Perceiving Goals and Actions in Individuals with Autism Spectrum Disorders

Tiziana Zalla · Nelly Labruyère · Nicolas Georgieff

Published online: 19 February 2013
© Springer Science+Business Media New York 2013

Abstract In the present study, we investigated the ability to parse familiar sequences of action into meaningful events in young individuals with autism spectrum disorders (ASDs), as compared to young individuals with typical development (TD) and young individuals with moderate mental retardation or learning disabilities (MLDs). While viewing two videotaped movies, participants were requested to detect the boundary transitions between component events at both fine and coarse levels of the action hierarchical structure. Overall, reduced accuracy for event detection was found in participants with ASDs, relative to participants with TD, at both levels of action segmentation. The performance was, however, equally diminished in participants with ASDs and MLDs under the coarse-grained segmentation suggesting that difficulties to detect fine-grained events in ASDs cannot be explained by a general intellectual dysfunction. Reduced accuracy for event detection was related to diminished event recall, memory for event sequence and Theory of Mind abilities. We hypothesized that difficulties with event detection result from a deficit disrupting the on-line processing of kinematic features and physical changes of dynamic human actions. An impairment at the earlier stages of the event encoding process might contribute to deficits in episodic memory and social functioning in individuals with ASDs.

Keywords Action parsing · Intention · Episodic memory · Biological motion

Introduction

Autism is a neurodevelopmental disorder characterised by difficulties in social interaction, communication, as well as a restricted repertoire of interests and repetitive stereotyped behaviours (American Psychiatric Association 2000). An extensive literature has focused on a diminished social functioning, leading to the hypothesis that specific impairments in social cognition constitute the core features of autism spectrum disorders (ASDs) (e.g., Baron-Cohen 1995; Baron-Cohen et al. 1985, 1986; Frith 1989; Happé and Frith 1996). Using false belief tasks, previous studies have shown deficits in Theory of Mind (ToM), i.e., the ability to attribute beliefs and other mental states to oneself and to others, in individuals with ASDs, along with a preserved ability to understand goals and intentions (Baron-Cohen et al. 1986). Such an account predicts impairments only for belief reasoning, with goal and intention processing being subserved by preserved distinct cognitive systems. Nevertheless, subtle impairments in ASDs sometime extend beyond belief reasoning, to include difficulties with the attribution of goals and intentions (Phillips et al. 1998; D’Entremont and Yazbek 2007), recognition of human biological movements (Blake et al. 2003), anticipation (Cattaneo et al. 2007) and imitation of gestures of other people (Dewey et al. 2007; Smith and Bryson 1994; Williams et al. 2001).
Despite a considerable literature on disturbances in action execution and action understanding in individuals with ASDs (Hughes 1996; Russell and Jarrold 1998, 1999; Smith and Bryson 1994; Theoret et al. 2005; Zalla et al. 2006, 2010; Cattaneo et al. 2007; Fabbri-Destro et al. 2009; Stoit et al. 2011), the nature of these impairments is still controversial. Questions have also been raised as to whether these deficits reflect conceptual impairments, motor disturbances, perceptual, sensory-motor integration (Smith and Bryson 1994; Vanvuchelen et al. 2007) or a disorder in perception–action coupling due to a dysfunction of the mirror neuron system (Oberman and Ramachandran 2007).

So far, no studies have assessed the ability to detect discrete meaningful events during perception of dynamic action in individuals with ASDs. Impairments in attentional or perceptual processes occurring the first stages of the action encoding process might crucially contribute to deficits in high-level cognitive functions. If the perception process of action encoding does not take place correctly, it might originate a cascading effect affecting action memory, planning, imitation and theory of mind.

The understanding of other’s behaviour depends mostly on our ability to infer goals, motives, beliefs, and desires from the observed ongoing actions. Humans segment a flow of dynamic action into discrete events: when simply observing everyday behaviour, perceivers extract and identify distinct meaningful component events from the continuous flow of information, by means of a spontaneous and automatic process of segmentation (Newtson 1973, 1976; Baldwin and Baird 2001; Zacks and Tversky 2001; Zalla et al. 2003, 2004). Events can be encoded as discrete units at fine-grained and coarse-grained levels, as goal-directed events are hierarchically organized by goals and subgoals, with groups of fine-grained units clustering into larger event chunks (Hard et al. 2006). Previous research has revealed remarkable intersubjective agreement and reliability (Newtson 1976) across and within observers on locations of boundary breakpoints, on both recall and online detection of dynamic action units (Newtson 1973; Newtson et al. 1977).

Action segmentation is crucial for action comprehension of both familiar and novel combinations of actions, as well as for the continual anticipation of upcoming perceptual information (Schütz-Bosbach and Prinz 2007). Furthermore, this segmentation forms a basis for later memory (Zacks and Tversky 2001). Events extracted during ongoing activity as large ‘chunk’ units enable the viewer to maintain on-line compact representations of extended sequences of acts by decreasing working memory demand, and to store this information in long-term memory for later retrieval (Sirigu et al. 1995; Zacks and Tversky 2001). Further evidence in favour of this chunk processing comes from neurophysiological studies showing that during passive viewing of movies, selected brain areas transiently increase in activity at those moments that observers identify as event boundaries (Speer et al. 2003; Zacks et al. 2001, 2006). Previous research using the event segmentation procedure has shown that the perception of large meaningful constituent units is impaired in patients with frontal lobe lesions (Zalla et al. 2003) and patients with schizophrenia (Zalla et al. 2004), suggesting that conceptual action knowledge, schematically represented in long-term memory as ‘scripts’, might be selectively disrupted.

The processes by which observers automatically segment dynamic human action are, however, not well established. Previous studies have shown that judgments about event boundaries tend to coincide with the analysis of the actor’s intentions suggesting that ‘top-down’ conceptual knowledge about overall goals and intentions might facilitate the action structure detection (Baldwin and Baird 2001; Hard et al. 2006; Zacks 2004). However, there is also substantial evidence showing that when the activity being segmented is less familiar or not conceived as goal-directed, event parsing can be predicted from low-level movement cues (Zacks 2004; Hard et al. 2006). In these circumstances, the detection of both fine and coarse event breakpoints might be related to physical changes, ranging from kinematic (i.e., the actor’s body position and movement) (Newtson et al. 1977; Zacks et al. 2009) to movement parameters (i.e., acceleration of the objects, and their location relative to one another) (Zacks 2004; Hard et al. 2006).

Baldwin et al. (2008) suggested that knowledge of sequential probabilities works conjointly with conceptual inferential knowledge to assist adults in identifying meaningful constituent units. Sequential probabilities arise because some small-scale events are causally linked in achieving a goal and, thus, co-occur more frequently than others in a given context. When prior conceptual knowledge of intentions, goals, and causes is not available, low sequential probabilities across adjacent small-scale motion elements predict boundaries between distinct higher-level actions. In this case, a ‘bottom-up’ process, sensitive to the concomitant statistical regularities within a novel string of small-scale acts, enables observers to cluster small-scale segments into relevant higher-level action bundles.

A previous study using a picture arrangement task showed that children with ASDs had difficulties constructing scenarios representing sequences of goal-directed actions, while they exhibited a preserved ability to arrange mechanical-physical events (Zalla et al. 2006). Despite these difficulties, they were able to identify the overall goal of the action sequence. Interestingly, the greater number of sequence errors occurred predominantly midway in the
action sequences, supporting the hypothesis that impairment affects the ability to represent the internal structure of the action knowledge, e.g., the causal relationship between the component events and the overall goal. In a subsequent study (Zalla et al. 2010), we showed that children and adolescents with ASDs also exhibited difficulties in predicting the most likely outcome of a sequence of goal-directed actions and their performance did not improve when prediction of goals involved familiar actions belonging to their personal motor repertoire. When looking more closely at the type of errors, they predominantly chose temporally preceding sequence events, revealing difficulties with on-line prediction of the action sequential constituents.

So far, no study has investigated whether difficulties with action understanding might arise at the earlier stages of event encoding in ASDs. The current study addressed the issue of whether ASDs lead to difficulties with the encoding of the event knowledge by using a task in which participants were required to segment two movies showing an actor conducting everyday familiar activities. As the viewer’s attention to different levels of events can be oriented by instructions (Newtson 1973), while watching the film, participants were asked to detect meaningful events at two different levels of categorisation: (1) small-scale action units, in fine segmentation, and (2) large-scale action units, in coarse segmentation. Subsequently, to assess their ability to memorize the events belonging to each scenario, they were asked to recall in as much detail as possible everything they remembered about the movie in the correct sequence order. In a recognition task, they were required to decide whether some described actions had occurred in the scenario. As previously shown (Hanson and Hirst 1989, 1991), orienting people’s attention to the fine-grained structure of an activity improves recall memory performance. Specifically, Newtson and Engquist (1976) reported that after viewing a movie of an ongoing activity, participants remembered still pictures taken from event boundaries better than still pictures from the middle of events. Difficulties in event segmentation might thus explain diminished episodic and autobiographical memory often documented in this population (Crane and Goddard 2008; Lind and Bowler 2010).

Based on previous evidence of impaired processing of biological motion among persons with ASDs, we predicted difficulties with on-line detection of the action event structure at both levels of segmentation along with a reduced performance to recall event units, as compared to individuals with typical development and with moderate mental retardation or learning disabilities. This group of participants was included to ascertain that such impairments would be specific to this disorder and not related to a general cognitive disability. In addition, to assess whether these difficulties are related to social impairments, we performed correlation analyses between accuracy measures and scores measuring ToM abilities using two false belief tasks. The basic sensitivity to meaningful structure of dynamic action is a crucial prerequisite for later developmental emergence of several cognitive functions including episodic memory and a full-fledged intentional understanding.

### Methods

#### Participants

The demographic and clinical information about the three groups is displayed in detail in Table 1.

Fifteen individuals (14 males and 1 female) with autism spectrum disorders (ASDs), thirteen individuals (11 males and 2 females) with moderate mental retardation or learning disabilities (MLDs), and forty individuals (34 males and 6 females) with typical development (TD)

### Table 1 Demographic and clinical characteristics of the groups (years; months)

| Group | Participants | Chronological age | Full scale IQ | Verbal IQ | Performance IQ | Global mental age \(a\) | CARS-T | False belief tests |
|-------|--------------|-------------------|---------------|-----------|----------------|--------------------------|--------|-------------------|
| ASD (n = 15) | 158 (5.1) | 8.8–22.1 | 61.27 (7.77) | 52–77 | 60.93 (13.35) | 9.76 (3.60) | 31.9 (5.1) |
| MLD (n = 13) | 15.7 (7.9) | 7.1–28.4 | 58.08 (7.84) | 48–76 | 60.77 (8.54) | 8.95 (4.32) |
| TD (n = 40) | 15.5 (5.1) | 7.8–23.4 | – | 48–80 | – | 4–18 |

\(a\) As calculated from WISC data by the equation \(\text{MA} = \frac{\text{CA} \times \text{IQ}}{100}\)
participated in this study. Children and adolescents with ASDs were recruited at ITTAC (Institut de Traitement des Troubles de l’Affectivité et de la Cognition, Vinatier, Villeurbanne), a specialised clinical service for autism spectrum disorders.

The diagnosis based on DSM-IV criteria was made by a qualified child psychiatrist or paediatric neurologist using different sources of information, including an extensive standardised psychological evaluation, clinical observation, parent interviews concerning the child’s social, emotional, and behavioural functioning, review of autistic symptoms and developmental history, prior evaluations, and preschool and school records. Interviews with parents using the ADI-R (Autism Diagnostic Interview, Lord et al. 1994, French translation: Plumet et al. 1994) confirmed the diagnoses: elevated scores indicated problematic behaviour in the three following areas: reciprocal social interaction, communication and stereotyped behaviours. The cut-off points for the three classes of behaviour are reciprocal social interaction 10, communication 8, and stereotyped behaviours 3, respectively. All participants scored above the cut-offs points. In order to rate the severity of autism symptoms, the Childhood Autism Rating Scale (CARS; Schopler et al. 1988, French translation: Pry and Aussilloux 2000) was completed for each participant. Intellectual functioning was assessed by using the verbal and performance scores of one of the Wechsler Intelligence Scales (Wechsler Intelligence Scales for Children—Third or Fourth Edition: Wechsler 2003; Wechsler Adults Intelligence Scales—Third Edition, 1997) (Table 1).

Participants with moderate mental retardation or learning disabilities (MLDs) were recruited from the Ecole Louis Armand, a specialised school in Villeurbanne. All completed the Wechsler Intelligence Scales for Children—Fourth Edition (Wechsler 2003). Participants with TD, with no history of neurological or psychiatric disturbances, or learning difficulties, were recruited from several primary schools and colleges in Villeurbanne. They were matched with the two clinical groups for global mental age, and we assumed that their mental age corresponded to their chronological age.

The three groups did not differ in chronological age: participants with ASDs were matched with both participants with TD \((t(53) = 0.19; p = 0.84)\) and MLDs \((t(26) = 0.04; p = 0.96)\); participants with MLDs matched with participants with TD \((t(51) = 0.19; p = 0.92)\). The group with ADSs and MLDs were matched for IQ level (Full-scale: \(t(26) = 1.07, p = 0.29\); Verbal: \(t(26) = 0.038, p = 0.97\); IQ: \(t(26) = 1.44, p = 0.16\)).

We evaluated ToM abilities by using two false belief tasks: the Sally and Ann task (Baron-Cohen et al. 1985) and the Smarteries task (Perner et al. 1989). Eight of the fifteen participants with ASDs passed both false belief tests, while four participants passed only one of the two tests, and three of them failed on both tests. One participant with MLDs failed at both false belief tasks, one of them passed just one, and eleven passed both tasks. All participants with TD passed the Sally and Ann and the Smarteries tasks.

The present research has been approved by the local Ethical committee. All participants or parents of participants gave informed, signed consent for participation to this study, and the investigation has been conducted according to the principles expressed in the Declaration of Helsinki.

Materials and Procedure

Participants were presented with two soundless video scenarios showing an actor performing familiar everyday activities on a colour screen. The movies were in DV format \((720 \times 576\text{ pixels}, 25\text{ images per second})\) and were presented in an \(11.8 \times 14.8\text{ cm}\) window in the centre of a grey screen on PC computers with 17-in. (43.18-cm) CRT monitors, with the participants seated in a chair at a comfortable distance (approximately 60 cm). According to the general concept of a script (Schank and Abelson 1977), the scenes represented typical events following a natural temporal order.

The first movie was an 87 s long scenario presenting a woman brushing her teeth (“Brushing teeth” scenario; 2,178 images; 1 image/40 ms) as follows: “A woman enters a bathroom. She goes up to the washbasin, takes a toothbrush and a tube of toothpaste. She opens the tube and puts toothpaste on the toothbrush. She closes the tube and puts it on the washbasin. She turns on a tap to wet the toothbrush and brushes her teeth. Then she puts down the toothbrush and picks up a glass that she fills with water. She rinses her mouth out three times and puts the glass back down. She puts her fingers under the water and cleans her mouth. She takes the toothbrush, washes it and puts it back down. She turns off the tap. Finally she takes a towel and dries her hands and her mouth, puts the towel down and goes out of the bathroom.”

The second movie was a 66 s scenario presenting a woman drinking a coke (1,650 images; 1 image/40 ms) as follows: “A woman enters a kitchen and goes up to a refrigerator. She opens the door of the refrigerator, looks inside and takes out a Coke bottle. She closes the refrigerator and puts the bottle on a table. She opens the door of a cupboard, takes out a glass, puts it on the table, and closes the cupboard. She takes the Coke bottle, opens it and pours Coke into the glass, then closes the bottle. She puts the bottle back into the refrigerator. She takes the glass and drinks.”

In order to familiarise the participants with the segmentation task, an initial 73 s movie depicting a man...
setting a table for lunch was used for practice. The camera
was maintained in a fixed head-high perspective without
cuts or camera movement throughout the filming so as to
capture the naturalistic experience of watching an ongoing
activity.

The participants were informed that the purpose of the
experiment was to understand how people perceive and
remember events in everyday situations. They were told
that their memory for events would be tested by presenting
them with a short film after a practice film. Subjects were
asked to mark the end of each event and the beginning of
the following one, by pressing a response button under
three orientation conditions: (1) spontaneous segmentation;
(2) small-event segmentation; and (3) large-event seg-
mentation. Before the start of each film, they were asked
to press the response button as soon as they saw a visual
signal. This first response initiated the program used to
record button presses and the number of event boundar-
ies identified on each viewing. A couple of examples for both
small and large event units were provided before the
practice section. A “small event” is “picking up a fork” in
the “Setting a table” practice movie and a “large event” is
“setting the table”.

The presentation order of the two conditions (large and
small) was counterbalanced across subjects. Half of the
subjects in each group were requested to press for small
events and half were told to press for large events. At the
end of the presentation of each scenario, participants
completed a Free Recall task and a Recognition task con-
sisting of a six (true/false) item questionnaire about the
discrete events that actually occurred (true item) or could
have occurred (false item) in the movies.

To estimate psychomotor slowing in the ASDs and
MLD groups, participants’ reaction times (RTs) were
recorded using a simple response time task, prior to the
experimental sessions. Participants were instructed to press
a key as quickly as possible when a visual stimulus
appeared in the centre of the computer monitor. A total of
21 trials were performed. A t test revealed that the ASD
group (median = 559.8 ± 312.8; t = 5.11; p < 0.0001)
and the MLD group (median = 442.5 ± 208.1; t = 3.98;
p = 0.0002) were both significantly slower than the TD
group (median = 301.3 ± 52.7), while they did not differ (t = 1.15; p = 0.26). We calculated individual slowness
with respect to the median RTs of the normal control
group. Thus, in order to partial out the effect of psycho-
motor slowing, the median values in individual response
time increment so obtained were subtracted from the time
taken to detect the action boundaries. The corrected reac-
tion times were used to determine the breakpoints detected
and to estimate the mean individual temporal distance from
the boundaries. This correction was justified by the greater
variability in RTs in participants with ASDs and MLDs.

Previous studies have reported that IQ is related to the
speed of information processing, with the lower IQ being
related to slower RTs among adults and children with and
without intellectual disability (Kail 1992; Grudnik and
Kranzler 2001). Slower performance in individuals with
mental retardation have been explained by either an
impaired general central processing system that affects all
aspects of cognitive functioning or a particular cognitive
process (e.g., stimulus encoding, response selection or
execution, or attentional mechanisms). Furthermore,
Brewer and Smith (1984) showed that individuals with
mental retardation adopt speed-accuracy trade-off strate-
gies and that this contributes to the slower and more var-
iable performance.

Coding Procedure

Similar to a procedure used in a previous study (Hard et al.
2006), ten external judges were asked to rate video frames
for several movement parameters: start, stop, change
direction, orientation or speed, take or leave an object (1 if
a movement change was present, 0 if it was absent) so as to
identify points of maximal changes. Thus, the average
number of total movement changes that occurred for each
frame was calculated for each judge. As boundaries con-
sisted of several images characterised by greater scores in
movement changes, images were aggregated into 1-s
interval. Hence, boundary breakpoints for small and large
action units were identified by asking the judges to detect
transitions between events (the end of the episode and the
beginning of the following one) at the fine and course unit
levels. Only breakpoints (i.e., aggregations of 25 images)
on which the consensus among judges, through the press
responses procedure, was equal or superior to 70 % were
regarded as boundaries and constituted a ‘prototypical
script’. Since boundaries do not systematically correspond
to large movement changes, the subsequent parsing pro-
cedure allowed selecting those breakpoints that were per-
ceived as event boundaries even though changes were not
maximal. Overall, boundary breakpoints corresponded to
significantly greater movement changes (M = 3.51, sd: 0.87)
as compared to non-breakpoints (M = 2.10, sd: 0.77)
t(56) = 8.30; p < 0.0001).

Thirty-six boundaries in the small-oriented condition,
and ten in the large-oriented condition were retained for the
“Brushing teeth” scenario; twenty-one boundaries in the
small-oriented condition and seven in the large-oriented
condition were retained for the ‘Drinking coke’ scenario.
The detection criterion consisted in inclusion of any
response that fell within the transition between events, as
defined on the basis of the above described procedure.
Data Collection

Performance of participants was monitored by two experimenters. The following variables were analysed on the two scenarios:

**Segmentation.** Mean number of presses and mean event unit length were calculated under the three conditions (spontaneous, small and large). For each participant, the mean length of the event units identified was calculated by dividing the length of the movie by the number of presses.

**Boundary Detection.** Accuracy measures were computed as hit rates (proportion of correct boundaries identified out of the total number of presses), false alarm rate (proportion of incorrect boundaries identified out of the total number of presses), sensitivity index (d') and B response bias (beta values) under the small-oriented and large-oriented condition instructions. As described above, a prototypical script for each condition—*small* and *large*—was established on the basis of the judgement provided by the external judges who were blind to the purpose of the study.

**Temporal Distance from the Prototypical Boundaries.** We calculated the temporal distance from each boundary breakpoint in the sequence. As explained above, in order to eliminate the effect of the psychomotor factor affecting individuals with mental retardation, the individual increment in the response time was subtracted from the time taken to respond.

**Free Recall Test.** After viewing each movie, participants were asked to verbally report what they had seen in as much detail as possible, and in the same temporal order as that presented. The number of events correctly identified and their order of recollection were recorded by the experimenters and participants’ responses were scored by three independent raters (Inter-rater reliability: r = 0.90). No time limitation was imposed.

**Recognition Test.** After the Free Recall task, subjects were given a 6-item questionnaire. They were told that for each item they should decide whether the described actions had occurred in the movie. The number of correct responses was recorded.

Data Analysis

Statistical analyses included repeated measures ANOVA, with group (ASD, MLD, TD) and condition (type of segmentation: spontaneous, small, large) as between-participants independent variables. Pearson correlation analysis, and unpaired T Test. The Scheffé test was used for post hoc comparisons.

Accuracy measures were computed as a proportion of correct boundaries identified out of the total number of presses produced by each participant for each condition in the two scenarios. Data were analysed using the indices of item discrimination D’ and response bias B. The sensitivity index (D’) was computed from measurements of the hit rates (i.e. correct choice of boundaries) and false alarms (i.e. incorrect choice of response) as follows: $d' = z(FA) - z(H)$, where FA is the false alarm rate and H is the hit rate.

According to this procedure, a d’ value equal to 1.0 represents maximum accuracy; a value of 0.5 indicates chance-level performance. B values greater than zero suggest the tendency to respond ‘yes’; B values less than zero represent a conservative bias, i.e., tendency to respond ‘no’. In both cases, the greater the B score, the greater the bias. Proportions were subjected to arcsine transformation in order to meet criteria for parametric analyses. For these statistics, the alpha level for acceptance was set at 0.05.

Results

**Segmentation.** Repeated measures ANOVA yielded highly significant effects of group (F(2,65) = 7.7; p = 0.001) and condition (F(1,65) = 364.8; p < 0.0001), as well as a significant Group X Condition interaction (F(2,65) = 14.3; p < 0.0001). Overall, the instructions modulated the number of presses. The condition effect was due to the lower number of presses produced in the large orientation condition as compared to both the spontaneous (mean difference = 10.8; p < 0.0001) and the small orientation (difference = 14.1; p < 0.0001) conditions across the three groups. The number of presses in the small orientation condition was also significantly greater that in the spontaneous orientation condition (mean difference = −3.4; p < 0.0001).

The group difference was due to participants with TD producing a greater number of presses than participants with ASDs (mean difference = −2.6; p = 0.001) while participants with MLDs did not differ from the TD (mean difference = −0.95; p = 0.61) nor from the ASD (p = 0.26) groups. The interaction effect was due to the lower number of presses produced by the group with ASDs in the small oriented condition in comparison to both the TD (p < 0.0001) and MLD (p = 0.008) groups, while the TD and the MLD groups did not differ (p = 0.25). In the large oriented condition, the group with ASDs produced a lower number of presses than the TD group (p = 0.02) while the MLD group did not differ from the ASD nor from the TD groups (p > 0.05). The group with ASDs also produced a lower number of presses than the TD group in the spontaneous condition (mean diff. = −3.5; p = 0.005) while they did not differ from the MLD group (p = 0.71). The difference between the MLD and the TD did not reach significance (p = 0.09) (Table 2).
For mean event length, we found a main effect of condition \((F(2, 65) = 196.6; p < 0.0001)\) and significant interaction group \(\times\) condition effect \((F(1, 65) = 4.5; p = 0.01)\). Event unit length was significantly longer for the large segmentation than for the small one \((\text{mean difference} = -11.3, p < 0.0001)\). The significant interaction was due to ASD group identifying longer event units in the small oriented segmentation condition than both the MLD \((p = 0.003)\) and the TD \((p < 0.0001)\) groups, while difference between the MLD and TD groups was not significant \((p = 0.29)\), nor were any group differences in the large oriented segmentation condition \((p > 0.05)\). Within each group, there was a variability in the mean length of event units; for the small and large conditions respectively, ASD participants ranged from 3.5 to 9.5 s, and from 7.6 to 22.7 s, MLD participants ranged from 3.1 to 7.2 s and from 7.6 to 25.5 s, and TD participants ranged from 2.7 to 5.1 s and from 10.1 to 31 s (Table 2).

Boundary Detection

**Hit Rate.** The repeated measures ANOVA on the proportion of detected boundaries for the two scenarios revealed significant main effects of group \((F(2, 65) = 22.3; p < 0.0001)\) and condition \((F(1, 65) = 30.5; p < 0.0001)\) and a significant Group \(\times\) Condition interaction \((F(2, 65) = 12.5; p < 0.0001)\) (Fig. 1). Overall, the TD group performed significantly better that the ASD group \((\text{mean difference} = -0.18; p < 0.0001)\) and the MLD \((\text{mean difference} = -0.1; p = 0.005)\) groups. The group with MLDs performed slightly better that the group with ASDs, but the difference did not reach significance \((\text{mean difference} = -0.08; p = 0.07)\). The interaction effect revealed that the ASD group detected fewer boundaries than both groups with TD \((\text{mean difference} = -0.11; p = 0.002)\) and with MLDs \((\text{mean difference} = -0.11; p = 0.02)\) under the small oriented condition, while no difference was found between MLD and TD \((\text{mean difference} = -0.08; p = 0.99)\). In the large oriented condition, the TD group performed significantly better than both participants with ADSs \((\text{mean difference} = -0.25 p < 0.0001)\) and MLD \((\text{mean difference} = -0.20; p < 0.0001)\) while the difference between the ASD and MLD was not significant \((\text{mean difference} = -0.05; p = 0.53)\) (Table 2; Fig. 1).

Participants identified significantly more large boundaries that small ones \((\text{mean difference} = -0.06; p = 0.0006)\). However, the interaction effect specified that

### Table 2

| Groups | ASD | MLD | TD |
|--------|-----|-----|----|
| Number of presses | Spontaneous: 14.30 (4.39) | Spontaneous: 15.38 (3.17) | Spontaneous: 17.86 (3.17) |
| Small: 15.50 (4.61) | Small: 19.84 (3.60) | Small: 21.73 (3.07) |
| Large: 7.03 (2.64) | Large: 6.04 (2.53) | Large: 5.26 (1.58) |
| Event length (sec) | 5.52 (1.71) | 4.17 (1.12) | 3.66 (0.52) |
| Small: 13.21 (5.14) | Large: 15.33 (6.39) | Large: 16.35 (5.39) |
| Hit rates | 0.65 (0.15) | 0.76 (0.09) | 0.76 (0.08) |
| Large: 0.52 (0.11) | Large: 0.57 (0.19) | Large: 0.77 (0.11) |
| False alarms | 0.29 (0.11) | 0.22 (0.11) | 0.20 (0.06) |
| Large: 0.41 (0.16) | Large: 0.37 (0.18) | Large: 0.21 (0.09) |
| D’ | 1.07 (0.83) | 1.63 (0.73) | 1.66 (0.51) |
| Large: 0.30 (0.62) | Large: 0.66 (1.24) | Large: 1.85 (0.72) |
| Temporal distance from boundaries | -1.10 (0.31) | -0.94 (0.17) | -0.91 (0.13) |
| Large: -0.89 (0.34) | Large: -1.28 (0.36) | Large: -1.02 (0.26) |
| Free recall (number of events) | 8.07 (3.01) | 8.8 (3.78) | 13.36 (2.62) |
| Recognition (number of events) | 5.9 (0.2) | 5.8 (0.33) | 5.9 (0.16) |
this was due to ASD and MLD detecting significantly more boundaries on the small condition relative to the large one ($p < 0.01$), while TD identified a comparable proportion of correct boundaries across the two conditions (mean difference $= 0.01$; $p = 0.71$).

**False Alarm Rate.** The repeated measures ANOVA on false alarm rate showed highly significant effects of group ($F(2,65) = 18.9; p < 0.0001$) and condition ($F(1,65) = 22.5; p < 0.0001$) and significant Group X Condition interaction ($F(2,65) = 6.4; p = 0.003$). Both ASD and MLD group performed worse than the TD ($p < 0.0001$ and $p = 0.003$, respectively), while the ASDs and the MLD groups did not differ ($p = 0.29$).

When we conducted group comparisons for each condition separately, participants with ASDs exhibited a greater false alarm rate in the small ($p = 0.005$) and large conditions ($p < 0.0001$), as compared to participants with TD, while they did not differ from participants with MLDs (all $p > 0.05$). MLD group’s performance was significantly lower relative to TD group in the large oriented condition ($p = 0.0007$), but not in the small oriented one ($p = 0.69$) (Table 2). Overall, false alarm rate was greater in the large oriented condition than in the small one ($p = 0.0009$). However, within group comparisons revealed that this was the case for the ASD ($p = 0.02$) and the MLD ($p = 0.02$) groups, while false alarm rate for participants with TD was similar across conditions ($p = 0.61$).

**D’ (Sensitivity) Index.** The repeated measures ANOVA on D’ (sensitivity) index revealed significant main effects of group ($F(2,65) = 19.7; p < 0.0001$) and condition ($F(1,65) = 17.6; p < 0.0001$), as well as a group by condition interaction ($F(2,65) = 11.6; p < 0.0001$). D’ was greater in participants with TD, relative to the group with ASDs ($p < 0.0001$) and MLD ($p = 0.008$), while the ADS and MLD groups did not differ ($p = 0.12$). However, under the small condition $d’$ was lower for participants with ASDs relative to both participants with TD (mean diff. $= -0.60; p < 0.01$) and MLD (mean diff. $= -0.57; p < 0.05$), whereas the latter two groups did not differ (mean diff. $= -0.03; p = 0.99$). Under the large condition, $d’$ was significant higher for participants for TD participants relative to both participants with ASDs ($p < 0.0001$) and MLDs ($p = 0.0001$), while participants with ASDs and those with MLDs did not differ ($p = 0.51$).

*Overall, $d’$ was greater in the small condition than in the large one (mean difference $= 0.24; p = 0.03$). However, the significant interaction specified that this was the case only for the groups with ASDs (mean difference $= 0.76; p = 0.008$) and MLDs (mean difference $= 0.97; p = 0.02$), but not for TD group (mean difference $= -0.19; p = 0.17$).*

**B (Bias) Values.** The Repeated measures ANOVA on B response bias showed no main effects of group ($F(2,65) = 0.4; p = 0.65$) and condition ($F(1,65) = 0.006; p = 0.93$), nor group X condition interaction ($F(2,65) = 0.31; p = 0.73$). The three groups showed a similar positive bias, i.e. a tendency to respond that the boundary was presented.

**Temporal Distance from the Prototypical Boundaries.** We observed a significant main effect of group ($F(2,65) = 3.86; p = 0.02$) and a significant Group X Condition interaction ($F(1,65) = 7.54; p = 0.001$). There was no significant effect of condition ($F(2,65) = 2.23; p = 0.14$). Post-hoc Scheffé tests indicated that under the small oriented condition the temporal distance was greater in participants with ASDs compared to TD ($p = 0.007$) while the difference between the ASD and the MLD groups did not reach significance ($p = 0.09$). Under the large condition, participants with MLDs exhibited a significantly longer temporal distance relative to both TD ($p = 0.026$) and the ASD ($p = 0.004$) groups while the latter ones did not differ ($p = 0.4$) (Fig. 2).

**Recognition Task.** Unpaired T Test revealed that participants with TD recalled a significantly greater number of events as compared to participants with ASDs ($t(52) = -6.23; p < 0.0001$) and participants with MLDs ($t(50) = -4.83; p < 0.0001$). No significant difference was observed between the ASD and the MLD groups ($t(26) = -0.56; p = 0.58$).

The number of events generated out of order was significantly greater for participants with ASDs ($M = 0.33, SD 0.55$), relative to those with TD ($M = 0.98, SD: 0.22, p = 0.033$), while their performance did not differ from that of participants with MLDs ($M = 0.38, SD: 0.41; p = 0.79$). The MLD’s performance was lower than that of the TD group ($p = 0.002$) (Table 2).

**Fig. 2** Temporal distance (in sec.) from the prototypical boundaries under the small and the large conditions for the three groups (ASD, MLD, TD). * $p < 0.05$; ** $p < 0.0001$
with MLDs performed significantly lower than participants with TD ($p = 0.02$) (Table 2).

**Correlation Analyses.** We computed correlation analyses, using the Pearson Product Moment test, between hit rates and the number of events recalled under the two segmentation conditions. Significant correlations between hit rates and the number of events recalled were found in the three groups (ASD: $r = 0.52$, $z = 1.97$; $p < 0.05$; MLD: $r = 0.61$, $z = 2.21$; $p = 0.03$; TD: $r = 0.39$, $z = 2.47$; $p = 0.01$) in the small segmentation condition. In the large segmentation condition, hit rates and the number of events recalled correlated only in participants with ASDs ($r = 0.62$, $z = 2.52$; $p = 0.01$) (Fig. 3).

**Additional Analyses.** To assess whether difficulties in event segmentation was related to the ASD group’s performance on the ToM tests (Sally & Ann and the Smarties), we compared the group of 8 ASD subjects who passed the two ToM tests (ToM$^+$) and the group of 7 ASD subjects who failed at least one of the two tests (ToM$^-$) on accuracy performance (D’ index). The mean D’ was greater for the ToM$^+$ group than for the ToM$^-$ group under the small (Mean difference = 0.81; $t(13) = 2.1$; $p = 0.05$) and large (Mean difference = 0.84; $t(13) = 3.5$; $p = 0.004$) segmentation conditions.

**Discussion**

The aim of the present study was to investigate the ability of a group of children and adolescents with ASDs to parse continuous flows of dynamic action into meaningful event units. During perception of the two movies, we monitored participants’ attention to the fine and coarse levels of the action structure. Previous research have demonstrated that instructions affect how information is encoded in episodes by orienting subjects’ attention towards different levels of the action knowledge structure, showing that during on-line perception the observer encodes and organises event information hierarchically (Newton 1973; Black and Bower 1979; Hanson and Hirst 1989; Zacks et al. 2001). The present results reveal that when participants were required to detect transitions between one action and another, they identified significantly more units at the fine-grained event level relative to the coarse-grained event level condition, confirming that performance was modulated by instructions and that participants were able to allocate attentional resources throughout the tasks.

However, when we compared the groups’ performance under specific conditions, participants with ASDs exhibited lower number of presses and longer event length than both control groups in the fine-grained action segmentation. The diminished performance in participants with ASDs was evidenced by the lower proportion of boundaries correctly identified, the greater proportion of false alarms, and the lower d’ sensitivity index, as well as the longer temporal distance from the boundaries under the fine-grained action segmentation, relative to the two control groups, despite a similar bias across the two segmentation conditions. In contrast, participants with MLDs exhibited a low level of performance in terms of hit rates, false alarms, D’ sensitivity index and a longer temporal distance from the boundaries only under the course-grained segmentation condition, revealing specific difficulties in detecting large action boundaries. Hence, although participants with ASDs showed a general impairment in action segmentation, they exhibited specific difficulties in parsing fine-grained actions. Indeed, accuracy was lower in the course-grained segmentation for both ASD and MLD groups, relative to the TD group. They also showed poor performance in the free recall task, in terms of fewer events recalled and the greater number of errors for sequential order, as compared to participants with TD.

Remarkably, the number of recalled events correlated positively with the proportion of hit rate under the fine-grained oriented condition in all participant groups. This finding is consistent with previous studies showing that orienting people’s attention to the fine-grained structure of an activity plays a critical role in event encoding by improving recall memory performance (Hanson and Hirst 1989, 1991; Zacks and Tversky 2001), while its effect on recognition memory remained unclear (Lassiter et al. 1988; Lassiter and Slaw 1991). It is possible that impairments at the early stage of action event encoding might affect episodic memory recollection in this population. As documented by previous studies, diminished performance on
measures of episodic and autobiographical memory was often reported in ASDs (Crane and Goddard 2008; Lind and Bowler 2010). Interestingly, a more recent study suggested that inefficient initial information encoding and organisation, rather than storage and retrieval, are the primary factors that limit episodic memory in ASDs (Southwick et al. 2011). The present findings also provide evidence that impairment in event detection might be related to ToM abilities. Indeed, the sub-group of ASD participants who obtained lower scores on the two False Belief tasks identified fewer small and large events than the group of ASD participants who succeeded both False Belief tasks.

Event parsing is considered one of the fundamental mechanisms by which human cognition collects, organises and stores the large variety of information extracted from the environment. Developmental studies indicate that 9–11 month old infants spontaneously parse ongoing behaviour along boundaries that coincide with the initiation and completion of the actors’ intentions (Baldwin and Baird 2001; Baldwin et al. 2001; Saylor et al. 2007). Event Segmentation Theory claims that observers build internal event models of the current situation by integrating information over the recent past to generate predictions about future perceptual input (Zacks et al. 2007). Event segmentation results from the continual anticipation of future outcomes; it serves the encoding of events in long-term memory, the updating of working memory, and the learning of new procedures.

Previous research provided consistent evidence that events are spontaneously encoded at multiple timescales and hierarchically organised (Zacks and Swallow 2007). Although event parsing is measured by asking participants to identify event boundaries, which separate natural and meaningful units of action stream explicitly (Newtson 1973), functional neuroimaging data has revealed that activity in brain regions including temporo-parietal cortex and lateral frontal cortex increases transiently at event boundaries when naïve observers passively view ongoing actions (Speer et al. 2003; Zacks et al. 2001, 2006). Nevertheless, the processes underlying dynamic action parsing are still a controversial issue.

It has been argued that fine and course segmentations might rely on partially distinct mechanisms reflecting two kinds of chunking modes: a low level automatic perceptually driven procedure that operates directly at the earlier stages of the encoding of kinematic features of the movement, and a high level procedure, in which goals and course-grained action units are constructed using inferential mechanisms sensitive to knowledge about human behaviour and situational context, under the subjects’ deliberate control (Gobet et al. 2001).

Interestingly, previous studies have shown that neurological and psychiatric disorders associated with frontal cortex dysfunction might selectively impair coarse-grained unit segmentation, along with a relatively intact ability to perceive fine-grained units (Zalla et al. 2003, 2004). Along the same line, it has been suggested that the processing of movement features plays a major role in segmenting fine-grained units, whereas coarse segmentation might depend more strongly on conceptual features (e.g., Zacks and Tversky 2001).

However, there is also consistent evidence that physical discontinuities in the behaviour stream cue both coarse and fine action boundaries with the coarse boundaries corresponding to greater changes in movement features, as compared to the finer ones (Hard et al. 2006, 2011). Infants perceive certain combinations of movements, such as reaching for and grasping an object as a unified act (Woodward 1998) and segment continuous behaviour according to points of major goal completion (Baldwin et al. 2001).

Although both individuals with ASDs and MLDs exhibited problems with the action parsing, it is likely that the disrupted underlying mechanisms would be somewhat different in the two groups. Previous results indicated that children with mental disability might be able to make basic perceptual discriminations along with impairments in the perception of more complex visual motion cues (Virji-Babul et al. 2006). In accordance with this view, the inability to parse large chunk units of events in our participants with MLDs might be related to the reduced effect of top-down explicit conceptual knowledge during on-line action perception. The fact that ASD participants are impaired at both fine and coarse-grained levels of action segmentation rather supports the hypothesis of a general deficit in processing distinctive movement features and physical changes within human action. Indeed, previous studies among children with ASDs have documented difficulties with a large variety of false belief tasks and pretend play, but no major difficulties have been reported in understanding of others’ goals, desires and intentions (Baron-Cohen et al. 1986; Carpenter et al. 2001; Hamilton 2009; Zalla et al. 2006, 2010), though this knowledge might be sometimes acquired by using compensatory cognitive strategies (Fabbri-Destro et al. 2009; Boria et al. 2009).

Interestingly, perception of biological motion is often compromised in people with ASDs (Blake et al. 2003; Freitag et al. 2008; Gepner and Mestre 2002; Klin et al. 2009; Cook et al. 2009) and consistent evidence supports the notion that difficulties in understanding others’ goals in ASDs often arise when this information has to be inferred from the kinematic features of movements (Blake et al. 2003; Schmitz et al. 2003; Klin et al. 2009; Cook et al. 2009). Gepner and Mestre (2002) has suggested that impairments in biological motion perception among...
individuals with ASDs could be related to difficulties with on-line processing of rapid movements or changes in biological dynamic action. Overall, enhanced imitative performances have been observed when the facial and body movements to be imitated are slowed down, supporting the notion of the rapid visual-motion perception integration impairment in autism (Lainé et al. 2011). According to this view, an abnormal functional connectivity (i.e. under- or overconnectivity) between several brain areas involved in movement and temporal processing and perception–action coupling, would disrupt the perception of biological motion. This abnormal brain circuit includes the magnocellular pathway, the dorsal stream, the cerebellum, the superior temporal sulcus and the mirror neuron system (see Gepner and Feron 2009).

Theoretical models have been proposed to account for impairments in both action understanding and action execution in ASDs. In particular, internal forward models have suggested that the diminished ability to encode action units might reflect impairments of internal simulation models subserving both action monitoring and action prediction in children and adolescents with ASDs (Wolpert et al. 1995). These models posit that for every motor command generated during movement execution, effector copies are used to predict the sensory consequences of a given motor command, which are then compared to the actual sensory outcome. According to the Common Coding Theory, forward models also allow observers to simulate the actions of others in their own sensory-motor systems (Blakemore and Decety 2001; Blakemore et al. 2002; Wolpert et al. 1995; Rizzolatti and Craighero 2004). Thus, an existing agent’s action repertoire and predictive mechanisms in the motor system that exploit the same efference copy produced by the motor command during action execution serve to understand others’ actions (Frith et al. 2000). Recently, disrupted internal modelling in ASDs has been shown to be responsible for poor performance in joint action and social coordinative behaviour (Stoit et al. 2011).

Interestingly, Boria and collaborators (2009) reported that when children with ASDs failed to use motor information coming from the agent’s hand shape, they based their judgment concerning the agent’s functional knowledge or on contextual or social information present in the environment. On the same vein, it has been suggested that children with ASDs are unable to perceive and organise chains of motor actions via a visual modality, while possessing a relatively preserved inferential mechanism sensitive to explicit knowledge and situational context (Cattaneo et al. 2007; Fabbri-Destro et al. 2009). These results are in agreement with our previous studies showing difficulties with on-line prediction of ensuing actions (Zalla et al. 2010) or with the sequential ordering of events directed to objects, even when the overall goal was correctly identified (Zalla et al. 2006). All together, these findings suggest that even when individuals with ASDs exhibit relatively preserved conceptual action knowledge, they might encounter difficulties in on-line processing of motor action components.

In conclusion, the present study shows that difficulties in action understanding might arise at the earlier stages of event encoding in ASDs. Individuals with ASDs exhibited an overall reduced accuracy in event detection, but their performance was specifically impaired when asked to detect fine-grained action units, as compared to both control groups. Based on previous evidence, we hypothesized that such difficulties might arise from an impaired on-line processing of movements and physical changes underlying boundary detection and event encoding. An impairment at the earlier stages of the event encoding process might contribute to deficits in episodic memory and social functioning in individuals with ASDs.

Alternatively, different lines of research have suggested that impairment in action understanding in ASDs might be due to the inability to process intentional and referential information conveyed by eye gaze, facial expressions or body postures (see Piero et al. 2006; Vivanti et al. 2011). Further research is needed to assess the extent to which difficulties understanding others’ actions in individuals with ASDs are due to a diminished sensitivity to intentional cues or rather reflect a genuine impairment in on-line processing of the kinematic features of human action.

Acknowledgments We gratefully acknowledge the commitment of the participants and their families to the pursuit of research in autism. This research was supported by the Agence Nationale de la Recherche (Grant Programme « Neurosciences, Neurologie et Psychiatrie) to TZ. The authors wish to thank Richard Carter, Jérome Sackur and the anonymous reviewer for their very helpful comments on an earlier draft of this article.

References

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders, 4th edition DSM-IV-TR (Text Revision). Washington, D.C.: American Psychiatric Press.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. Cognition, 106(3), 1382–1407.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic action. Trends in Cognitive Sciences, 5, 171–178.

Baldwin, D., Baird, J.,aylor, M., & Clark, A. (2001). Infants parse dynamic human action. Child Development, 72, 708–717.

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

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., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. British journal of Developmental Psychology, 4, 113–125.
Theoret, H., Halligan, E., Kobayashi, M., Fregni, F., Tager-Flusberg, Stoit, A. M. B., van Schie, H. T., Riem, M., Meulenbroek, R. G. J., Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of
Russell, J., & Jarrold, C. (1998). Error-correction problems in autism:
Russell, J., & Jarrold, C. (1999). Memory for actions in children with
Saylor, M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007).
Southwick, J. S., Bigler, E. D., Froehlich, A., Dubray, M. B.,
Smith, I. M., & Bryson, S. E. (1994). Imitation and action in autism:
Childhood
Schank, R. C., & Abelson, R. P. (1977).
Scripts, plans, goals and
Schmitz, C., Martineau, J., Barthelemy, C., & Assaiante, C. (2003).
Motor control and children with autism: Deficit of anticipatory function? Neuroscience Letters, 348(1), 17–20.
Schopler, E., Reichler, R. J., & Renner, B. R. (1988). Childhood autism rating scale. Los Angeles: Western Psychological Services (French translation: Pryn, R., & Aussiloux, C. (2000)).
Schütz-Bosbach, S., & Prinz, W. (2007). Prospective coding in event representation. Cognitive Processing, 8(2), 93–102.
Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals and understanding. Hillsdale, NJ: Erlbaum.
Schmitz, C., Martineau, J., Barthelemy, C., & Assaiante, C. (2003). Motor control and children with autism: Deficit of anticipatory function? Neuroscience Letters, 348(1), 17–20.
Zalla, T., Labruyere, N., & Georgieff, N. (2006). Goal-directed action problems in children and adolescents with autism. Developmental Neuropsychology, 21(2), 287–295.
Zalla, T., Pradat-Diehl, P., & Sirigu, A. (2003). Perception of action boundaries in patients with frontal lobe damage. Neuropsychologia, 41(12), 1619–1627.
Zalla, T., Verlut, I., Franck, N., Puzevat, D., & Sirigu, A. (2004). Perception of dynamic action in patients with schizophrenia. Psychiatry Research, 128, 39–51.