Children with autism spectrum disorder spontaneously use scene knowledge to modulate visual object processing

Lien Van Eylen a,b,c Peter De Graef d,e, Jean Steyaert b,c,f, Johan Wagemans b,d and Ilse Noens a,b,g

a Parenting and Special Education Research Unit, KU Leuven, Leuven, Belgium;

b Leuven Autism Research, KU Leuven, Leuven, Belgium;

c Department of Child Psychiatry, UPC-KU Leuven, Leuven, Belgium;

d Laboratory of Experimental Psychology, KU Leuven, Leuven, Belgium;

e Department of Applied Psychology, Thomas More University College, Antwerp, Belgium

f Department of Clinical Genetics, University Hospital Maastricht, Maastricht, the Netherlands;

g Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, USA

Corresponding author:
Lien Van Eylen
Department of Child Psychiatry
University Hospital of Leuven
Herestraat 49 – box 7003, 3000 Leuven, Belgium
Tel. +32 16/34.25.03 or +32 498/04.60.09.
E-mail address: Lien.VanEylen@ppw.kuleuven.be
Abstract

Several studies have investigated contextual influences on visual object processing in individuals with autism spectrum disorder (ASD) and failed to find reduced context effects. However, these studies did not properly account for local inter-object effects and/or the influence of post-perceptual processes, leaving it unclear whether individuals with ASD display equally large global top-down effects of context, or whether they rely on a more local mechanism producing the same result. In this study, an eye-movement paradigm was used to investigate implicitly induced context effects on visual object processing in children with ASD compared to matched controls. To find out whether the context effects in the ASD group were, to the same extent, due to early top-down influences on object recognition, we also examined the interaction between context effects and the ease of object recognition. Both groups displayed equally large context effects and congruent contextual information facilitated object recognition to the same extent in both groups. This indicates that the context effects in the ASD group did not result from the operation of a more local, less top-down mechanism. These findings contradict predictions based on the weak central coherence account. However, a good alternative to explain all inconsistencies is currently lacking.

Keywords: autism spectrum disorders, weak central coherence, context effects, scene perception, eye-movements.
1. Introduction

Autism spectrum disorders (ASD) are characterized by the co-occurrence of impairments in social reciprocity and communication, combined with restricted and repetitive patterns of interests and activities (American Psychiatric Association, 2000). According to the weak central coherence (WCC) account, some of these symptoms are related to or even caused by their atypical information processing. This account claims that individuals with ASD show reduced central coherence, or global processing, meaning that the normal tendency to process information in its context and integrate information for higher-level meaning is diminished, favoring piecemeal or local processing (Frith, 1989). In the ASD literature, the evidence for increased local processing is strong while that for reduced global processing is more mixed, especially for ASD individuals with (at least) normal intelligence (for a review, see Happé & Frith, 2006). To account for these inconsistencies, the WCC theory was further refined (Happé & Booth, 2008; Happé & Frith, 2006). Firstly, the authors state that diminished global processing is not a ‘core deficit’ but rather a ‘processing bias’ or ‘cognitive style’, which can be overcome when explicitly instructed to do so. Thus, reduced global processing may only become apparent when individuals with ASD are not guided to attend to global information. Accordingly, they encourage the use of open-ended tasks to capture spontaneous modes of processing information. Secondly, they make a distinction between local and global coherence, emphasizing that individuals with ASD may only display diminished global coherence, but intact local coherence. Local coherence refers to item-to-item processing (chaining), or intra-domain coherence, whereas global coherence refers to inter-domain coherence, requiring more top-down modulation of information. Taken together, the revised WCC account suggests that individuals with ASD will mainly show reduced global processing on open-ended implicit tasks requiring global coherence.

One way to test this revised WCC account is by investigating implicitly induced context effects on visual object processing in ASD individuals. In what follows we will first discuss context effects on visual object processing in typically developing (TD) individuals. Afterwards, shortcomings of studies that have investigated these effects in ASD will be mentioned and an alternative paradigm will be presented.

Contextual modulation of visual object processing is demonstrated by numerous studies in TD individuals (for a review, see Bar, 2004; Henderson & Hollingsworth, 1999). A consistent effect found across all of these studies is contextual facilitation of object processing: an object in a plausible scene context is processed faster than that same object in an implausible context. This is evidenced by a shorter naming latency (Boyce & Pollatsek, 1992) and a shorter fixation duration (De Graef, Christiaens, & d’Ydewalle, 1990) for scene-consistent objects. Additionally, eye-movement studies showed that context influences the spontaneous
allocation of attention. One of the first studies investigating this found that scene-inconsistent objects were fixated earlier and more often than consistent objects (Loftus & Mackworth, 1978). The same finding was also reported by Brockmole and Henderson (2008), but overall, these results are quite inconsistent and depend on the specific paradigm used (for a review, see De Graef, 2005a; Henderson & Hollingworth, 1999).

Different models try to account for the scene-context effects on visual object processing (De Graef, 2005b). According to Henderson and Hollingworth (1999), they can be divided into three groups: the perceptual schema model, the priming model and the functional isolation model. The idea underlying each of these models is that perception of a scene leads to the activation of the corresponding scene schema. A scene schema contains generalized expectations about which objects are likely to be in a particular scene and which objects are not (e.g., in a kitchen, a car is normally not present, while a fridge is). Activation of this schema subsequently alters and facilitates the processing of objects present in that scene. However, the specific level at which object processing is influenced differs for each of the models (Henderson & Hollingworth, 1999). According to the **perceptual schema model**, the activation of the scene schema facilitates the first stages of the subsequent perceptual analysis of schema-consistent objects. The **priming model** proposes that the effect of context occurs at a later stage: ‘activation of the scene schema primes the stored representations of schema-consistent objects’ (Henderson & Hollingworth, 1999, p. 261). This reduces the amount of perceptual information necessary to identify a primed object. So, both of these models predict contextual facilitation of object identification, meaning that scene consistent objects will be recognized faster than inconsistent objects. On the contrary, the **functional isolation model** suggests that context does not influence object identification, but only influences the later ‘post-perceptual’, semantic phase. According to this model, object identification results from a purely bottom-up visual analysis. When an object is embedded in a scene, this will activate the scene-schema and trigger an additional processing step aimed at integrating the object in the scene representation. The activated schema thereby acts as an interpretational template for the output of the object identification process: objects that fit the template merely need to be tagged as present, those that do not will require more extensive processing in order to be integrated (De Graef, 2005b).

One aspect that is often neglected is that not only global scene information (as described above), but also local object information induces context effects (De Graef, 2005b; De Graef, De Troy & d’Ydewalle, 1992). A scene typically contains several semantically related objects. When seeing one object, this leads to the activation of the corresponding object-cohort (i.e., a prototypical representation of objects that are episodically and/or semantically related) and also facilitates the processing of cohort-consistent objects. It is important to mention...
this because these object-to-object context effects are the result of intra-level or local coherence and require less top-down processing than inter-level, scene-to-object effects.

Furthermore, although post-perceptual processes can indeed produce context effects, there is convincing evidence that congruent contextual information does facilitate object recognition in TD individuals, thereby refuting the functional isolation model (Bar, 2004). Brain imaging studies have provided further insight in the underlying cortical mechanisms and have proven that a low-spatial frequency representation of the scene is projected early and rapidly from the visual cortex to the parahippocampal and the prefrontal cortex, resulting in a guess about the most likely context. This information is then fed back to the inferior temporal cortex, activating a set of object representations that are likely to occur in that particular scene. As such, the recognition of scene-consistent objects is facilitated by substantially limiting the number of object representations that need to be considered (Bar, 2004; Bar et al., 2006).

Based on this overview it is clear that different mechanisms can lead to context effects on visual object processing. In summary, the effects of context can arise due to global scene information and/or local object information. Furthermore, context effects can result from early top-down processes facilitating object identification and/or post-perceptual processes facilitating integration within the activated schema.

Regarding individuals with ASD, several studies have investigated contextual influences on visual object processing and found no behavioral differences when comparing a TD and an ASD group with normal intelligence. However, the problem is that all studies reporting equally strong context effects in both groups, failed to properly account for local inter-object effects and/or the influence of post-perceptual processes. It is therefore currently unclear whether ASD individuals (in some situations) display equally large global top-down effects of context, contradicting the WCC account, or whether they rely on a different, more local mechanism which produces the same result. For example, in a contextual priming study of Lopez and Leekam (2003), local inter-object priming instead of global scene-to-object priming might have induced the context effects in ASD individuals. Concerning the effect of post-perceptual influences, Henderson and Hollingworth (1999) provide a detailed evaluation for different paradigms. Based on this evaluation, it is clear that post-perceptual processes could explain the context effects found in the eye-movement study of Loth, Goméz and Happé (2011), as well as in the change blindness paradigm used by Fletcher-Watson, Leekam, Turner and Moxon (2006). In both studies, memory requirements play a role, which enlarge the effect of these post-perceptual processes (Henderson & Hollingworth, 1999). A possible explanation for the lack of group differences in both of these studies is that
individuals with ASD may use a different strategy and that context effects in this group occur at a later, post-identification stage, compared to a TD group. This suggestion was also formulated by Pijnacker et al. (2010), based on an ERP study of context sensitivity in ASD. They found that adults with so-called high-functioning autism had a normal response pattern at a behavioral level but a later ERP response. So, although the functional isolation model has been rejected as an explanation for context effects in TD individuals, this model might still apply for individuals with ASD.

In the present study we try to overcome the shortcomings of the previous studies by investigating global top-down effects of context in children with ASD compared to TD children. In line with suggestions based on the WCC theory, this was done by using an open-ended implicit task. More specifically, we used a scene perception task, based on a paradigm of De Graef et al. (1990). During this task, eye-movement patterns were recorded while participants explored line drawings of real-world scenes. All scenes contained one semantically consistent object, one semantically inconsistent object, and a variable number of non-objects (i.e., meaningless closed figures with an object-like appearance). The task was to count the number of non-objects in the scene. By showing stimuli with a variable number of non-objects and asking individuals to count them, they are motivated to scan the whole scene and saccade from one object-like form to the next. In doing so, they will also fixate existing objects in the scene at which point they will only have to determine whether the pictured object matches an object representation in long term memory or not (whether it is a non-object or not). Thus, maximum performance can be achieved without contextual processing. Additionally, participants are not explicitly asked to name the objects (as is often the case in contextual priming studies). Registration of eye-movement patterns in our study therefore provides an unobtrusive, on-line and implicit measure of context effects on visual object processing. Consistent with the literature on scene-context effects in the general population, we expect that inconsistent objects will be fixated longer, faster and more often than consistent objects.

To ensure that we measured global scene-to-object effects, the contribution of local inter-object effects were minimized. This was done by adding only two objects to the scene-background, which were semantically inconsistent with one another. So, seeing one object could not prime the representation of the other object. Additionally, we wanted to make sure that the observed context effects were (at least in part) due to early top-down influences on object recognition and could not be entirely contributed to post-perceptual processes. On the one hand, this was done by choosing a task without mnemonic requirements (unlike the change blindness task, for instance). On the other hand, this was done by investigating the interaction between context effects and the
ease of object recognition. One factor influencing object recognition speed is the viewpoint in which the object is perceived (Boutsen, Lamberts, & Verfaillie, 1998; Palmer, Rosch, & Chase, 1981). We therefore presented each object in its best and worst recognizable viewpoint and we expected to find longer fixation durations for objects in their worst recognizable viewpoint compared to the same objects in their best view. Concerning the interaction between context and viewpoint effects, two scenarios are possible. If the expected context effects merely reflect post-perceptual processes, then the ease of object recognition would not be influenced by the context and no context by viewpoint interaction should emerge. This prediction is in line with the functional isolation model. However, if the context facilitates object recognition, then the effect of context is expected to be larger when object recognition is harder (Bar, 2004), so when objects are presented in their worst view. This would be expected based on the perceptual schema and the priming model.

Concerning group differences in context effects, the WCC account predicts less top-down facilitation of object recognition in the ASD group. We therefore expect smaller context effects in the ASD group and a smaller interaction between context and viewpoint effects, compared to the TD group. This should result in a three-way group by context by viewpoint interaction. If the context by viewpoint interaction is absent for the ASD group, this means that the functional isolation model applies for these individuals. Since viewpoint effects rely more on bottom-up local processing, no group differences are expected for these effects.

2. Method

2.1 Participants

Twenty children with ASD (17 boys, 3 girls) and 20 TD children (17 boys, 3 girls) took part in this study. All children were between 8 and 14 years old and had verbal, performance and full-scale IQ scores above 80. Intelligence was assessed with a shortened, Dutch version of the Wechsler Intelligence Scale for Children, Third Edition (WISC-III<sub>NL</sub>; Kort et al., 2005), consisting of four subtests: Vocabulary, Similarities, Picture Completion and Block Design (Sattler, 2001). None of the participants took psychotropic medication. The participants were group-wise matched for gender, chronological age, verbal IQ (VIQ), performance IQ (PIQ), and full-scale IQ (FSIQ; see Table 1).

(Insert Table 1 about here)

Individuals with ASD were recruited through the Flemish Autism Association. They all received a formal diagnosis made by a multidisciplinary team according to DSM-IV-TR criteria (APA, 2000). All children were enrolled in special schools or in mainstream schools with additional guidance. The control group was
recruited through schools, personal contacts and advertisements. None of these TD children presented any neurological or psychiatric disorder or had a first degree family member with a developmental or neurological disorder.

2.2 Materials

During the scene perception task black-and-white line drawings of real-world scenes were presented. Each scene contained one semantically consistent object, one semantically inconsistent object and a variable number (0-3) of non-objects. In addition, each of these existing objects was shown in its best and worst recognizable viewpoint. To determine these views for each object, a pilot study was performed with 17 TD children (aged 8-14 years), who did not participate in the main experiment. During this study, 40 objects were shown in seven different depth-rotated views (a depth rotation of 0°, 15°, 30°, 45°, 60°, 75° and 90°; created by Germeys, 2004). The participants had to name the objects as quickly as possible. A voice key was used to record naming latencies, which provides a measure of goodness of view (shorter naming latencies reflect faster recognition and therefore better views). The results showed main effects of viewpoint and of object, and an interaction between the two. Based on these findings we selected 22 objects for the main experiment with large viewpoint effects and determined the best and worst viewpoint for each. In this pilot, we also presented 30 scenes (created by van Diepen & De Graef, 1994) to test which scenes could be recognized and named by the children. Twenty-six scenes were recognized by all children and 22 of these scenes were used for the main study. More information about this pilot is available on request.

In total, 44 stimuli were presented. The 22 objects shown were grouped in 11 pairs and each object pair was shown in two scenes: once in a scene where object A was semantically consistent and object B was semantically inconsistent and once in a scene where the object-scene consistency was reversed. Since all objects were shown in a consistent and in an inconsistent scene, the effect of object-scene consistency (or the context effect) could not be due to differences in lower-level visual saliency of the objects. In addition, each object pair was shown two times in both of the scenes: with both objects in their best or worst view. We also made sure that the location of the midpoint of each object within a particular scene in each of its views was the same and this without violation of two other object-context relations: support and interposition. Additionally, there was always a distance of five visual degrees between the consistent and the inconsistent object within a scene, so that fixations on either object are clearly distinguishable. Furthermore, the objects were never presented in the center of the screen. See Figure 1 for an example of one object pair in each of its presentations. The construction of the
stimuli was completed by the insertion of varying numbers of non-objects (selected from the set created by van Diepen & De Graef, 1994), in such a way that the percentage of stimuli with zero, one, two or three non-objects was equal (in each case 25%). In total, 90 non-objects were presented.

To record eye-movement patterns, the stimuli were presented on a 17-inch monitor of the Tobii T120 eye tracker (with a spatial resolution of 1240 x 768). Eye-movement data were collected at a frame rate of 120Hz. This online record was broken down into separate fixations by means of the built-in Tobii fixation filter with a 35 pixel radius (see Tobii Studio 2.0 user manual, 2010). The coordinates of the fixation were then compared to the coordinates of the existing objects in the displays, by determining whether fixations fell inside or within 0.5° of the smallest rectangle enveloping the whole object. Afterwards, the following fixation parameters for these objects were determined: first fixation duration, total fixation duration (i.e., the summated duration of all fixations on the object), time to first fixation (i.e., the latency between the beginning of a trial and the first object fixation), and fixation count (i.e., the total number of refixations on the object).

2.3 Procedure

All participants were tested individually in a quiet room at the University Hospital ‘Gasthuisberg’ in Leuven. They were told that during this task images would be presented containing a number of non-objects and that they had to count the number of non-objects in the image. Non-objects were described as strange figures that could not be named and were probably never seen before. In order to illustrate the concept and to make sure that participants understood it, they were given a page containing existing and non (existing) objects and for each of these objects they had to indicate whether it was a non-object or not. Afterwards, participants were seated in front of the monitor at a distance of 60cm, with their heads fixated on a head-and-chin rest to minimize head movements and to keep viewing distance constant. Subsequently, the lights were turned off and the task was initiated. First, a nine-point calibration of the Tobii was performed. Next, participants completed six practice trials before the 44 experimental trials were presented. During the trials, a fixation cross was first presented in the center of the screen for 500ms and participants were instructed to fixate it. Then, a stimulus was shown for 10s. A pilot study had indicated that 10s was sufficient for a complete exploration of the scene for children between 8 and 14 years old. A fixed exposure duration was used to create similar viewing conditions for all scenes and to prevent participants from terminating a scene presentation before they had completely examined it. Following this scene presentation, individuals had to say how many non-objects were present in the scene. During the practice trials, the same scene was then shown again and individuals had to indicate the non-objects.
allowing us to correct mistakes. The experimental trials were grouped in four blocks of 11 trials per block, so that the same existing objects did not appear twice in a block and that the number of objects shown in a particular view (best vs. worst) was equated within a block. All stimuli within a block were presented in random order. Between each block, the light was switched on and a short break was provided. Before the start of the next block, the nine-point calibration was repeated.

Informed consent was obtained from the participants’ parents, according to the study protocol, approved by the Ethical Committee of the KU Leuven. After testing, participants could choose a reward for their participation (e.g., a film ticket, a comic book, marbles, etc.).

2.4 Data analyses

A t-test was used to analyze group differences in the reported number of non-objects. To analyze the eye-movement data, the fixation parameters described earlier were taken as dependent variables and the effect of the following independent variables was explored: the between-subjects factor ‘group’ (ASD vs. TD) and the within-subject factors ‘context’ (consistent vs. inconsistent) and ‘viewpoint’ (best vs. worst). Since there were several missing values, a multilevel analyses was performed (using the MIXED procedure in SAS 9.3). For each dependent measure the same multilevel model was tested including the main and interaction effects of all independent variables. A significance level of $p < 0.05$ (two-sided) was adopted and a Tukey-Kramer correction was used for pairwise contrasts.

Before applying the model, all dependent variables were log-transformed to obtain more normally distributed data. Afterwards, data within each category (group: ASD vs. TD; context: consistent vs. inconsistent and viewpoint: best vs. worst) were inspected for outliers. An outlier was defined as a value with a Z-score higher than 3 or lower than -3. All analyses were performed including and excluding outliers. Both ways of analyzing yielded the same results. Therefore, only the analyses on the full participant sample will be reported. The data shown in the graphs and in the tables result from the non-transformed variables.

3. Results

Although children with ASD counted on average more non-objects ($M = 102$, $SD = 10.66$) than TD children ($M = 98$, $SD = 5.29$), this difference was not significant ($t(38) = 1.65; p = 0.11$).
Regarding the eye-movement data on the existing objects, we found a significant main effect of context for all fixation parameters. Objects shown in an inconsistent context, compared to the same objects in a consistent context, had a longer first fixation duration \((F(1,38) = 20.97, p < 0.001)\), a longer total fixation duration \((F(1,38) = 50.28, p < 0.001)\), a shorter time to first fixation \((F(1,38) = 10.65, p < 0.01)\) and a higher fixation count \((F(1,38) = 20.38, p < 0.001)\).

Additionally, we tested whether our context manipulation was successful for all objects in the TD group. This was done by calculating the difference score between the consistent and the inconsistent context for all objects and all parameters separately. Four objects did not show the expected context effects on any of the parameters. We therefore excluded these objects from subsequent analyses. Doing this enlarged the context effects \((p < 0.001\) for all parameters). See Table 2 for the corresponding mean values and standard errors.

(Insert Table 2 here)

There was also a significant main effect of viewpoint for all fixation parameters, except for the variable ‘fixation count’ \((F(1,38) = 1.14, p = 0.29)\). Objects in their worst recognizable viewpoint had a longer first fixation duration \((F(1,38) = 15.52, p < 0.001)\), total fixation duration \((F(1,38) = 12.15, p < 0.01)\) and time to first fixation \((F(1,38) = 11.48, p < 0.01)\) compared to the same objects in their best view (see Table 2).

Additionally, an interaction between context and viewpoint effects was found for the variables first fixation duration \((F(1,38) = 13.79, p < 0.001)\) and time to first fixation \((F(1,38) = 4.60, p < 0.04)\) (see Figure 2 and 3, respectively), with a larger context effect for objects presented in their worst view, as expected. Pairwise contrasts revealed that for the time to first fixation, the context effects were significant for both views (both \(p < 0.001\)). However, for the first fixation duration, we only found a significant context effect for objects in their worst view \((p < 0.001)\), and not for objects in their best view \((p = 0.99)\). Furthermore, for this parameter, the viewpoint effect was only significant in an inconsistent context \((p < 0.001)\), not in a consistent context \((p = 0.99)\). No interaction between context and viewpoint effects was observed for the parameters total fixation duration \((F(1,38) = 0.95, p = 0.34)\) and fixation count \((F(1,38) = 0.80, p = 0.38)\).

(Insert Figure 2 and Figure 3 about here)

Concerning the main effect of group, there was no significant group difference for any of the fixation parameters. So overall, children with ASD did not differ from TD children regarding first fixation duration \((F(1,38) = 0.01, p = 0.91)\), total fixation duration \((F(1,38) = 1.10, p = 0.30)\), time to first fixation \((F(1,38) = 3.22, p = 0.08)\) and fixation count \((F(1,38) = 0.04, p = 0.85)\) (Table 2). Additionally, there were no significant interactions between group and viewpoint (for the first fixation duration \(F(1,38) = 0.18, p = 0.68\), see Figure 2;
for the total fixation duration $F(1,38) = 2.56, p = 0.12$; for the time to first fixation $F(1,38) = 0.13, p = 0.72$, see Figure 3; and for the fixation count $F(1,38) = 3.13, p = 0.08$, nor between group and context (for the first fixation duration $F(1,38) = 0.01, p = 0.92$, see Figure 2; for the total fixation duration $F(1,38) = 0.08, p = 0.78$; for the time to first fixation $F(1,38) = 1.11, p = 0.30$, see Figure 3; and for the fixation count $F(1,38) = 0.55, p = 0.46$), indicating that both viewpoint and context effects were similar in the ASD compared to the TD group.

Despite the absence of a significant group by context interaction, additional analyses were performed to investigate whether this interaction effect might be dependent on the viewpoint, by testing the three-way interaction between group, context and viewpoint for all fixation parameters. None of these interactions appeared to be significant (for the first fixation duration $F(1,38) = 0.16, p = 0.69$, see Figure 2; for the total fixation duration $F(1,38) = 0.48, p = 0.49$; for the time to first fixation $F(1,38) = 0.27, p = 0.61$, see Figure 3; and for the fixation count $F(1,38) = 0.58, p = 0.45$). This indicates that group differences in context effects were absent for both views, but also that the interaction effect of context and viewpoint was similar for both groups.

4. Discussion

In the present study, children with ASD and typically developing controls were explicitly asked to count the number of non-objects present in a scene. Both groups performed equally well on this task. At the same time, we measured their eye-movements to investigate implicitly induced global top-down effects of context (consistent vs. inconsistent) and more bottom-up effects of viewpoint (best vs. worst recognizable view) on visual object processing. Overall, we found the expected context and viewpoint effects, which demonstrates that both manipulations were successful. More specifically, regarding the context effects, objects shown in a consistent context were fixated shorter, later and less often than the same objects shown in an inconsistent context. These findings indicate that objects that are semantically consistent with the scene are easier to recognize and/or easier to integrate within the scene (Henderson & Hollingworth, 1999). According to De Graef (2005b), the later fixation on scene consistent objects is due to the enhancement of extra-foveal processing of these objects, reducing the urgency to actually fixate them. The higher fixation count for inconsistent objects probably reflects additional post-perceptual processes trying to integrate the object within the activated scene-schema. When integration fails, viewers will re-fixate that object more often to verify its identity and repeat the object-scene integration. Concerning the viewpoint effects, we found that an object presented in its best view was fixated shorter and faster than that same object in its worst view. These findings confirm that objects in their best view are easier to recognize. A possible explanation for the faster fixation is that the figure-ground segregation occurs
faster when objects are presented in their best view, making them a more salient fixation target if you are on the look-out for objects. No viewpoint effect was found for the number of fixations. This is because this parameter is mainly determined by post-perceptual processes (object-scene integration), which are not influenced by the object viewpoint.

Another finding is that the size of the context effects depends on the viewpoint of the objects, with larger context effects for objects presented in their worst view. Interestingly, for the first fixation duration, the context effect was only significant for objects in their worst view. This is in line with Bar (2004), who states that context effects are particularly evident when context is needed to disambiguate meaning. When objects are perceived in ideal conditions, recognition can occur within less than 150ms (Kirchner, & Thorpe, 2006; Thorpe, Fize, & Marlot, 1996). This process is so rapid and efficient that it can hardly benefit from early top-down facilitation. However, when object recognition becomes harder (e.g. for non-canonical views), congruent contextual information can facilitate object recognition (Bar, 2004). Furthermore, objects presented in their worst view had a longer first fixation duration than objects in their best view, but only when the objects were semantically inconsistent with the scene. These findings clearly demonstrate that a congruent context facilitates object recognition and thus provide evidence against the functional isolation model.

The above mentioned findings indicate that our paradigm worked, but the main aim of this study was to investigate whether ASD and TD children process visual object information differently and particularly whether contextual facilitation of object recognition occurs to the same extent in both groups. Overall, no group differences were found for any of the fixation parameters. Additionally, both groups displayed equally large viewpoint effects. This was expected based on the WCC account, since viewpoint effects rely more on bottom-up local processing. However, opposite to predictions based on this theory, both groups also showed equally large context effects on visual object processing. In this study, we can be confident that these equally large context effects in the ASD group are not due to the operation of a more local, less top-down contextual mechanism. On the one hand, the effects of local inter-object priming are minimized in our task, so, if children with ASD would only rely on local context information, then smaller context effects would be observed compared to TD children. This more local mechanism also implies that at least one object needs to be identified before influencing the processing of subsequent objects. However, since we already found context effects on the first fixation duration of the objects, which were equally pronounced in both groups, it is not very likely that the context effects in the ASD group are more due to local inter-object priming. On the other hand, we can exclude
the possibility that the context effects in the ASD group are merely the result of post-perceptual processes, because the interaction between context and viewpoint effects is equally large for both groups. This means that in this task a congruent context facilitates object recognition to the same extent in children with ASD compared to TD individuals, refuting the functional isolation model for both groups. In sum, these findings provide clear evidence against the WCC account, which predicted smaller global, top-down context effects in ASD measured with an open-ended implicit task.

Another theory trying to account for differences in visual processing in ASD is the Enhanced Perceptual Functioning (EPF) theory (Mottron, Dawson, Soulieres, Hubert, & Burack, 2006). Although there is a great degree of overlap between the EPF and the WCC account, they have a different emphasis and in some cases lead to different hypotheses. Central in both models is the superior local processing in ASD. However, reduced global processing is conceptualized differently. According to the EPF theory, global/contextual processing is mandatory for typically developing individuals, even when it is detrimental to performance, whereas ‘higher-order control over perception is not mandatory in autism when it interferes with performance of tasks that can be more economically processed locally or using a low-level processing mode’ (Mottron et al., p.39). This suggests that individuals with ASD will only show reduced global processing when using global information is detrimental or irrelevant for task performance. In our paradigm, using contextual information helps to identify an object-like figure as an existing object or a non-object and thus helps to perform the task. Based on the EPF theory, we therefore expected equally large context effects in the ASD and TD group. Our findings are in line with this prediction.

However, several studies have provided evidence against this theory by showing equally large context effects in ASD when this leads to detrimental performance on the task (Maras & Bowler, 2011) and reduced contextual processing in tasks where contextual information improved performance (Kamio, Robins, Kelley, Swainson & Fein, 2007). Furthermore, a problem with this theory is that it does not provide a clear description of the circumstances leading to reduced global processing in ASD. So overall, the EPF theory does not provide a better alternative for the WCC account to explain the inconsistent findings regarding contextual processing in ASD.

There is growing evidence that individuals with ASD have no general deficit in contextual processing. Individuals with ASD are often impaired in seeing the bigger picture in daily life, and several studies report
reduced contextual influences in this population (Benson, Piper & Fletcher-Watson, 2009; Loth, Gómez & Happé, 2008), whereas other studies demonstrate intact contextual processing in ASD under different circumstances (Fletcher-Watson et al., 2006; López & Leekam, 2003; Loth, Gómez & Happé, 2011). However, the specific conditions under which contextual processing is reduced remain to be specified. Klin and Jones (2006) suggest that ASD individuals particularly show deficits when using social stimuli. Recently, this claim is supported by the finding that adults with ASD show diminished top-down processing for social (faces) but not for non-social stimuli (objects or houses; Loth, Gómez & Happé, 2010). Benson, Piper and Fletcher-Watson (2009) also reported atypical saccadic scanning of a scene in ASD by using a social instruction. Another suggestion is made by López and Leekam (2003). They propose that individuals with ASD are not generally impaired to use context, but only have specific difficulty with complex verbal stimuli and in particular when using sentence context is needed to disambiguate meaning. Furthermore, it is not unlikely that the extraction of meaning occurs through a different interplay of bottom-up and top-down processing in simplified, schematic line drawings (used in our study) compared to richer, more naturalistic stimuli (e.g. photos or videos). For example, Loth, Gómez and Happe (2008) found reduced contextual influences on detecting object changes in people with ASD when using color photographs. Yet another possibility is that memory recollection might induce group differences. In line with this, studies have demonstrated that recalling information from memory is less influenced by context in ASD (Bowler, Gaigg & Gardiner, 2008; Loth, Gómez & Happé, 2011; but see Maras & Bowler, 2011 for negative findings). Furthermore, it is possible that differences between TD and HFA individuals are very subtle and often missed when only looking at the behavioral outcomes of a certain process and not at the underlying neurological mechanisms. For instance, in an ERP study on context sensitivity in ASD, a normal response pattern was found at the behavioral level, but group differences were found in neural processing (Pijnacker et al., 2010).

Besides the above mentioned differences in paradigms, findings might also depend on the specific ASD population. Since our sample was restricted to individuals with ASD and a normal intelligence, our findings are probably not universal for the entire ASD population. Several population characteristics that might be important are age, IQ, and severity of ASD symptoms. For example, Pijnacker et al. (2010) found deviating ERP responses on a context sensitivity task in a high-functioning autism group, but not in participants with Asperger syndrome. A limitation of our study is that we miss a quantitative measure of autism severity. All ASD participants had a formal ASD diagnosis made by a multidisciplinary team and received special guidance at school. Additionally, they performed worse on a more controlled version of the Wisconsin Card Soring Task, measuring cognitive
flexibility. This task was administered to the same sample, before or after the scene perception task. The ASD group made on average more perseveration errors \( (p = 0.04) \) compared to the TD group and had a higher switch cost in reaction time \( (p = 0.04) \), with only significant group differences on switch trials \( (p = 0.04) \), not on maintain trials \( (p = 0.33) \) (for more information about the task and results of a more extended sample see Van Eylen et al., 2011). These findings demonstrate that the same ASD group does differ from the TD group on another task, although they performed equally well on the scene perception task. However we do not know how severe their ASD symptoms are. It could be that group differences in our scene perception task would emerge when studying younger, lower functioning individuals with more severe ASD traits.

Based on this overview, it is clear that further research is needed to gain more insight in the specific factors leading to reduced contextual processing in ASD individuals.

To conclude, our study demonstrates that individuals with ASD do not show reduced top-down processing on an open-ended implicit task requiring global coherence. These findings provide evidence against the WCC account. Our results are more in line with predictions derived from the EPF account, but this framework is also not without its problems and is also not capable of explaining all current inconsistencies. Although there is growing evidence that individuals with ASD have no general global processing deficit, further research is needed to elucidate the determining factors.

5. Acknowledgements

This study was funded by a doctoral fellowship from the Fund for Scientific Research (FWO Flanders) to Lien Van Eylen, a grant from the Research Council of the KU Leuven (IDO/08/013) to Jean Steyaert, Johan Wagemans, and Ilse Noens, and from the Methusalem program (METH/08/02) to Johan Wagemans. We are especially grateful to Peter Catteeuw from the Movement Control and Neuroplasticity Research Group of the KU Leuven for lending us the Tobii T120. Furthermore, we thank all children who participated and the following master students from the Faculty of Psychology and Educational Sciences for their assistance in data collection: Ine Thys, Sofie Cromheeke and Goedele Degraeuwe.

Conflict of interest

The authors declare that they have no conflict of interest.
6. References


Figure Captions

Figure 1. An illustration of the four scenes containing the object pair ‘goat – suitcase’. In the two upper displays the object pair is shown in the farm were the goat is the consistent object and the suitcase is inconsistent and in the two lower displays the objects are placed in the office scene where the object-scene consistency is reversed. The left and right displays contain the object pair in its best and worst viewpoint, respectively. Each scene also contains 0-3 non-objects.

Figure 2. Mean first fixation duration (in ms) as a function of context (consistent versus inconsistent), viewpoint (best versus worst) and group (ASD versus TD).

Figure 3. Mean time to first fixation (in ms) as a function of context (consistent versus inconsistent), viewpoint (best versus worst) and group (ASD versus TD).
Figure 2.

Mean first fixation duration (in ms)

Viewpoint

- consistent_ASD
- inconsistent_ASD
- consistent_TD
- inconsistent_TD
Table 1

**Age and IQ scores (mean and standard deviation) for the ASD and the TD group, as well as t-scores and p-values demonstrating that both groups are matched for age, VIQ, PIQ and FSIQ**

<table>
<thead>
<tr>
<th></th>
<th>ASD group</th>
<th></th>
<th>TD group</th>
<th></th>
<th>t_{(df = 38)}</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td>10.2</td>
<td>1.82</td>
<td>10.6</td>
<td>2.11</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>VIQ</td>
<td>105.5</td>
<td>13.14</td>
<td>112.3</td>
<td>12.24</td>
<td>1.68</td>
<td>0.10</td>
</tr>
<tr>
<td>PIQ</td>
<td>105.6</td>
<td>14.53</td>
<td>107.0</td>
<td>14.44</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td>FSIQ</td>
<td>105.5</td>
<td>12.27</td>
<td>109.6</td>
<td>10.30</td>
<td>1.14</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th></th>
<th>First fixation duration (in ms)</th>
<th>Total fixation duration (in ms)</th>
<th>Time to first fixation (in ms)</th>
<th>Fixation count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>consistent</td>
<td>315.2 (7.2)</td>
<td>1244.6 (31.8)</td>
<td>2347.3 (57.0)</td>
<td>3.6 (0.07)</td>
</tr>
<tr>
<td>inconsistent</td>
<td>362.8 (10.8)</td>
<td>1664.9 (37.5)</td>
<td>1642.9 (44.4)</td>
<td>4.6 (0.09)</td>
</tr>
<tr>
<td><strong>Viewpoint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>best</td>
<td>316.8 (7.15)</td>
<td>1306.5 (28.9)</td>
<td>1897.8 (51.1)</td>
<td>4.1 (0.08)</td>
</tr>
<tr>
<td>worst</td>
<td>362.5 (11.1)</td>
<td>1614.2 (40.6)</td>
<td>2078.1 (52.6)</td>
<td>4.2 (0.09)</td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>333.8 (7.1)</td>
<td>1454.0 (35.75)</td>
<td>2043.8 (52.9)</td>
<td>4.2 (0.09)</td>
</tr>
<tr>
<td>TD</td>
<td>345.3 (11.1)</td>
<td>1464.7 (35.5)</td>
<td>1930.6 (50.8)</td>
<td>4.0 (0.08)</td>
</tr>
</tbody>
</table>