reject offers of more than half of the pie, than their ND counterparts. Sally and Hill attribute the variation in Rs behaviour to ASD's 'relative abilities in discerning the intentions of the offerer' (p. 93). They report that the behaviour of both their ND and ASD participants is related to their ToM, with Ps's offers being larger for individuals who passed their ToM test than for those who failed it. They speculate that 'the denial of an inapt or inequitable offer, and the acceptance of a fair proposal may depend on the bargainer's theory-of-mind' (p. 94). More recently, Takagishi, Kameshima, Schug, Koizumi, and Yamagishi (2010) have also emphasised the importance of mentalising abilities in UG bargaining. They studied a sample of 68 pre-schoolers in a face-to-face UG setting and found that differences in Ps behaviour were better explained by differences in ToM than differences in age.

Our investigation extends this emerging line of research by comparing the behaviour of ASD and ND Rs in a series of one-shot mini-UGs, and exploring the role of differences in ToM. We find significant differences in rejection behaviour between ASD and ND participants. While our control ND Rs show the typical pattern of rejections consistent with reciprocal behaviour (Falk et al., 2003; Harbaugh, Krause, & Liday, 2003; Sutter, 2007), our ASD Rs react little to differences in the varying alternative allocation. When we split our full sample according to their performance in our ToM test, we find that Rs who pass the test show patterns of behaviour consistent with reciprocity, while Rs who fail the test are much less responsive to differences in the alternative split that could have been chosen but was not. However, this difference is mostly driven by participants with ASD, as ND Rs who fail our ToM test still show the typical pattern of rejections. This split-sample analysis

[4] *Inequality aversion* (e.g. Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) is another popular explanation for rejection by Rs in the UG. We consider its implications for our specific setting in Section 2.

suggests that differences in rejection rates between ND and ASD participants are not mediated by differences in ToM, but may be due to some other characteristic specific to ASD participants which correlates with their ToM. We speculate that the relative irresponsiveness of Rs to changes in the alternative split in our mini-UGs might be related to ASD's *executive functioning* deficits (e.g. Ozonoff, Pennington, & Rogers, 1991).

The remainder of the paper is organised as follows. In Section 2, we present the simple games that have been used by FFF to study intention-based motives, and describe our tweak on their design. We give more details on our experimental procedures in Section 3, and present our results in Section 4. In Section 5, we discuss our findings and offer some concluding remarks.

## 2. Experimental design

Our experiment is based on the design of FFF, who study the four games shown in Fig. 1.

In each of the four games, P can choose between two allocations of 10 experimental points between himself and R. Let [*p*, *r*] denote an *allocation* in which P gets *p* points and R gets *r* points. All four games in Fig. 1 feature the same [8, 2] allocation, but vary in terms of the other allocation that is available to P, which we denote as $[10 - x, x]$, where *x* is P's *offer* to R in that particular allocation. In FFF's games, *x* takes values of 5, 8, 2 and 0. As standard in UGs, P's proposed allocation is implemented if R accepts, otherwise both P and R get nothing.

Our main interest is in the behaviour of Rs with respect to the [8, 2] allocation. The standard prediction is that a self-interested R will always accept this allocation, on the grounds that 2 is better than nothing. If R has other-regarding preferences of some kind, however, she may reject P's offer of 2 points.

One first reason for rejecting the [8, 2] allocation is *inequality aversion* (e.g. Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). If that allocation were implemented, R would be in a situation of disadvantageous inequality: if she dislikes this kind of inequality strongly enough, then she might prefer the equal, but inefficient, [0, 0] allocation. With respect to inequality aversion, FFF's games have two interesting properties. First, since the effect of inequality aversion is a function of the degree of inequality of the allocation that is being evaluated, whether there are other, more or less unequal, possible allocations is irrelevant. Therefore, if inequality aversion is the only cause for the rejection of the [8, 2] allocation, the rate at which that allocation is rejected should be the same across the four games. Second, since in the [8, 2]-UG the two allocations available to P are the same, the rate at which that allocation is rejected in that game provides a baseline estimate for the level of inequality aversion in a particular sample.

The rejection of the [8, 2] allocation can also be driven by *reciprocal* considerations. Various reciprocity models have been proposed in the literature (e.g. Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Rabin, 1993), which define reciprocity and model its implications in different ways. A detailed discussion of those models and whether or not they are compatible with FFF's findings goes beyond the scope of this article. For the purposes of our analysis of R's behaviour, we will limit ourselves to say that she acts reciprocally if she rejects the [8, 2] allocation because she thinks that P has been *unkind* to her by making that offer (*negative* reciprocity), or if she accepts it because she thinks that P has been *kind* (*positive* reciprocity).

Given the simple structure of FFF's mini-UGs, in most cases relatively unambiguous inferences about whether R may perceive P as being kind or unkind to her by choosing the [8, 2] split can be made by looking at the $[10 - x, x]$ allocation.

In the [5, 5]-UG, the [8, 2] allocation looks bad in comparison to the [5, 5] allocation, because P could have chosen an equal split, but did not. When R receives an offer of 2 in the [5, 5]-UG, she can attribute a negative, unkind, intention to P, and be more inclined to reject the [8, 2] allocation than she would be in a game such as the [8, 2]-UG, in which P has no role in determining the inequality of the proposed allocation.

Analogously, when the other alternative is [2, 8], R can see that P faces the choice between some degree of inequality at his advantage and the same degree of inequality at his disadvantage. Offering 2 when offering 8 was possible can still be interpreted as unkind, but less so than in the [5, 5]-UG, as being kind would have come at greater cost. Inferring this kind of intentions leads to expect a rejection rate between those of the [5, 5]-UG and [8, 2]-UG.

Finally, when the alternative allocation is [10, 0], an offer of 2 can be interpreted as a kind gesture, and therefore reduce the rejection rate for the [8, 2] allocation relative to the baseline [8, 2]-UG. While interpreting a rejection of the [8, 2] allocation as an effect of negative reciprocity is relatively uncontroversial in the [5, 5]-UG and [8, 2]-UG, because R sacrifices 2 points in order to punish P for his unkind behaviour, the situation is less clear-cut in the [10, 0]-UG. Since both reciprocal concerns and material payoff maximisation tend to offset inequality aversion and increase the likelihood that an offer of 2 is accepted, it is not possible to attribute a lower rejection rate in the [10, 0]-UG *exclusively* to positive reciprocity.

These qualitative predictions were confirmed in FFF's experiment. The [8, 2] allocation was rejected by 44.4%, 26.7%, 18% and 8.9% of participants in the [5, 5]-, [2, 8]-, [8, 2]-, and [10, 0]-UG respectively. Sutter (2007) reports a similarly declining (though slightly flatter) pattern for the same set of games played by ND children.

We extend FFF's design by comparing the behaviour of Rs who are able to detect intentions and Rs who are not. This sets our approach apart from the other two that have been used in the literature, and provides a new and further angle on the study of intentional behaviour in the UG. We will test the following:

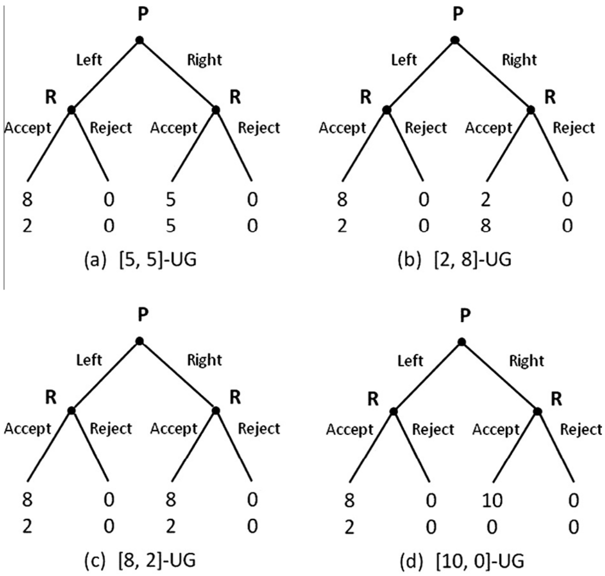


**Fig. 1.** FFF's mini-ultimatum games.

**Prediction:** *If the rejection of the [8, 2] allocation is intention-driven, individuals who are able to detect intentions should behave according to the pattern observed by FFF, while individuals who are unable to detect intentions should behave in the same way in the four mini-UGs.*

In order to control for Rs' ability to detect intentions, we use a sample comprising both ND and ASD children. We conduct our analysis at two levels. We first compare the patterns of rejection rates for the [8, 2] allocation between ND and ASD. We then conduct the same analysis by comparing children who pass the ToM test (the *High ToM* group) with those who do not (the *Low ToM* group) and investigate whether any differences in behaviour between ND and ASD children are mediated by differences in ToM.

## 3. Experimental procedures

We recruited a total of 60 ND children and 20 ASD children. The ND children were on average ten years old, were recruited from fifth grade in six primary schools, and had no certified psychiatric or developmental disorder. ASD children were contacted through the Centre for Pervasive Developmental Disorders of the "Brotzu" Hospital in Cagliari (Italy). Inclusion in the ASD group required a previous DSM-IV diagnosis of an ASD made by a licensed clinician experienced in the assessment and diagnosis of autism, a diagnosis based on the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999). The ASD participants' average age at the time of the experiment was about 12 years and a half, and their average IQ – 87.55 (s.d. 19.04) – made their average cognitive functioning comparable to that of a ten-year old ND child.[5] The participants were all males, considering the higher prevalence of ASD in this gender. The experimental sessions took place in Cagliari between May and November 2011.[6]

Each individual in our sample took part in the four mini-UG described in Section 2, plus another six games not reported here.[7] The four mini-UGs always came one after the other, in randomised order. Whether the [8, 2] option was shown to the left or to the right was counterbalanced across subjects. At the end of the experiment, once all games had been played, one was selected at random to determine the subject's earnings.

Each individual played the four mini-UGs in the same role, either as P, or as R. Since our main interest is on Rs' responses to the [8, 2] allocation for different $[10 - x, x]$ alternatives, all 20 ASD subjects were assigned the role of R. Of the 60 ND children, 40 played as P and 20 played as R. This ensures that we have 40 Ps and 40 Rs, which is essential for the implementation of our reward mechanism (see below).

Although our main interest is in the behaviour of Rs, the decisions of Ps allow us to carry out an important consistency check in our data. The decisions of Ps in FFF's data show that the percentage of individuals choosing the [8, 2] allocation in

[5] Because our ASD children attended the hospital on a regular basis, we had access to their recent IQ assessment made by a licenced clinician, which typically takes more than an hour to administer. For ND children, we did not have the possibility to take equivalent IQ measures because the experiment was conducted during school hours. Therefore, we recruited ND children with no certified psychiatric or developmental disorder of an age for which the typical IQ values are in line with those of our ASD children.

[6] Before the experiment was conducted, we received the written consent of the parents of both ND and ASD children.

[7] These included two trust games, the hide and seek game and three coordination games. The results of these games are not reported here.

the [5, 5]-UG is roughly 30%, rising to about 70% in the [2, 8]-UG, and 100% in the [10, 0]-UG (the data of the [8, 2]-UG cannot be unambiguously interpreted). A similar pattern is observed by Sutter (2007) with children. If our sample conforms to the findings reported in the literature with respect to the pattern of [8, 2] choices, we can be reassured that the test we carry out on the behaviour of Rs is picking up the effects of interest and is not the result of other unobservable features of our sample.

For players in the R role, we used the *strategy method*. We asked them to decide whether they would accept or reject each of the possible allocations that P could choose between (in each game, [8, 2] and $[10 - x, x]$ respectively). Relative to sequential play, the strategy method has the advantage of producing a larger number of observations, which is of particular importance for us given the rarity of ASD subjects. It also simplifies the implementation of our incentive scheme. We acknowledge that its use may somewhat weaken the effects of reciprocal behaviour that would be possible in a sequential implementation, but we judged that the benefits in terms of sample size were worth this risk.[8] Since our test requires a comparison of children who can and cannot attribute intentions, and the strategy method is kept constant across these groups, its use should not affect our qualitative conclusions.

Given the age of our participants, we did not reward them using money, but with trading cards (Yu-Ghi-Oh and Pokémon, see the Appendix for an example of one of the game displays). Each P was matched with an R. Since the logistical arrangements of the experimental trials involving children affected by ASD required that only one subject played the games at any given time (these trials took place at the hospital during one of the children's routine appointments), and since the Rs are the subjects we are mostly interested in for our tests, we adopted a procedure intended to ensure that the salience of the incentives was maximised for these subjects. We first collected the data for all Ps in all games. Subsequently, we collected the data for all Rs. These subjects knew that the child they were playing with had already made their decisions. They selected which game was to be played for real soon after completing the experiment, and were paid straight away on the basis of their own decision and the decision made by a randomly assigned P, which was communicated via a computer terminal. The payment of the Ps was delayed until the decisions of the Rs were collected and processed. In order to justify this delay, these subjects were (truthfully) told that the child they were playing with was from a different school, so their payment would occur on a different day. They could not tell from the text of the instructions exactly when the person they had been matched with would take part in the experiment (see Appendix for details).

All children made their decisions in the presence of an experimenter, were shown a sample game before the beginning of the experiment, and could ask clarification questions. It was common knowledge that the other players were also children. ASD children were accompanied by their personal tutor who helped ensure they understood the rules of the games.

In order to assess the participants' ToM competence, we use the 'ice cream van' version of the second-order false belief test (Perner & Wimmer, 1985).[9] This test is based on a story in which two children, John and Mary, are together at the park when the ice cream van arrives. John would like to have an ice cream, but since he has left his money at home he goes back to get it after being reassured by the ice cream man that he will be at the park all afternoon. While John is away, however, the ice cream man leaves the park to go to sell ice cream outside the church. John sees him and discovers where he is going. Later, Mary goes to John's house to look for him, but she is told that he has gone to buy ice cream. The second-order false belief question is 'where does Mary think that John has gone to buy ice-cream?' The correct answer is 'park', because Mary does not know that John has met the ice cream man while he was moving, so she thinks that John thinks that the ice cream van is still in the park (see the Appendix for details).

ND children typically pass this test by the age of six (Hayashi, 2007; Perner & Wimmer, 1985), but it is not uncommon for them to fail it, especially in particular cultures (Lecce & Hughes, 2010).[10] ASD children are much more likely to fail (Baron-Cohen, 1989; Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). The test was individually administered to all participants before they played the games.

The second-order false belief test fits well with our study also because of our use of the strategy method. In general, the attribution of intentions by second movers in sequential games requires that the individual is able to engage forms of

[8] The strategy method was also employed by FFF and Sutter (2007), who find clear evidence of reciprocal behaviour. Recent evidence shows that, while in public good games the use of the strategy method seems not to be problematic (Fischbacher, Quercia, & Gächter, 2012), it can significantly lower observed rates of trustworthiness (Casari & Cason, 2009).

[9] Other studies, including Sally and Hill (2006), use a different version of the test, the so-called 'birthday puppy story' (Sullivan, Zaitchik, & Tager-Flusberg, 1994), which is a simplified version of the ''ice-cream van'' developed by Perner and Wimmer (1985). The Sullivan et al. (1994) task is characterised by a deceptive context, and has been shown to elicit better performance in ND children, thanks to a simplified procedure (Hayashi, 2007) and an additional question about the content of the belief (Miller, 2012). However, its wording is more complex than Perner and Wimmer's, and the beneficial effects of deception on second-order false belief understanding are not so clear (Miller, 2012). Given that individuals with ASD are known to have difficulties with deception (Russell, Mauthner, Sharpe, & Tidswell, 1991) and understanding of verbal language (Volkmar, Paul, Klin, & Cohen, 2005), Perner and Wimmer's task should reduce the possible interfering effects of these variables. In our study, we used coloured vignettes representing the most salient parts of the story to aid the participants' understanding.

[10] Performance in first-order belief tasks such as the Sally-Anne test (i.e. Baron-Cohen et al., 1985; Takagishi et al., 2010) is usually at ceiling in individuals over five years old (Wellman, Cross, & Watson, 2001). ToM abilities in school children are better tested with second-order belief tasks, aimed to probe recursive ToM understanding: inferences about a belief about a belief (Liddle & Nettle, 2006). For this reason, given the age bracket of our participants, we employed a second-order false belief test.

counterfactual reasoning analogous to those required to successfully complete the test. When a second mover (B) observes a strategy choice by a first mover (A), she might infer that A *intended* to be kind to her. But for this to be the case, B must be able to anticipate that A thinks that his action will be perceived by B herself as kind (a second-order belief). This is even more important when the strategy method is used, because of the conditional nature of the question that is being asked.[11]

## 4. Results

Before discussing our main findings we present the results of the second-order false belief test. These are shown in Table 1.

While more than half of ND participants pass the test (52.5% of Ps and 60% of Rs), only 25% of ASD participants do so. This is in line with previous findings.[12]

### 4.1. Proposers' decisions

We start by looking at the behaviour of Ps. As explained in the previous Section, we can compare the offers of these ND children to those reported in the literature as a way of checking the appropriateness of our sample. The relevant data are reported in panel A of Table 2.

The proportion of subjects offering the [8, 2] split rather than the [10 − $x$, $x$] alternative is 7.5% in the [5, 5]-UG, 60% in the [2, 8]-UG, and 100% in the [10, 0]-UG. We use the Cochran's $Q$ test to test the null hypothesis that the proportion of participants offering the [8, 2] split is the same in the three games. The last column of panel A shows that this hypothesis can confidently be rejected. As indicated by the McNemar tests reported in panel B of Table 2, the rejection rate in the [5, 5]-UG differs significantly from those of the other two games. With the possible exception of the [5, 5]-UG, these figures are very similar to the results reported by FFF for adults and by Sutter (2007) for children. We take this as an indication of the appropriateness of our sample for the purposes of our study.

Panels C and D of Table 2 report the breakdown of responses for High ToM and Low ToM Ps respectively. The patterns of [8, 2] offers do not differ between the two subgroups, suggesting that low ToM abilities do not prevent ND Ps from anticipating that unfair offers are likely to be rejected when fairer alternatives are available.

### 4.2. ND vs. ASD responders

We now turn to the behaviour of Rs. For each of the four games, Table 3 reports the number of Rs rejecting the [8, 2] allocation and the alternative [10 − $x$, $x$] allocation, with the corresponding proportions, separately for ND participants (panel A) and ASD participants (panel C). The last column of each of these panels reports the test (based on Cochran's $Q$) of the null hypothesis that rejection rates are the same in all games, against the alternative that they differ in at least one game. Panels B (for ND Rs) and D (for ASD Rs) report pairwise comparisons of rejection rates between games based on the McNemar test, under the main diagonal for the rejection of the [8, 2] allocation, and above it for the rejection of the [10 − $x$, $x$] allocation.

Panel A shows that ND Rs display the typical pattern of rejections that has been attributed to reciprocal behaviour. The [8, 2] allocation is rejected by 75% of Rs in the [5, 5]-UG, 45% in the [2, 8]-UG, 20% in the [8, 2]-UG and 5% in the [10, 0]-UG.[13] The Cochran's test shows that these rejection rates are significantly different overall, with most pairwise comparisons between rejection rates registering as highly significant (see panel B).[14]

On the other hand, the behaviour of ASD Rs is rather different. The [8, 2] offer is rejected by 60% of Rs in the [5, 5]-UG, 60% in the [2, 8]-UG, 65% in the [8, 2]-UG and 45% in the [10, 0]-UG. These rejection rates are not significantly different from each other overall (see Cochran's test in panel C) nor in any of the pairwise comparisons.[15]

The stark difference between the patterns of rejection rates in the two groups is depicted in Fig. 2.

In line with our prediction, the curve for ND Rs is clearly downward sloping, while that for ASD Rs is virtually flat. This indicates that, relative to ND participants, ASD participants' rejection of the [8, 2] allocation varies much less across the four games.

[11] Hedden and Zhang (2002) suggest that these kinds of abilities are a more general pre-requisite for the play of most forms of strategic games, including simultaneous-move games.

[12] Some readers may find it surprising that almost half of the ND participants fail the second-order false belief test. However, there seems to be a substantial variation in test results across ages and cultures. In a recent comparison of 70 British and 70 Italian 5–6 year-old children, Lecce and Hughes (2010) found that significantly more Italian children failed the second-order false belief test (50 vs. 30 in the British sample). The Italian ND children that we study are older and perform better, but their ToM – as measured by our test – does not appear to be fully developed in the whole sample.

[13] The absolute values of these rejection rates are in some cases rather different from FFF's, but it has to be remembered that there are several differences between our design and theirs that can be responsible for this (e.g. the age of the participants, the use of monetary rewards in FFF's experiment and non-monetary rewards in ours, possible cultural factors, etc.).

[14] The [10 − $x$, $x$] allocation is rejected by 5% of responders in the [5, 5]-UG, nobody in the [2, 8]-UG and everybody in the [10, 0]-UG (see the second row of panel A). These rejection rates are significantly different overall on a Cochran's $Q$ test, because of the large difference between the [10, 0]-UG and the other two (see panel B above main diagonal).

[15] The pattern of rejections of the [10 − $x$, $x$] allocation by ASD Rs is analogous to that of ND Rs (see second row of panel C). Three observations are missing.

**Table 1**
Results of the second-order false belief test.

| ND Proposers | | ND Responders | | ASD Responders | |
|---|---|---|---|---|---|
| Pass | Fail | Pass | Fail | Pass | Fail |
| 21 | 19 | 12 | 8 | 5 | 15 |
| 52.5% | 47.5% | 60.0% | 40.0% | 25.0% | 75.0% |

**Table 2**
Proposers' decisions.

| *(A) All Proposers (N = 40):* | [5, 5]-UG | | [2, 8]-UG | | [10, 0]-UG | | Sig.[a] |
|---|---|---|---|---|---|---|---|
| | Obs. | % | Obs. | % | Obs. | % | |
| Offers [8, 2] | 3 | 7.5 | 24 | 60.0 | 40 | 100.0 | *** |

| *(B) Pairwise statistical tests[b]:* | [5, 5]-UG | | [2, 8]-UG | |
|---|---|---|---|---|
| | $\chi^2$ | Sig. | $\chi^2$ | Sig. |
| [2, 8]-UG | 17.64 | *** | | |
| [10, 0]-UG | 37.00 | *** | 16.00 | *** |

| *(C) High ToM Proposers (N = 21):* | [5, 5]-UG | | [2, 8]-UG | | [10, 0]-UG | | Sig.[a] |
|---|---|---|---|---|---|---|---|
| | Obs. | % | Obs. | % | Obs. | % | |
| Offers [8, 2] | 2 | 9.5 | 14 | 66.7 | 21 | 100.0 | *** |

| *(D) Low ToM Proposers (N = 19):* | [5, 5]-UG | | [2, 8]-UG | | [10, 0]-UG | | Sig.[a] |
|---|---|---|---|---|---|---|---|
| | Obs. | % | Obs. | % | Obs. | % | |
| Offers [8, 2] | 1 | 5.3 | 10 | 52.6 | 19 | 100.0 | *** |

a – Significance level based on Cochran's Q test: *** = 1%; ** = 5%; * = 10%.
b – Significance levels in McNemar test: *** = 1%; ** = 5%; * = 10%.

**Table 3**
Responders' decisions (ND vs. ASD).

| *(A) ND (N = 20):* | [5, 5]-UG | | [2, 8]-UG | | [8, 2]-UG | | [10, 0]-UG | | Sig.[a] |
|---|---|---|---|---|---|---|---|---|---|
| | Obs. | % | Obs. | % | Obs. | % | Obs. | % | |
| Rejects [8, 2] | 15 | 75.0 | 9 | 45.0 | 4 | 20.0 | 1 | 5.0 | *** |
| Rejects $[10 - x, x]$ | 1 | 5.0 | 0 | 0.0 | – | – | 20 | 100.0 | *** |

| *(B) Pairwise statistical tests (ND)[b]:* | [5, 5]-UG | | [2, 8]-UG | | [8, 2]-UG | | [10, 0]-UG | |
|---|---|---|---|---|---|---|---|---|
| | $\chi^2$ | Sig. | $\chi^2$ | Sig. | $\chi^2$ | Sig. | $\chi^2$ | Sig. |
| [5, 5]-UG | – | – | *1.00* | | – | – | *19.00* | *** |
| [2, 8]-UG | 2.67 | | – | – | – | – | *20.00* | *** |
| [8, 2]-UG | 11.00 | *** | 7.00 | *** | – | – | – | – |
| [10, 0]-UG | 14.00 | *** | 10.00 | *** | 3.00 | * | – | – |

| *(C) ASD (N = 20):* | [5, 5]-UG | | [2, 8]-UG | | [8, 2]-UG | | [10, 0]-UG | | Sig.[a] |
|---|---|---|---|---|---|---|---|---|---|
| | Obs. | % | Obs. | % | Obs. | % | Obs. | % | |
| Rejects [8, 2] | 12 | 60.0 | 12 | 60.0 | 13 | 65.0 | 9 | 45.0 | |
| Rejects $[10 - x, x]$ | 0 | 0.0 | 2 | 11.8 | – | – | 16 | 94.1 | *** |

| *(D) Pairwise statistical tests (ASD)[b]:* | [5, 5]-UG | | [2, 8]-UG | | [8, 2]-UG | | [10, 0]-UG | |
|---|---|---|---|---|---|---|---|---|
| | $\chi^2$ | Sig. | $\chi^2$ | Sig. | $\chi^2$ | Sig. | $\chi^2$ | Sig. |
| [5, 5]-UG | – | – | *2.00* | | – | – | *16.00* | *** |
| [2, 8]-UG | 0.00 | | – | – | – | – | *14.00* | *** |
| [8, 2]-UG | 0.14 | | 0.14 | | – | – | – | – |
| [10, 0]-UG | 1.29 | | 1.29 | | 2.67 | | – | – |

a – Significance level based on Cochran's Q test: *** = 1%; ** = 5%; * = 10%.
b – Significance levels in McNemar test: *** = 1%; ** = 5%; * = 10%. Tests for [8, 2] decisions under main diagonal, tests for $[10 - x, x]$ decisions above main diagonal.

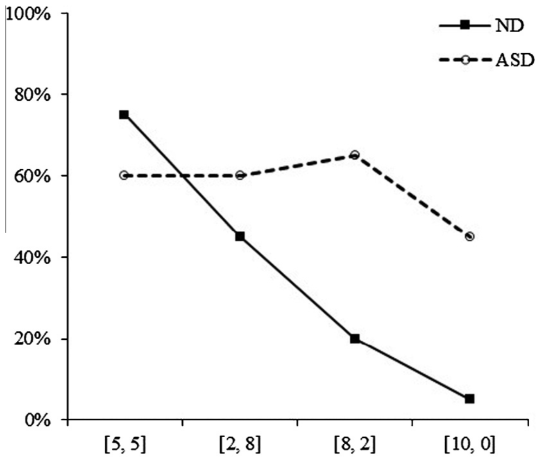


**Fig. 2.** Rejection rates in the four mini-ultimatum games for ND and ASD participants.

To further explore the difference between the two groups, we ran a random effects logit regression of the probability of rejecting the [8, 2] allocation on a constant, a dummy variable for the ASD group and two interaction terms between ASD and game (ASD $\times$ Game) and ND and game (ND $\times$ Game).[16] The constant captures the baseline tendency to reject in the ND, while the ASD dummy tells us whether the propensity to reject is different in this group. The remaining two variables are meant to reflect the slopes of the two curves in Fig. 2. Since in logit models the estimated coefficients cannot be interpreted directly as marginal effects, we will mainly be concerned with their sign and significance. This analysis allows us to carry out a more direct comparison of the two groups. The results are shown in Table 4.

The first thing to notice is that the ASD dummy is negative and significant. This means that, other things being equal, ASD individuals are less likely to reject the [8, 2] offer than ND individuals. The two interaction terms are also negative, in line with the patterns we observe in Fig. 2. However, while the interaction term is large and strongly significant for the ND group, it is much smaller in size and not significant in the ASD group. The hypothesis that the coefficients of the two interaction terms are equal can be confidently rejected ($p. < 0.01$).

### 4.3. High ToM vs. Low ToM responders

So far, we have highlighted a clear difference in behaviour between the ND and ASD Rs consistent with the hypothesis that reciprocal behaviour is driven by intention detection. In this subsection we try to understand the mechanism behind this behaviour by contrasting the behaviour of Rs who passed the ToM test and those who did not.

Table 5 looks at how rejection rates for the [8, 2] split vary with performance in the ToM test. In panel A, we pool ASD and ND participants, while in panel B we report subgroups obtained by dividing our sample by ToM *and* diagnosis. The data in panel A are plotted in Fig. 3 below.

This picture shows a marked responsiveness of rejection rates to changes in the $[10 - x, x]$ alternative for High ToM participants, but only a weak responsiveness for Low ToM participants. Table 5 shows that [8, 2] rejection rates vary significantly in the former group, but not in the latter.

In order to understand whether ToM differences may be mediating the differences we observed between our ND and ASD Rs' rejection patterns, we can take a look at the numbers in panel B of Table 5, which split our sample in four subgroups: 12 High ToM ND participants, 5 High ToM ASD, 8 Low ToM ND and 15 Low ToM ASD. Unfortunately, these numbers are too small for us to conduct meaningful statistical tests at the level of the individual subsample (hence the '–' in the last column of panel B), or to derive definitive conclusions. However, the pattern that seems to emerge is that, with the exception of the Low ToM ASD participants, all groups show a substantial degree of responsiveness to changes in the alternative offer in the direction usually observed with this design.

Table 6 reports the results of a random effects logit regression of [8, 2] rejections over two- and three-way interaction terms and a constant. The constant can be thought of as representing the underlying overall tendency to reject in the whole sample. The two-way interaction terms reflect differences in average rejection rates between the ND High ToM participants and the other three subgroups, and show that the ASD Low ToM participants are relatively less prone to reject overall ($p. < 0.05$; the other two terms are not significant). The four three-way interaction terms indicate directional changes in rejection rates across the four mini-UGs. Except for the ASD $\times$ LowToM $\times$ Game interaction, which is not statistically significant, all other coefficients are significantly negative ($p. < 0.05$ in three cases, $p. < 0.01$ in two), indicating a significant decrease of rejection rates in the usual direction. Moreover, while the coefficients for the two three-way interaction terms for ASD participants are significantly different from each other ($p. < 0.05$), the null hypothesis that the two coefficients

[16] Here 'Game' takes value 1 for the [5, 5]-UG, 2 for the [2, 8]-UG, 3 for the [8, 2]-UG and 4 for the [10, 0]-UG.

**Table 4**
Random Effects Logit Estimates (ND vs. ASD).

| | Coeff. | Std. Dev. | *Z* | *p*. |
|---|---|---|---|---|
| Constant | 4.441 | 1.320 | 3.36 | 0.001 |
| ASD | −3.220 | 1.516 | −2.12 | 0.034 |
| ASD × Game | −0.268 | 0.262 | −1.02 | 0.308 |
| ND × Game | −2.230 | 0.560 | −3.99 | 0.000 |

*Note*: ASD × Game = ND × Game: $p = 0.0011$.

**Table 5**
Effect of ToM on [8, 2] rejections.

| | [5, 5]-UG | | [2, 8]-UG | | [8, 2]-UG | | [10, 0]-UG | | Sig.[a] |
|---|---|---|---|---|---|---|---|---|---|
| | Obs. | % | Obs. | % | Obs. | % | Obs. | % | |
| *(A) [8, 2] rejections by ToM* | | | | | | | | | |
| High ToM ($N = 17$) | 13 | 76.5 | 11 | 64.7 | 5 | 29.4 | 1 | 5.9 | *** |
| Low ToM ($N = 23$) | 14 | 60.9 | 12 | 52.2 | 12 | 52.2 | 9 | 39.1 | |
| *(B) [8, 2] rejections by ToM and diagnosis* | | | | | | | | | |
| High ToM ND ($N = 12$) | 9 | 75.0 | 6 | 50.0 | 2 | 16.7 | 0 | 0.0 | – |
| High Tom ASD ($N = 5$) | 4 | 80.0 | 5 | 100.0 | 3 | 60.0 | 1 | 20.0 | – |
| Low ToM ND ($N = 8$) | 6 | 75.0 | 5 | 62.5 | 2 | 25.0 | 1 | 12.5 | – |
| Low Tom ASD ($N = 15$) | 8 | 53.3 | 7 | 46.7 | 10 | 66.7 | 8 | 53.3 | – |

a – Significance level based on Cochran's *Q* test: *** = 1%; ** = 5%; * = 10%.

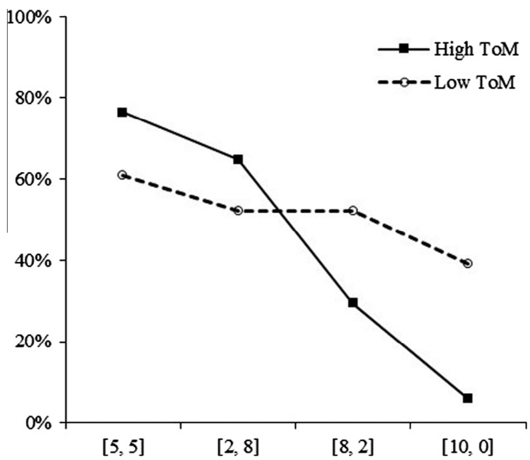


**Fig. 3.** Rejection rates in the four mini-ultimatum games for High ToM and Low ToM participants.

**Table 6**
Random Effects Logit Estimates of diagnosis and ToM interactions.

| | Coeff. | Std. Dev. | *Z* | *p*. |
|---|---|---|---|---|
| Constant | 5.073 | 1.823 | 2.78 | 0.005 |
| ASD × LowToM | −5.004 | 2.078 | −2.41 | 0.016 |
| ASD × HighToM | 0.486 | 2.954 | 0.16 | 0.869 |
| ND × LowToM | −0.880 | 2.363 | −0.37 | 0.710 |
| ASD × LowToM × Game | 0.144 | 0.311 | 0.46 | 0.643 |
| ASD × HighToM × Game | −1.703 | 0.750 | −2.27 | 0.023 |
| ND × LowToM × Game | −1.897 | 0.677 | −2.80 | 0.005 |
| ND × HighToM × Game | −2.719 | 0.820 | −3.32 | 0.001 |

ASD × LowToM × Game = ASD × HighToM × Game: *p*. = 0.0234.
ND × LowToM × Game = ND × HighToM × Game: *p*. = 0.3770.

**Table 7**
[8, 2] Rejection rates for different game positions.

| | Position in sequence | | | | Fisher's exact test $p$-value |
|---|---|---|---|---|---|
| | 1st | 2nd | 3rd | 4th | |
| *All responders*[a] ($N$ = 37) | | | | | |
| [5, 5]-UG | 63.6% | 72.7% | 62.5% | 71.4% | 1.000 |
| [2, 8]-UG | 50.0% | 66.7% | 57.1% | 37.5% | 0.701 |
| [8, 2]-UG | 57.1% | 50.0% | 23.1% | 44.4% | 0.444 |
| [10, 0]-UG | 33.3% | 28.6% | 22.2% | 8.3% | 0.557 |

[a] Three observations are missing.

**Table 8**
[8, 2] Rejections at the individual level.

| | Number of [8, 2] rejections | | | | | | | | | | Mean | Std. Dev. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | | 1 | | 2 | | 3 | | 4 | | | |
| | Obs. | % | Obs. | % | Obs. | % | Obs. | % | Obs. | % | | |
| All observations ($N$ = 40) | 6 | 15.0 | 11 | 27.5 | 10 | 25.0 | 6 | 15.0 | 7 | 17.5 | 1.93 | 1.33 |
| ND High ToM ($N$ = 12) | 3 | 25.0 | 3 | 25.0 | 4 | 33.3 | 2 | 16.7 | 0 | 0.0 | 1.42 | 1.08 |
| ND Low ToM ($N$ = 8) | 1 | 12.5 | 3 | 37.5 | 2 | 25.0 | 1 | 12.5 | 1 | 12.5 | 1.75 | 1.28 |
| ASD High ToM ($N$ = 5) | 0 | 0.0 | 0 | 0.0 | 3 | 60.0 | 1 | 20.0 | 1 | 20.0 | 2.60 | 0.89 |
| ASD Low ToM ($N$ = 15) | 2 | 13.3 | 5 | 33.3 | 1 | 6.7 | 2 | 13.3 | 5 | 33.3 | 2.20 | 1.57 |

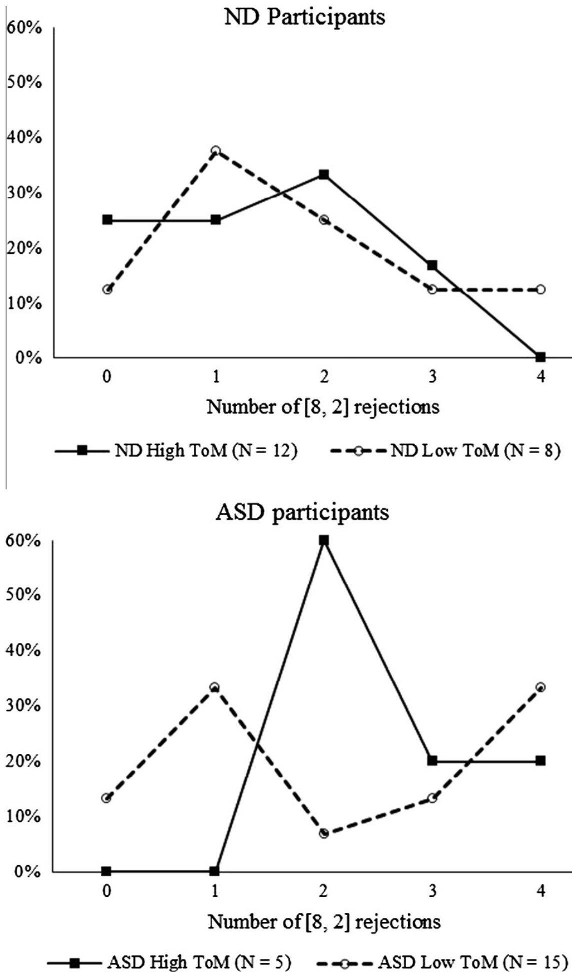


**Fig. 4.** Number of [8, 2] rejections at the individual level.

involving the corresponding terms for the ND participants are equal cannot be rejected. This regression analysis confirms the impression that our main results are driven by the behaviour of the ASD Low ToM participants.[17] This interpretation is also in line with the fact that differences in ToM do not seem to matter for the behaviour of our (ND) Ps as shown in Panels C and D of Table 2.

### 4.4. Learning and individual heterogeneity

Before we turn to a general discussion of our findings, we complete our analysis by considering the issues of learning and individual heterogeneity, which can be important given the within-subject nature of our design.

Table 7 reports rejection rates for the [8, 2] allocation in the four mini-UGs depending on whether they appeared in the first, second, third or fourth position of the mini-UG sequence, pooling over all Rs. We are interested in whether there are any systematic trends in rejection rates. Overall, that does not seem to be the case. For most games, there is not a clear trend. Only in the [10, 0]-UG there is a slight downward trend that, however, is not statistically significant in a Fisher's exact test.[18]

We now turn to consider whether there is individual-level heterogeneity in our participants' inclination to reject the [8, 2] allocation. Table 8 reports the numbers, and the corresponding percentages, of participants who never reject, reject once, twice, three or four times over the four mini-UGs, followed by the mean and standard deviation of the number of rejections, pooling over all Rs and by subgroups.

The breakdown by subgroups is particularly interesting, so we show it in Fig. 4, separately for ND and ASD participants.

While both High ToM and Low ToM ND Rs display a unimodal pattern, in which intermediate numbers of rejections are relatively more frequent than either high and low numbers, High ToM and Low ToM ASD participants have divergent behaviours. The High ToM are similar to the ND participants (the more coarse shape of the distribution is due to the fact that there are only 5 of them), but the Low ToM have a bimodal distribution, in which low and high numbers of rejections are relatively more frequent than intermediate ones. This suggests a more repetitive tendency to either (almost) always accept or (almost) always reject, consistent with the findings by Sally and Hill (2006), and may be useful in the understanding of the interaction between ASD and Low ToM that we have highlighted in the previous section.

## 5. Discussion and conclusion

In this article, we have used a sample comprising children affected by ASD to shed further light on the role of perceived intentions in ultimatum bargaining. We have extended FFF's mini-UG design and compared the rejection behaviour of ND and ASD Rs.

In line with our expectations based on ASD individuals' deficit in attributing intentions and mental states to other individuals, we have found that ASD Rs are much less responsive than ND Rs to changes in the alternative to the relatively unequal [8, 2] allocation which was the focus of our analysis. Given ASD individuals' typical ToM deficits, ToM differences were the obvious candidate for explaining the observed behavioural differences between ASD and ND participants. In our sample as a whole, it is indeed the case that individuals who fail our ToM mind test are relatively irresponsive to what might be interpreted as an intention perceived on the basis of alternative allocations which were not chosen. However, this difference is mainly driven by the 15 ASD participants who fail the ToM test, while the remaining 5 who pass it show behaviour more in line with that of the (High and Low ToM) ND participants. We now speculate on the potential reasons for this unexpected result.

One possibility is that, while ToM differences are important for the phenomena that we study, our test is not sufficiently fine-grained to measure ToM precisely. Mentalising abilities cover a wide range of different skills that develop progressively over a number of years. The dichotomous nature of our test (and other equivalent tests used in the literature) may not capture this richness to a sufficient level of detail.

An alternative possibility is that some other factor is responsible for the interaction between ASD and ToM that we seem to observe. The link could stem from ASD individuals' deficits in *executive functioning* – i.e. the group of abilities that maintain an appropriate problem-solving set for goal-directed behaviours (Alvarez & Emory, 2006). These deficits often result in repetitive and perseverative responses (e.g. Rumsey, 1985), which would be consistent with the bimodal distribution of individual-level rejections that we showed in Fig. 4 for Low ToM ASD participants. The correlation between ToM impairments and executive functioning found by Ozonoff et al. (1991) in ASD individuals, but not in the ND control group, is consistent with this conjecture. In future research, it might be of interest to explore the importance of executive functions in strategic decision-making.

[17] In the Appendix, we report the results of a mediation analysis in the spirit of Kenny (2013) and MacKinnon and Dwyer (1993), which points to the same conclusion (although this analysis does not take into account the lack of independence resulting from each individual being observed four times). We are interested in whether differences in ToM (a dichotomous variable – $M$) mediate the observed effect of ASD (a dichotomous independent variable – $X$) on rejection rates (a dichotomous dependent variable – $Y$) across our four games. The approach is based on regressing $Y$ over $X$, $M$ over $X$, and $Y$ over $M$ and $X$, and allows to decompose the total effect of $X$ on $Y$ into a direct effect (from $X$ to $Y$) and an indirect effect (from $X$ to $M$ to $Y$). In all regressions, we use Game as a covariate. This approach indicates that, in our case, most of the differences between ASD and ND Rs are the result of a significant direct effect, while the indirect (mediating) effect of ToM is very small and not statistically significant.

[18] A breakdown by subgroups based on both diagnosis and ToM performance is presented in the Appendix. The numbers of observations are very small, but no general trend is detectable there either.

Overall, our results are in line with Sally and Hill's (2006) conclusion that, for the kind of strategic interactions that we study in the laboratory, ToM differences may be less important than what one could expect *a priori*.

More broadly, they may be indicative of the potentially important distinction between *menu dependence* – i.e. behaviour that is affected by the set of options that are available – and *intention dependence* – i.e. behaviour that is affected by the decision-maker's perception of the intentions behind their co-players' actions. While in some cases menu dependence may arise as a consequence of intention detection (and therefore also being interpretable as intention dependence), this may not necessarily be the case. The attraction and compromise effects in the individual decision-making literature (e.g. Huber, Payne, & Puto, 1982; Simonson, 1989) clearly indicate that menu dependence may occur in non-interactive settings. Therefore, it is not unconceivable that in the context of strategic interactions, menu dependence may arise from direct comparisons between the characteristics of the available options, without any consideration for intentions (e.g. Sandbu, 2007). Such a 'menu dependence' effect could explain why our ND and High ToM ASD participants (as well as adults and adolescents in other implementations of the FFF design) react systematically to changes in the set of available options, independently of intention detection. As we suggested above, the different result in the Low ToM ASD subsample could be due to other factors, such as the correlation between ToM impairment and executive functioning. The exploration of the individual-specific factors and contextual conditions that promote menu dependence in interactive settings is a very interesting question that we leave to future research.

## Acknowledgements

We would like to thank Gianluigi Mazziti e Pietro Maria Peruzzi for valuable research assistance; Robert Sugden, Daniel Zizzo, Benedetto Gui, Luigi Mittone and Luca Stanca for their feedback on previous versions of the manuscript; Gordon Brown, Robin Cubitt, Simon Gächter, Peter Hammond, Adam Harris, Andrew Oswald, Elke Renner, Martin Sefton, Chris Starmer, two anonymous referees and the participants in various conferences and workshops in Trento, Copenhagen, Florence, Chicago, Halifax, Innsbruck, Rome, Warwick and Nottingham for their comments and suggestions. Financial support from the Economic and Social Research Council of the UK (award no. RES-051-27-0248) and the Regione Autonoma della Sardegna (L.R. 7/2007, award no. CRP3-223) is gratefully acknowledged.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.joep.2015.05.001.

## References


Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review, 16*(1), 17–42.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, 30*(2), 285–297.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37–46.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (1993). *Understanding other minds. Perspectives from autism*. Oxford: Oxford University Press.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (2000). *Understanding other minds. Perspectives from developmental cognitive neuroscience*. Oxford: Oxford University Press.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes, 63*(2), 131–144.

Bolton, G. E., & Ockenfels, A. (2000). ERC – A theory of equity, reciprocity and competition. *American Economic Review, 90*, 166–193.

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior, 10*, 95–121.

Brandts, J., & Solà, C. (2001). Reference points and negative reciprocity in simple sequential games. *Games and Economic Behavior, 36*, 138–157.

Camerer, C. (2003). *Behavioral game theory*. Princeton: Princeton University Press.

Casari, M., & Cason, T. N. (2009). The strategy method lowers measured trustworthy behaviour. *Economics Letters, 103*, 157–159.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*, 817–869.

Cooper, D. J., & Kagel, J. H. (2013). Other-regarding preferences: A selective survey of experimental results. In J. H. Kagel & A. E. Roth (Eds.). *Handbook of experimental economics* (Vol. 2). Princeton University Press.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47*, 268–298.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry, 41*, 20–26.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*, 293–315.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817–868.

Fischbacher, U., Quercia, S., & Gächter, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology, 33*(4), 897–913.

Güney, Ş., & Newell, B. R. (2013). Fairness overrides reputation: The importance of fairness considerations in altruistic cooperation. *Frontiers in Human Neuroscience., 7*, 252. http://dx.doi.org/10.3389/fnhum.2013.00252.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*(4), 367–388.

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior. *Journal of Economic Psychology, 11*, 417–449.

Harbaugh, W. T., Krause, K., & Liday, S. J. (2003). *Bargaining by children*. University of Oregon Economics Working Paper No. 2002-4. SSRN: http://ssrn.com/abstract=436504 or http://dx.doi.org/10.2139/ssrn.436504.

Hayashi, H. (2007). Young children's understanding of second-order mental states. *Psychologia, 50*(1), 15–25.

Hedden, T., & Zhang, J. (2002). What do you think I think you think? Strategic reasoning in matrix games. *Cognition, 85*, 1–36.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9*(1), 90–98.

Kenny, D. A. (2013). *Mediation with dichotomous outcomes*. <http://davidakenny.net/doc/dichmed.pdf> Accessed 29.10.14.

Lecce, S., & Hughes, C. (2010). 'The Italian job'? Comparing theory of mind performance in British and Italian children. *British Journal of Developmental Psychology, 28*, 747–766.

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology, 4*(3–4), 231–246.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, CA: Western Psychological Services.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*, 144–158.

Miller, S. A. (2012). *Theory of mind: Beyond the preschool years*. New York: Psychology Press.

Nelson, W. (2002). Equity or intention: It is the thought that counts. *Journal of Economic Behavior & Organization, 48*(4), 423–430.

Oosterbeek, H., Sloof, R., & Van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics, 7*, 171–188.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*(7), 1081–1105.

Perner, J., & Wimmer, H. (1985). John thinks that Mary thinks that... attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437–471.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*, 1281–1302.

Roth, A. E. (1995). Bargaining experiments. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton: Princeton University Press.

Rumsey, J. M. (1985). Conceptual problem-solving in highly verbal, nonretarded autistic men. *Journal of Autism and Developmental Disorders, 15*(1), 23–36.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "Window" task as a measure of strategic deception in pre-schoolers and autistic subjects. *British Journal of Developmental Psychology, 9*(2), 331–350.

Sally, D., & Hill, E. (2006). The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of Economic Psychology, 27*, 73–97.

Sandbu, M. E. (2007). Fairness and the roads not taken: An experimental test of non-reciprocal set-dependence in distributive preferences. *Games and Economic Behaviour, 61*(1), 113–130.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research, 16*(2), 158–174.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*, 395–402.

Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology, 28*(1), 69–78.

Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., & Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *Journal of Experimental Child Psychology, 105*, 130–137.

Volkmar, F., Paul, R., Klin, A., & Cohen, D. J. (2005). *Handbook of autism and pervasive developmental disorders* (3rd ed.). New York: Wiley & Sons.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin, 124*(3), 283–307.