J Autism Dev Disord
DOI 10.1007/s10803-012-1588-7

The Role of Causal and Intentional Judgments in Moral Reasoning in Individuals with High Functioning Autism

Marine Buon · Emmanuel Dupoux · Pierre Jacob · Pauline Chaste · Marion Leboyer · Tiziana Zalla

Abstract In the present study, we investigated the ability to assign moral responsibility and punishment in adults with high functioning autism or Asperger Syndrome (HFA/AS), using non-verbal cartoons depicting an aggression, an accidental harm or a mere coincidence. Participants were asked to evaluate the agent's causal and intentional roles, his responsibility and the punishment he deserves for his action. Adults with HFA/AS did not differ in judgments of suffering and causality from adults with typical development. However, subtle difficulties with judgments of intentional action and moral judgments were observed in participants with HFA/AS. These results are discussed in the light of emerging studies that deal with integrity of moral reasoning in individuals with autism spectrum disorders.

Keywords Moral judgment · Theory of mind · Causal reasoning · Autism

M. Buon (✉) · E. Dupoux
Laboratoire de Sciences Cognitives et de Psycholinguistique, Centre Nationale de la Recherche Scientifique, Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France
e-mail: marinebuon@gmail.com

M. Buon · P. Jacob · T. Zalla (✉)
Institut Jean Nicod, Centre Nationale de la Recherche Scientifique, Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France
e-mail: tiziana.zalla@ens.fr

P. Chaste · M. Leboyer · T. Zalla
INSERM U 955, IMRB, University Paris Est Creteil, AP-HP, Henri Mondor-Albert Chenevier Hospitals, Department of Psychiatry, Fondation FondaMental, French National Science Foundation, Creteil, France

Introduction

It is well documented that individuals with autism spectrum disorders (ASDs) are impaired in understanding and predicting others' behavior in terms of desires and beliefs (Baron-Cohen et al. 1985; Leslie and Frith 1987; Baron-Cohen 1989). There is also a wide consensus that this deficit can explain the severe difficulties in social interaction and communication often reported in ASDs. The ability to interpret others' actions in terms of mental states is critical to evaluate others and especially to form moral judgment about agents implicated in a harmful situation. Of all mental states, the concept of 'intentional action' is particularly important for the assignment of judgments of blame and praise: an agent who caused harm intentionally is blamed more severely than an agent who caused harm accidentally (Lagnado and Shannon 2008). Several experimental studies showed that Theory-of-Mind (ToM), defined as the ability to attribute mental states to oneself and others (Premack and Woodruff 1978), is crucial to form moral judgments. Baird and Astington (2004), for example, showed that in children with typical development, the ability to distinguish *intentional* from *accidental* harm is correlated with their ability to pass the standard False Belief Task (Wimmer and Perner 1983). Young and Saxe (2009) also reported that the activity of the right temporo-parietal junction (TPJ), a region implicated in reasoning about beliefs (Saxe et al. 2004), correlates with how much adults tended to exculpate an agent who had previously harmed someone accidentally, confirming that ToM processes play an important role in moral reasoning.

However, ToM impairments in people with High Functioning Autism (HFA) and Asperger Syndrome (AS) are not always manifest. Unlike subjects with low functioning autism, they usually pass first-order and second-

Published online: 04 July 2012

order ToM tasks (Bauminger and Kasari 1999; Bowler 1992; Happe 1994), likely using compensatory verbal strategies (Happe 1995). More advanced ToM tasks might reveal difficulties in reasoning about others' mental states in individuals with HFA or AS. For example, in the Strange Stories test (Happe 1994), in which subjects have to understand irony, sarcasm, bluff and double-bluff, HFA/AS participants are able to assign mental states, but fail to use them in a contextually appropriate way. In the Faux Pas task (Baron-Cohen et al. 1999; Zalla et al. 2009), participants have to judge a particular case of non-intentional action, reflecting an involuntary, socially inappropriate behavior. As reported by Zalla et al. (2009), faced with "faux pas" stories, a significant proportion of adults with HFA or AS judge that the agent acted with the intention to humiliate or offend the other person, where no malicious intent was involved.

ToM impairments might sometimes appear in people with HFA or AS when they are presented with non-verbal stimuli or when they are not explicitly prompted to attribute mental states. For example, using non-verbal animations depicting interactions between geometrical shapes, Castelli et al. (2002) reported that people with HFA or AS use less mental state terms and are less accurate in describing animations eliciting mentalistic interpretations, as compared to participants with typical development. Several studies have reported that individuals with HFA/AS are impaired in low level processing of socially relevant visual cues when faced with more ecological stimuli. For instance, the agent's eye gaze not only acts as an indicator of where another person is looking, but also plays a crucial role in communication and in reading others' thoughts and intentions (Baron-Cohen et al. 1997; Jellema et al. 2000). Qualitative impairment in eye-contact behaviour and atypical fixation patterns during the perception of facial stimuli are commonly mentioned in clinical and observational reports about individuals with ASDs (Baranek 1999; Charman et al. 1997; Volkmar and Mayes 1990). Recent studies with eye-tracking devices have revealed shorter fixation time on facial features, especially on the eye region, as compared to individuals without autism (Klin et al. 2002; Pelphrey et al. 2002; Vivanti et al. 2011). Senju et al. (2009) showed that, although adults with Asperger syndrome can pass a large variety of standard verbally instructed ToM tasks, they do not anticipate in they own eye movements where an agent searching for an object will be looking, if that agent has a false belief about where that object is. This reveals difficulties with the spontaneous encoding of socially relevant information and with the automatic on-line computation of others' mental states.

Overall, these findings suggest that although people with HFA and AS are able to pass first and second order ToM tasks, they might exhibit some impairments in the ability to encode others' mental states and reason about socially complex situations when stimuli and the paradigm used are sensitive enough. In this context, it appears particularly relevant to assess whether individuals with ASDs are impaired in making judgments of intentionality or in using information about intentions for moral reasoning.

With respect to the ability of people with HFA/AS to use information about the agents' mental states in moral judgments, results are rather divergent across experiments. Grant et al. (2005) presented verbal stories to children with autism where a little boy burns his brother's hand on purpose or accidentally, and stories where the same boy breaks an object intentionally or by accident. The results showed that, like control participants, children with autism judge intentional harm more severely than accidental harm and they judge the negative consequences worse when a human, rather than an object, is involved. However, when asked to justify their moral judgments, children with autism give justifications of poor quality or repeated elements of the story. According to the authors, the inability of children with autism to offer appropriate justifications might result from their lower mental age or, alternatively, from impairments in more complex information processing and executive control.

Recently, Zalla et al. (2009) presented participants with instances of so-called "faux pas", in which a speaker non-intentionally or accidentally hurts her addressee's feelings by making a socially inappropriate speech-act, whose meaning is partly based on the speaker's ignorance or false belief about a fact related to the addressee. Zalla et al. (2009) report that, unlike healthy individuals, individuals with HFA or AS tend to over-interpret as intended a non-intentional by-product of a speaker's communicative action, based on the speaker's ignorance or false belief. Similarly, Moran et al. (2011) reported that ToM impairments in individuals with HFA affect their moral judgments, as they are less willing than adults with typical development to exculpate agents with innocent intentions who accidentally caused harm. In this experiment, participants were presented with instances of accidental harm caused by an agent who puts white poison in her friend's coffee, while falsely believing that the white powder is sugar. Thus, there was a mismatch between information about the negative outcome of the agent's action (the agent poisoned her friend) and information about the agent's innocent intention (which could be inferred from her false belief that the white power was sugar). In this case, participants with HFA blamed the agent of accidental harm more severely than control participants.

Noteworthy, however, Zalla et al. (2009) and Moran et al. (2011)'s interpretations point to different impairments. While Moran et al. (2011) argued that since all participants with HFA/AS were able to pass first and

second order ToM tasks, they can encode others' mental states, but have difficulties in using such information in concert with other types of information, in Zalla et al.'s study, individuals with AS were impaired in their ability to infer the speaker's belief from contextual information and to use it to provide the correct interpretation of the social situation. One possible interpretation of these findings is that people with HFA/AS's judgment of intentional action may be influenced by evaluative considerations of morally aversive consequences (Knobe 2005).

Finally, a recent study using verbal stories depicting complex causal chains between an agent's action and a negative outcome found that participants with HFA/AS are more prone than controls to make use of mentalistic factors such as the agent's intention and negligence in their judgment of blame (Channon et al. 2011).

However, the fact that, in this experiment, the agent's intentions were explicitly stated, could have biased the participants' judgments by highlighting the relevant information (i.e., the agents' psychological state) and hidden their difficulties in spontaneous encoding of such information. It should be noted that such overt descriptions of the agents' psychological states is quite rare in everyday life, where the agent's mental states (i.e., desires, intentions and beliefs) need to be inferred from the situational context or from the perceptual cues available in the environment.

Taken together, these findings are divergent and it remains unclear whether individuals with HFA or AS are impaired in their ability to use relevant intentional cues to make judgments of intentionality or whether they fail to make use of this information for moral reasoning.

The present study addressed this issue. For this purpose, participants were presented with non-verbal cartoon scenarios representing a character being harmed in three different situational contexts in which the agent' causal role and her/his intention to harm varied systematically: (1) scenarios depicting an agent intentionally harming the victim; (2) scenarios in which an agent accidentally harmed the victim and (3) scenarios depicting an agent performing actions unrelated to the victim's suffering event. An important feature of our task is that, unlike many studies that use narratives, participants are not provided with mental states in verbal form, but have to infer them, in real time, from the character's actions, gaze and bodily postures. After each scenario, they were asked to evaluate the victim's physical distress, to judge the agent's causal and intentional roles in harming the victim, as well as the agent's responsibility and the amount of punishment she/he deserved. Importantly, the presence of the agent's causal role and the victim's suffering as dependent measures prevented participants from focusing preferentially on the intentional cues. We expected difficulties in making appropriate intentional and moral judgments in individuals with HFA to increase in those circumstances in which outcome and intention information are incongruent.

Method

Ethics Statement

The present research has been approved by the local Ethical committee (Inserm, Institut Thématique Santé Publique; C07-33). All participants signed informed consent before volunteering for this study, and all investigation has been conducted according to the principles expressed in the Declaration of Helsinki.

Participants

Two groups of adults participated to the study. A group of sixteen adults with a clinical diagnosis of HFA or AS according to DSM-IV R (American Psychiatric Association 2000) and to ASDI (Asperger Syndrome Diagnostic Interview, Gillberg et al. 2001) were recruited from Albert Chenevier Hospital in Creteil. All diagnoses were made by clinicians experienced in the field of autism independently of the present study. Diagnoses were based on observations of the participants and interviews with parents or caregivers and confirmed using the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000) and the ADI-R (Autism Diagnostic Interview; Lord et al. 1994). The cut-off points for the three classes of behavior are reciprocal social interaction (10), communication (8), and stereotyped behaviors (3), respectively. All participants scored above the cut-offs in the algorithms of the two instruments. As part of the checking process, the French translation of *A-TAC* (Autism, tics, AD-HD and other co-morbidities; Gillberg and Cederlund 2005) was completed by the parents. This screening questionnaire is focused on a number of abilities, conducts and behaviors in child's functioning as compared to his or her peers. Parents were asked to report any problem or specific characteristic during any period of life, even when this is no longer present. For each participant, an intellectual quotient (IQ) was calculated using the third version of the Wechsler Adult Intelligence Scale (WAIS III; Table 1).

Sixteen control participants (CP) were taken from the general population and matched with the HFA/AS group on Total, Verbal and Performance IQ, gender, chronological age and education level (Table 1). For CP, IQ was estimated by using two verbal (Similarities, Arithmetic) and two performance subtests (Digit symbol coding, Picture

Table 1 Means (and standard deviations) of demographic and clinical data for participants with High Functioning Autism or Asperger Syndrome (HFA/AS) and control participants (CP)

	Group		Group comparisons
	HFA/AS	CP	
Chronological age	26.8 (7)	24.2 (8)	$t(30) = 0.9$, $p > .1$
N (female:male ratio)	3/13	3/13	–
Education	14.25 (3.4)	14.23 (1.69)	$t(22,8) = 18$, $p > .1$
Total IQ	98(20)	102.31(10)	$t(21,8) = 0.2$, $p > .1$
Verbal IQ	100 (21)	101.8 (9.5)	$t(25) = 1.4$, $p > .1$
Performance IQ	95(17)	102 (9)	$t(24,52) = 1.3$, $p > .1$
ADI [B_fCJJ]*	18.8 (7.3); 12.1 (6.9); 6.7 (3.9)		

Between group differences were computed using independent sample *t* test

[B] = reciprocal social interaction, [C] = communication [D] = stereotyped behaviours

Completion) of the WAIS III (Cyr and Brooker 1984; Van Spaendonck et al. 1996).

Material

We designed 27 cartoons with the animation software Flash professional 8.0. In each cartoon, an agent is performing an action when the victim (Mr. Red) arrives. Each cartoon animation belongs to one of the three conditions; each condition differing with respect to the causal role of the agent regarding Mr. Red's distress and the agents' intention to harm.

In the *intentional condition*, the agent sees (as indicated by his gaze direction and body orientation) Mr. Red approach and he intentionally harms him. In the *accidental condition*, the agent faces away from Mr Red when Mr Red is approaching (he thus cannot see him) and he accidentally hurts him. In the *coincidental condition*, the agent faces away from Mr Red and since he just stops acting when Mr. Red hurts himself, he patently does not cause Mr Red's distress.

Each condition was presented six times across tree types of stories (see Table 2): (1) *Swinging stories* (the agent is swinging on a rope or goes on the swing while the victim, Mr Red, collides with him when he approaches him); (2) *Throwing stories* (the agent is throwing coconuts, rocks or apples while some of those strike the victim on the head); (3) *Mechanistic stories* (the agent is playing with a mechanical device which hurts the approaching victim). For each condition, three different types of cartoon story were designed.

Procedure

All participants were individually tested in a quiet distraction-free room either at the Albert Chenevier Hospital in Créteil or at the Ecole Normale Supérieure in Paris. Each cartoon animation was presented twice in a pseudo-random order on a portable computer. The task instructions were as follows: "In this experiment, you will watch a serie of twenty-seven cartoons on the computer screen. Each cartoon will be shown twice. After watching each cartoon, you are requested to answer the five questions displayed on this page. You have to follow the order of the cartoon presentation and you cannot come back to your previous answers". After each cartoon presentation, participants were required to answer the following five questions on a five point scale running from 1 (not at all) to 5 (completely):

1. The Suffering Question: "Did the victim suffer?"
2. The Causality Question: "Did the agent cause the victim's suffering?"
3. The Intentionality Question: "Did the agent have the intention to harm?"
4. The Responsibility Question: "Is the agent responsible for the victim' suffering?"
5. The Punishment Question: "What does the agent deserve?"

For the Punishment Question, answers were recorded using a scale running from 1 (rewards) to 7 (severe punishment).

Data Analysis

To assess group differences on each question, data were submitted to separate Two-Way repeated measures ANOVA with group (2: HFA/AS, CP) as between-subjects factor and condition (3: intentional, accidental, coincidental) as within-subjects factor. Fisher's exact tests were used for post-hoc analysis comparisons. For these statistics, the alpha level for acceptance was set at 0.05.

Table 2 Description and illustration of each condition as a function of the story type (swinging, launching, and mechanistic)

Swinging stories	Launching stories	Mechanistic stories
Baseline		
The agent is repeatedly swinging (on a swing or a rope)	The agent is repeatedly throwing objects (rocks, coconuts or apples)	The agent repeatedly activates a mechanism that guides the motion of an heavy object (bell, barrier, board)
Intentional condition *The agent sees of Mr Red when he is approaching. The agent stops his action and follows Mr Red's movements with his head and gaze…*		
He starts swinging again just when Mr Red is in front of him. He hits him.	He is throwing a new object at Mr Red and hits Mr Red on his head.	He activates the mechanism (the bell's lever here) just when Mr Red is under the heavy object (the bell)
		
Accidental condition *When performing his action, the agent's head is turned away from the path on which Mr Red is approaching. He cannot see Mr Red approach*		
He continues to swing in the presence of Mr Red. He hits him.	He throws a new object just when Mr Red is behind him. Mr Red receives the object on his head	He activates the mechanism just when Mr Red is under the object.
		
Coincidental condition *The agent's is turned away from the path on which Mr Red is approaching. He performs some actions (swinging, jumping, pulling) but his action has no causal role on Mr Red'suffering*		
He stops swinging few second before Mr Red's arrival. Mr Red loses his balance and falls down on his own.	He moves his harms but objects fall on their own. One object falls on Mr Red.	His action on the lever is decorrelated from the mechanism. The object falls on its own on Mr Red.
		

Each type of story contains three different versions

Results

The Suffering Question

When asked how much the victim was suffering we found a main effect of condition ($F(2,29) = 26.61$, $p < .0001$, $\delta^2 = 1$), but no effect of group ($F(1,30) = 1.74$, $p > .1$, $\delta^2 = .23$) and no significant interaction between group and condition ($F(2,29) = 2.12$, $p > .1$, $\delta^2 = .41$). The main effect of condition revealed that participants judged that the victim suffers less under the Coincidental condition that under the Accidental (mean difference = 0.34, $p < .0001$) and the Intentional (mean difference = 0.33, $p < .0001$) conditions while they judged that the victim experienced an equal amount of suffering in the Intentional and Accidental conditions (mean difference = 0.01, $p > .1$).

The Causality Question

When asked about the agent's causal role in provoking the victim's distress, we found a highly significant effect of condition ($F(2,29) = 273.36$, $p < .0001$, $\delta^2 = 1$) and a

marginally significant effect of group (F (1,30) = 3.91, $p = 0.06$, $\delta^2 = .47$) and a marginally significant group by condition interaction (F(2,29) = 2.41, $p < .09$, $\delta^2 = .46$).

The main effect of condition revealed that, overall, participants assigned a stronger causal role to the agent provoking the victim' distress in the intentional condition than to agents acting in both the accidental (mean difference = 0.48, $p < .001$) and the coincidental (mean difference = 2.89, $p < .0001$) conditions. Moreover, the agent in the accidental condition was assigned a stronger causal role than the agent in the coincidental one (mean difference = 2.41, $p < .0001$). Regarding the marginal effect of groups, post hoc analysis revealed that participants with HFA/AS tended to assign a stronger causal role to the accidental agent (mean between groups difference = 0.45, $p = .051$) and the coincidental agent (mean between groups difference = 0.46, $p = .08$) compared to control participants. However, as revealed by the marginal group by condition interaction, the two groups did not differ in judging the causal role of the agent from the intentional condition (mean between groups difference = 0.047, $p > .1$). As a result, while control participants considered that the intentional agent as having a greater causal role than the accidental one (mean difference = 0.73, $p < .001$) and that the accidental agent has a greater causal role than the coincidental one (mean difference = 2.41, $p < .0001$), participants with HFA/AS considered that the accidental agent has more causal role than the coincidental one (mean difference = 2.41, $p < .0001$) but did not distinguish the intentional and the coincidental ones on the basis of their causal role (mean difference between the intentional and the accidental agent = 0.23, $p > .1$; Fig. 1a).

The Intentionality Question

When asked whether the agent performed the action intentionally, a Two-Way repeated ANOVA yielded no main effect of group (F(1,30) = 0.92, $p > .1$, $\delta^2 = .14$), but a significant effect of condition (F(2,29) = 204,48, $p < .0001$, $\delta^2 = 1$) and a significant group by condition interaction (F(2,29) = 6.41, $p < .01$, $\delta^2 = .90$). The main effect of condition revealed that, overall, participants judged the intentional agent as acting more intentionally than both the accidental agent (mean difference = 2.45, $p < .0001$) and the coincidental agent (mean difference = 2.83, $p < .0001$) and that the accidental agent acted more intentionally than the coincidental agent (mean difference = 0.37, $p < .05$). However, this effect was qualified by a significant interaction indicating that the two groups differed in judging the agent' intentionality in both intentional and accidental conditions. Post hoc tests revealed that participants with HFA/AS judged the accidental agent as acting more intentionally than control participants (mean difference between groups = 0.69, $p < .05$). A marginal between group difference was also found on the intentionality rating of the Intentional agent, participants with HFA/AS tending to underrate the agent's intention to harm (mean difference between groups = 0.39, $p = .08$). Remarkably, although both groups judged that the intentional agents acted more intentionally than in the accidental one (HFA/AS, mean difference = 1.91, $p < .0001$; CP, mean difference = 3.05, $p < .0001$), only participants with HFA/AS tended to regard the accidental agent as acting more intentionally that the coincidental one (CP mean difference = 0.15, $p > .1$; HFA/AS, mean difference = 0.59, $p = 0.07$; Fig. 1b).

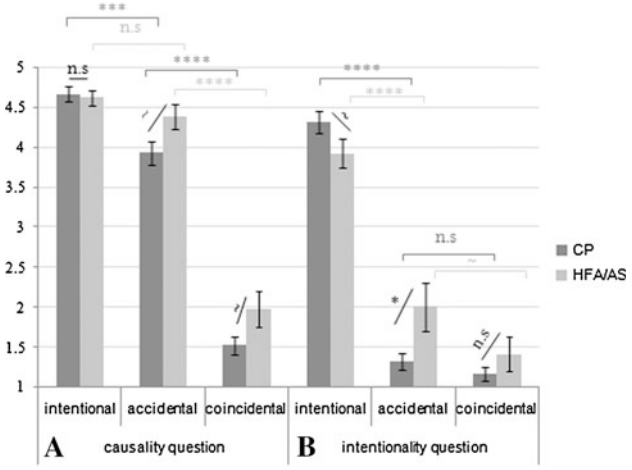

Fig. 1 Average results obtained to **a** the causality and **b** the intentionality questions as a function of the condition perceived and the experimental group. The main effects of group are displayed in *black*, and the main effects of conditions are displays in *dark gray* (for CP) and *light gray* (for HFA/AS). n.s, non significant; ∼, $.1 < p > .05$; *$p < .05$; **$p < .01$; ***$p < .001$

The Responsibility Question

When asked to evaluate the agent's responsibility, we found main effects of group (F(1,30) = 8.33, $p < .01$, $\delta^2 = .81$) and condition (F(2,29) = 270.52, $p < .0001$, $\delta^2 = 1$), as well as a significant group by condition interaction (F(2,29) = 9.57, $p < .001$, $\delta^2 = .98$). The main effect of condition revealed that, overall, participants judged the intentional agent as being more responsible than both the accidental (mean difference = 1.56, $p < .0001$) and the coincidental ones (mean difference = 2.87, $p < .0001$), and the accidental agent as being more responsible than the coincidental one (mean difference = 1.30, $p < .0001$). With respect to the main effect of group, post hoc tests revealed that participants with HFA/AS assigned more responsibility to the accidental agent

(mean difference = 0.92, $p < .01$) and tended to assign more responsibility to the coincidental agent (mean between group difference = 0.49, $p = 0.054$) than control participants. However, as revealed by the significant group by condition interaction, the two groups did not differ in judging the agent's responsibility in the intentional condition (mean difference = 0.144, $p > .1$). Despite this interaction, both groups judged the intentional agent as being more responsible that the accidental one (CP, mean difference = 2.01, $p < .0001$; HFA/AS, mean difference = 1.02, $p < .001$), and the accidental agent as being more responsible than the coincidental one (CP, mean difference = 1.08, $p < .0001$; HFA/ AS, mean difference = 1.52, $p < .0001$) (Fig. 2).

The Punishment Question

A Two-Way ANOVA conducted on the punishment question yielded no main effect of group ($F(1,30) = 1.62$, $p > .1$, $\delta^2 = .22$), but a main effect of condition ($F(2,29) = 248.91$, $p < .0001$, $\delta^2 = 1$) and a significant group by condition interaction ($F(2,29) = 6.85$, $p < .01$, $\delta^2 = .92$). The main effect of condition was due to participants judging the intentional agent as deserving more punishment than the accidental (mean difference = 1.38, $p < .0001$) and the coincidental (mean difference = 1.97, $p < .0001$) agents and the accidental agent as deserving more punishment than the coincidental one (mean difference = 0.58, $p < .0001$). Post hoc analyses revealed that the interaction effect was due to the two groups differing on judgment of punishment deserved by the accidental agent, group with HFA/AS considering that the accidental agent deserved more reprimand than the control group (mean difference between group = 0.57, $p = .01$). There was no between group differences on judgment of punishment deserved by the intentional agent (mean difference = 0.08, $p > .1$) and the coincidental one (mean difference = 0.11, $p > .1$). Despite this interaction, both groups reported that the intentional agents deserved more punishment than the accidental one (CP, mean difference = 1.71, $p < .0001$; HFA/AS, mean difference = 1.05, $p < .0001$), and that the accidental agents deserved more punishment than the coincidental ones (CP, mean difference = 0.35, $p < .01$; HFA /AS, mean difference = 0.81, $p < .01$) (Fig. 3).

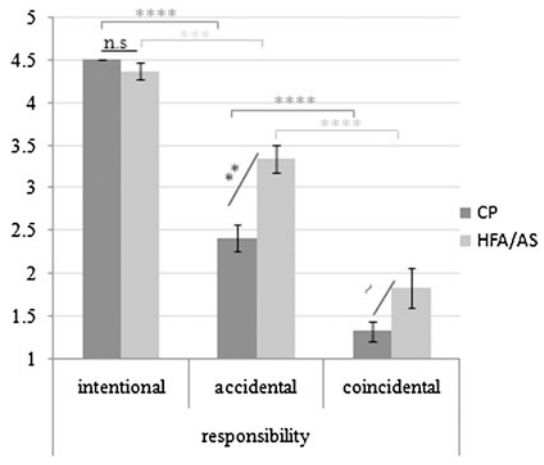

Fig. 2 Average results obtained to the responsibility question as a function of the condition perceived and the experimental group. The main effects of group are displayed in *black*, and the main effects of conditions are displays in *dark gray* (for CP) and *light gray* (for HFA/AS). n.s, non significant; ∼, $.1 < p > .05$; $*p < .05$; $**p < .01$; $***p < .001$; $****p < .0001$

Further Analyses

We reported between group differences on judgments of the agents' causal role, of the agent's intentional status, as well as on moral judgments of responsibility and punishment. We performed two additional analyses to establish whether differences in moral judgments (responsibility and punishment) would result from a single impairment in making judgment of intentionality or, alternatively, from two distinct deficits affecting separately mindreading and moral reasoning.

The first additional analysis allowed to disentangle the marginal interaction effect obtained on the causality question (i.e., control participants considered that the intentional agent is more causally involved in the victim' suffering than the accidental one, while participants with HFA/AS did not) from the effect on the intentionality question (i.e., HFA/AS participants tended to underestimate of the intentional agent and to overestimate the intention of the accidental agent). To this purpose, we first left out from the analysis all participants who showed a deviant intentionality rating of the agent, that is those who differed by more than two standard deviations from the control participants' mean intentional rating in the intentional (mean intentional rating = 4.32, s.d. = 0.53, cut-off = 3.26) and in the accidental (mean intentional rating = 1.32, s.d. = 0.4, cut-off = 2.13) conditions. Overall, two control participants and six participants with HFA/AS were thus removed from the analysis. In so doing, we assessed whether for the remaining ten participants with HFA/AS[1] and the fourteen control participants judgments of causality of the intentional agent differed from that of the accidental agent. Similarly to the previous analysis, a Fisher's exact test indicated that control participants

[1] This group included eight participants with a diagnosis of Asperger syndrome and two participants with a diagnosis of High Functioning autism.

considered that the agent in the intentional condition is more causally involved than the agent in the accidental one (mean difference = 0.76, $p < .0001$). The same analysis performed on the HFA/AS group did not yield a significant difference, suggesting that HFA/AS participants did not distinguish the two agents on the basis of their causal involvement (mean difference = 0.33, $p > .1$). This result showed that the marginal interaction we obtained on the causality judgment was not related to the effect we found on the intentionality judgment.

The second analysis aimed to disentangle the effects we obtained on the intentionality judgment from the one we obtained on moral judgments. Since significant differences on these dependant measures were mainly obtained on the accidental condition, we tested whether participants with HFA/AS who correctly detected the accidental agent's intentional state differed from control participants on their moral judgment with respect to this condition. Therefore, five participants with HFA/AS and one control participant were removed from the analyses because of their divergent intentionality judgments in the accidental condition. We left out individuals who differed by more than two standard deviations from control participants' mean intentional rating (see above for cut-off value). We then compared the remaining fifteen control participants and eleven participants with HFA/AS[2] on their judgments (causality, intentionality, responsability and punishment) in the accidental condition. A between-group t test on the intentionality judgment revealed no significant difference (mean difference = 0.10, $t(24) = 0.79$, $p > .1$). Importantly, we no longer found any group differences on the causality judgment (mean differences = 0.41, $t(24) = 1.51$, $p > .1$). However, a t test performed on the responsibility judgment revealed a significant between group difference (mean difference = 0.72, $t(24) = 3.29$; $p < 0.01$) suggesting that participants with HFA/AS were more prone than control participants to attribute responsibility to someone who harms someone else accidentally, even when they correctly interpreted the agent's intentional state. The between group t test performed on the punishment question also revealed that participants with HFA/AS attributed more punishment than control participants to the agent that caused the victim's harm accidentally (mean difference = 0.32; $t(24) = 2.47$, $p < 0.05$). As shown in Fig. 4, only one participant with HFA/AS (6%) is deviant on the intentionality judgment, five (31%) are deviant only in the responsibility judgment, four are deviant in both types of judgments, and seven of them scored within the normal range. Finally, only one HFA/AS participant (0.06%) was deviant on the punishment judgment and three (18%) were deviant in both judgment of punishment and judgment of intentionality (Fig. 4).

[2] This group included eight participants with a diagnosis of Asperger syndrome and three participants with a diagnosis of High Functioning Autism.

Discussion

The present study was designed to investigate the ability to form moral judgments in a group of adults with HFA/AS using a series of non-verbal cartoon scenarios in which the agent' intention to harm and his causal role in provoking the victim's distress varied systematically. Participants were asked to judge the same scenarios on several dimensions: the victim's distress, the agents' causal role, his intentions to harm, his responsibility and whether he deserved punishment for his action. Importantly, the causal role and the intentions were not provided in verbal format, as in several previous studies, but had to be inferred from purely non verbal cue, namely eye gaze and body movements.

This study revealed several effects. First, the two groups of participants delivered similar judgments of suffering: individuals with HFA/AS reported a spared sensitivity to the victim's distress, confirming previous studies showing preserved physiological responses and empathic processing when faced with others' distress (Blair 1996; Yirmiya et al. 1992). Secondly, we observed that participants with HFA/AS tended to differ from control participants in their causality judgments. Indeed, unlike the typically developed group, participants with HFA/AS did not integrate the intentional state of the perceived agents in their judgments

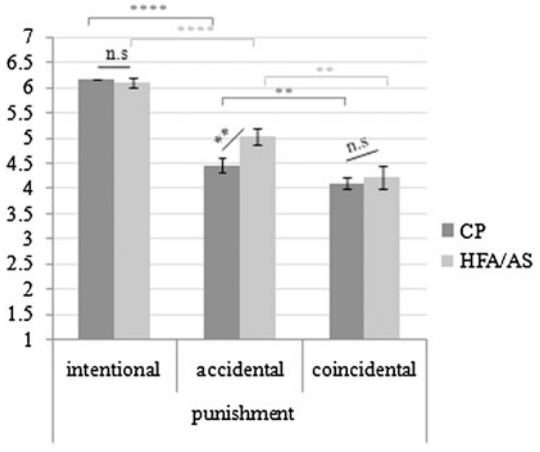

Fig. 3 Average results obtained to the punishment question as a function of the condition perceived and the experimental group. The main effects of group are displayed in *black*, and the main effects of conditions are displays in *dark gray* (for CP) and *light gray* (for HFA/AS). n.s, non significant; ~, $.1 < p > .05$; *$p < .05$; **$p < .01$; ***$p < .001$, ****$p < .0001$

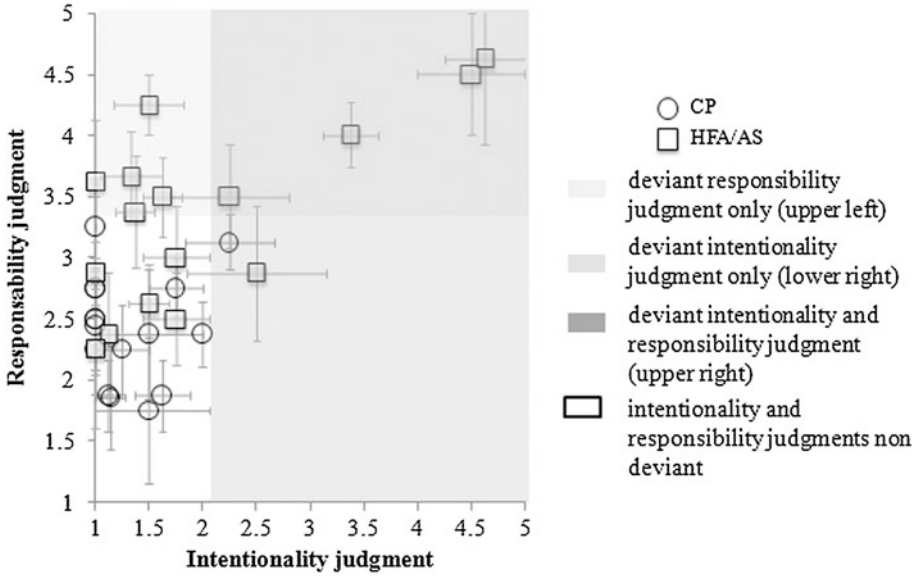

Fig. 4 Individual scores on the intentionality judgment (x axis) and the responsibility judgment (y-axis) for control (*circles*) and HFA/AS participants (*squares*). Each individual score is displayed with one standard error on each side. The control population two standard deviation Z-score cut-off is shown for intentionality and responsibility with *vertical* or *horizontal lines*, respectively, delimiting regions of impaired or non impaired performance

of causality. Thirdly, unlike typically developed adults, individuals with HFA/AS had difficulties in judging the agent's intention, since they tended to overrate the agent's intention to harm in the accidental condition and to underestimate the agent's intention in the intentional one. Finally, the two groups differed in their moral judgments. When asked to judge the agent's responsibility and estimate the punishment he deserved, individuals with HFA/AS regarded the agent as being more responsible in the accidental and coincidental conditions and punished him more severely than control participants in the accidental condition.

Diminished Sensitivity to Psychological States in Causality Judgments

With respect to the causality judgments, in accordance with several studies (Alicke 1992, 2000; Lagnado and Shannon 2008), we found that typically developing participants reported that the intentional agent was more strongly causally related to the victim's distress than the accidental agent. For instance, Alicke (2000) showed that a wife who gives to her husband a medication overdose *on purpose* is judged as being more causally involved in his death than if the overdose was given *by mistake*. By contrast, the present results disagree with a recent study showing the same propensity to assign a stronger causal role to an intentional than to an accidental agent in individuals with Asperger syndrome (Channon et al. 2011). In Channon et al.'s (2011) experiment, participants were presented with verbal stories in which the agent's mental states were explicitly stated and therefore made salient. The fact that we used visual stimuli that would not explicitly highlight the agent's mental states might contribute to explaining difficulties to use information about intentions for causality judgments in participants with HFA/AS. Importantly, this effect persists even when the analysis was performed only on those participants showing a deviant intentionality judgment. This finding suggests that the perceptual saliency of intentional cues rather than their verbal description, which could have hidden their difficulties in spontaneous encoding, is critical to allow individuals with HFA/AS to integrate intentional cues in their causality judgments spontaneously.

Distinct Impairments in Intentionality Judgment and Moral Reasoning

In the present study, participants with HFA/AS differed from control participants in their judgments of intentions, responsibility and punishment. Given the crucial role that the intentionality judgment plays in moral reasoning (Piaget 1965/1932; Cushman 2008; Alicke 2000; Alicke et al. 2008), it is likely that the overestimation of responsibility and the related amount of punishment are originated by the over-attribution of the intention to harm. However, inappropriate judgments of intentions alone cannot account for the inflated judgments of responsibility and punishment given by individuals with HFA/AS since even those participants who correctly recognized the agent's intentions judged more harshly than control participants the agent who harmed the victim accidentally. This is to say that, abnormal moral judgments in individuals with HFA/AS might result from two distinct impairments: the first one affecting more specifically the intentionality judgments and characterized by an over-attribution of intentions; the second one, affecting moral judgments and mainly characterized by an increased moral severity. We discuss these two patterns in turn.

Misattribution of Intentions: A Deficit in Intentionality Judgments

Difficulties with the intentionality judgment in our participants with HFA/AS are consistent with a large body of evidence showing low-level impairments in processing intentional cues, such as gaze orientation, emotional facial expressions and ascription of goals and intentions (Baron-Cohen 1995; Charman et al. 1997, Vivanti et al. 2011). In our task, gaze and head orientation are crucial for inferring the agent's intention (i.e., the intentional agent's gaze followed the approaching victim before harming him, while the accidental agent had his face turned away from the path taken by the victim). Thus, low-level impairments in processing intentional cues in participants with HFA/AS may explain their diminished performance in assigning the appropriate agents' intentions. Such difficulties, characterized in terms of an over-attribution of intentions to agents, mostly increased in the *accidental* condition, while there was only a marginal trend to under-attribute intentions in the intentional condition.

By using a series of "faux pas" stories, Zalla et al. (2009) also reported difficulties with the interpretation of accidental actions in a group of adults with HFA/AS. Although a faux pas is a non-intentional by-product of a speech act, they were more prone than control participants to consider that someone who committed a faux pas has done it with the malicious intention to hurt the listener's feelings. Importantly, this study suggested that deviant judgments of intentionality in participants with HFA/AS mostly occur when (1) the negative outcome of the agent's action is produced accidentally, (2) the agent's intention is inferred from verbally-presented situational contexts, and (3) the intentionality judgment is influenced by the moral valence of the outcome of the action.

The present study corroborates previous findings showing subtle difficulties with attribution of intentions to others in participants with HFA/AS and further confirms that difficulties arise when faced with conflicting and/or an aversive outcome, such as those exemplified in the accidental condition. Similarly, abnormal judgments of intentionality have been reported in situations in which subjects had to solve an apparent contradiction between the neutral intention and the accidental harmful outcome (Buccino et al. 2007; Nelson-LeGall 1985). It should be noted however, that in the present study abnormal intentionality judgments in participants with HFA/AS are neither completely restricted to accidental actions, nor present in all the experimental conditions.

Further research should address this issue using other cases of mismatch between the intention and the outcome, e.g. attempted harm (agent's intention to harm, no harmful outcome) or accidental help (the agent did not intend to help, helpful outcome). Furthermore, it would be helpful to use eye-tracking measures to investigate the ways in which intentional cues are visually processed in all these scenarios.

Moral Severity: An Intention Integration Deficit

The present results also show that participants with HFA/AS judged the agent who harmed the victim accidentally more severely than control participants, even when the agent's intention was correctly ascribed. In accordance with previous studies (Zalla et al. 2011; Zalla and Leboyer 2011), these findings suggest that even when information about the agent's psychological states is correctly inferred, it is not fully and flexibly used for moral reasoning in individuals with HFA/AS. In Zalla and Leboyer's (2011) study, individuals with autism were presented with scenarios where an agent's action yields an expected but unwanted positive side effect (e.g., helping the environment or winning a bull's-eye contest, Zalla and Leboyer 2011). Both HFA/AS and control participants correctly judged that the outcome was not intended. However, unlike control participants, participants with HFA/AS tended to *praise* the agent for his (unintentional) action, while failing to appeal to the agent's intention when asked to justify their moral judgments. Although the intentional states were correctly assigned, judgments of praise were not informed by the agent's intentions or desires in adults with HFA/AS.

Along the same line, Zalla et al. (2011) reported that adults with HFA provided inappropriate moral justifications and inadequate evaluation of the seriousness of normative transgressions. The authors concluded that individuals with ASDs failed to integrate and use relevant information about the agent's intentions and affective states in conscious moral reasoning. Similarly, individuals with HFA/AS have difficulties performing more 'advanced' ToM tasks requiring the use of social norms and mental contents for social reasoning (such as refraining to say something true not to hurt a friend (Baron-Cohen et al. 1999; Zalla et al. 2009), the detection of sarcasm, irony or bluff (Happe 1994), or the mentalistic interpretations of non-verbal perceived animations (Castelli et al. 2002). Taken together, these findings support the hypothesis that difficulties using mental state information for social reasoning in individuals with HFA/AS might reflect an under-reliance on the representation of the agent's intentional states. In accordance with this explanation, Young and Saxe (2009) showed that enhanced activation of TPJ, the region involved in mental states attribution, correlated with improved performance in moral judgment (e.g., the ability to exculpate an agent causing accidental harm). Young et al. (2010) used transcranial magnetic stimulation to disrupt transiently the neural activity in the

right TPJ during a task requiring moral judgment and showed that the capacity to use mental states (beliefs, intentions) in moral reasoning might be reduced in the case of attempted harms (e.g., agent intended but failed to do harm).

An alternative hypothesis points to an impairment in the domain of executive functions in individuals with ASDs, in that it is concerned with high level reasoning and complex problem solving (Ozonoff 1997; Russell 1997). This is in accordance with a large body of evidence showing impairments in response initiation, planning, inhibition and cognitive flexibility (i.e., the ability to engage and disengage actions in the service of overall goals) in individuals with HFA or AS (Ozonoff 1997; Pennington and Ozonoff 1996; Hill and Bird 2006).

In addition, developmental research on moral cognition has shown that children's use of information about intentions, acts and outcomes will be constrained by more general development of executive functions, in particular in the use of higher-order rules (Zelazo et al. 1996, 2002). According to this view, an impairment at the computational level of information processing would arise in people with ASD, when discordant epistemic states have to be combined into a single system of inferences. In evaluating an agent who caused accidental harm, one is faced with a conflict between the harmful outcome of the agent's action and the agent's innocent intention, but no conflict arises when intended harm is combined with a harmful outcome. Thus, it may well be that, in individuals with typical development, information about the agent's innocent intention is loaded on executive resources to inhibit the prepotent tendency to blame an agent for the harmful outcome of his action, and mitigate the blame on the basis of the agent's innocent intention (Young et al. 2007). It is likely that in circumstances of increased executive and attention demands, failure to inhibit the emotional responses elicited by the victim's distress, rather than the ability to infer an agent's intention, would explain difficulties with moral reasoning in individuals with HFA/AS. Similar difficulties in participants with HFA have been reported in studies using verbally presented harmful situations. For example, Moran et al. (2011) have showed that in judging accidental harm, participants with HFA, who successfully passed a standard false belief task, exhibited an under-reliance on information about a person's innocent intention, together with an over-reliance on the action's negative outcome. According to the authors, these findings reveal impairments in integrating conflicting information about mental state (e.g., neutral intentions) and (aversive) action outcome in moral judgment.

It is important to note, however, that the two hypotheses discussed above are not exclusive. Moran et al. (2011) hypothesized that what makes it hard for individuals with HFA/AS to respond to conflicting information about an agent's intention and the action outcome is the lack of a *robust* and fully flexible ToM, which might be necessary to *override* the prepotent response driven by emotionally salient information. While the lack of a robust ToM sustains the first explanation, failure to override prepotent responses advocates the executive dysfunction hypothesis. Further research is needed to assess the explanatory role of these two hypotheses by using specific tasks tapping on the ability to integrate intentionality judgments in moral reasoning, combined with validated measures of different components of executive functioning and ToM abilities in individuals with HFA.

Although these results are in agreement with previous findings, there are a couple of shortcomings with this study that need to be acknowledged here. The first issue concerns the lack of validated measures of executive function and ToM abilities which might help explaining the pattern of impairments observed in our clinical group, as well as the individual variability. The second limitation concerns the small sample size of our population which might weaken our conclusions and hidden further group differences.

Conclusions

The present findings show that two distinct patterns of impairments might contribute to explaining deviant judgments of responsibility and punishment in our participants with HFA/AS. Since these impairments are not found conjointly in the same individuals, this could be due to two separate sub-categories of HFA/AS: those whose difficulties in low-level processing of intentional cues would lead to misattribution of intentions and those whose difficulties using information about intentions in moral reasoning would lead to inflation of moral severity.

These results point to the critical importance of examining data from ASD population at the level of individuals, rather than relying exclusively on group data. The presence of sub-categories could go a long way toward explaining some of the variability in findings across autism studies and might be of considerable relevance for clinical practice. This study also emphasizes the need to distinguish and isolate different component processes that might be responsible for diminished abilities in complex cognitive functions, such as moral reasoning.

Acknowledgments We gratefully acknowledge the commitment of the participants and their families to the pursuit of research in autism. This research was supported by Fondation FondaMental and Fondation Orange to TZ and ML, and ANR-09-BLAN-0327 SOCODEV to MB, PJ and ED.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality & Social Psychology, 63*(3), 368–378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556–574.

Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin, 34*(10), 1371–1381.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition DSM-IV-TR (Text Revision). Washington, DC: American Psychiatry Association.

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development, 2*(103), 37–49.

Baranek, G. T. (1999). Autism during infancy: a retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of Autism and Developmental Disorder, 29*(3), 213–224.

Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay. *J Child Psychol Psyc, 30*(2), 285–297.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Book: MIT Press.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *J Child Psychol Psyc, 38*(7), 813–822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37–46.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorder, 29*(5), 407–418.

Bauminger, N., & Kasari, C. (1999). Brief report: theory of mind in high-functioning children with autism. *Journal of Autism and Developmental Disorder, 29*(1), 81–86.

Blair, R. J. (1996). Brief report: morality in the autistic child. *Journal of Autism and Developmental Disorder, 26*(5), 571–579.

Bowler, D. M. (1992). "Theory of mind" in Asperger's syndrome. *J Child Psychol Psyc, 33*(5), 877–893.

Buccino, G., Baumgaertner, A., Colle, L., Buechel, C., Rizzolatti, G., & Binkofski, F. (2007). The neural basis for understanding non-intended actions. *Neuroimage, 36*(Suppl 2), T119–T127.

Castelli, F., Frith, C., Happe, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(Pt 8), 1839–1849.

Channon, S., Lagnado, D., Fitzpatrick, S., Drury, H., & Taylor, I. (2011). Judgments of cause and blame: Sensitivity to intentionality in Asperger's Syndrome. *Journal of Autism and Developmental Disorder, 41*(11), 1534–1542.

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: an investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology, 33*(5), 781–789.

Cushman, F. (2008). Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353–380.

Cyr, J. J., & Brooker, B. H. (1984). Use of appropriate formulas for selecting WAIS-R short forms. *Journal of Consulting and Clinical Psychololgy, 52*, 903–905.

Gillberg, C., & Cederlund, M. (2005). Asperger syndrome: familial and pre- and perinatal factors. *Journal of Autism and Developmental Disorder, 35*(2), 159–166.

Gillberg, C., Gillberg, C., Råstam, M., & Wentz, E. (2001). The Asperger Syndrome (and high-functioning autism) Diagnostic Interview (ASDI): A preliminary study of a new structured clinical interview. *Autism, 5*(1), 57–66.

Grant, C. M., Boucher, J., Riggs, K. J., & Grayson, A. (2005). Moral understanding in children with autism. *Autism, 9*(3), 317–331.

Happe, F. G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorder, 24*(2), 129–154.

Happe, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843–855.

Hill, E. L., & Bird, C. M. (2006). Executive processes in Asperger syndrome: patterns of performance in a multiple case series. *Neuropsychologia, 44*(14), 2822–2835.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain and Cognition, 44*(2), 280–302.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*(9), 809–816.

Knobe, J. (2005). Theory of mind and moral cognition: exploring the connections. *Trends in Cognitive Sciences, 9*(8), 357–359.

Lagnado, D. A., & Shannon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition, 108*, 754–770.

Leslie, A. M., & Frith, U. (1987). Metarepresentation and autism: How not to lose one's marbles. *Cognition, 27*(3), 291–294.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr, Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorder, 30*(3), 205–223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorder, 24*(5), 659–685.

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., et al. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National academy of Sciences of the United States of America, 108*(7), 2688–2692.

Nelson-LeGall, S. (1985). Motive-outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology, 21*(2), 332–337.

Ozonoff, S. (1997). Components of executive function in autism and other disorders. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 179–211). New York: Oxford University Press.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorder, 32*(4), 249–261.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *J Child Psychol Psyc, 37*(1), 51–87.

Piaget, J. (1965/1932). *The moral judgement of the child*. New York: Free Press.

Premack, A., & Woodruff, P. W. (1978). Does the chimpanzee have a theory of mind? *The behavioral and brain sciences, 4*(2), 515–526.

Russell, J. (Ed.). (1997). *Autism as an executive disorder*. New York: Oxford University Press.

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus

responds to observed intentional actions. *Neuropsychologia, 42*(11), 1435–1446.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883–885.

Van Spaendonck, K., Berger, H., Horstink, M., Buytenhuijs, E., & Cools, A. (1996). Executive functions and disease characteristics in Parkinson's disease. *Neuropsychologia, 34*, 617–626.

Vivanti, G., McCormick, C., Young, G. S., Abucayan, F., Hatt, N., Nadig, A., et al. (2011). Intact and impaired mechanisms of action understanding in autism. *Developmental Psychology, 47*(3), 841–856.

Volkmar, F., & Mayes, L. (1990). Gaze behavior in autism. *Development and Psychopathology, 2*(1), 61–69.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128.

Yirmiya, N., Sigman, M. D., Kasari, C., & Mundy, P. (1992). Empathy and cognition in high-functioning children with autism. *Child Development, 63*(1), 150–160.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporo-parietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *PNAS, 107*(15), 6753–6758.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National academy of Sciences of the United States of America, 104*(20), 8235–8240.

Young, L., & Saxe, R. (2009). Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia, 47*(10), 2065–2072.

Zalla, T., Barlassina, L., Buon, M., & Leboyer, M. (2011). Moral judgment in adults with autism spectrum disorders. *Cognition, 121*(1), 115–126.

Zalla, T., & Leboyer, M. (2011). Judgment of intentionality and moral evaluation in individuals with high functioning autism. *Review of Philosophy and Psychology, 2*(4), 681–698.

Zalla, T., Sav, A. M., Stopin, A., Ahade, S., & Leboyer, M. (2009). Faux pas detection and intentional action in Asperger Syndrome. A replication on a French sample. *Journal of Autism and Developmental Disorder, 39*(2), 373–382.

Zelazo, P. D., Burack, J. A., Benedetto, E., & Frye, D. (1996). Theory of mind and rule use in individuals with Down's syndrome: A test of the uniqueness and specificity claims. *J Child Psychol Psyc, 37*(4), 479–484.

Zelazo, P. D., Jacques, S., Burack, J., & Frye, D. (2002). The relation between theory of mind and rule use: Evidence from persons with autism-spectrum disorders. *Infant and Child Development, 11*(11), 171–195.