Psychological Research
https://doi.org/10.1007/s00426-019-01223-0

ORIGINAL ARTICLE

# Do ostensive cues affect object processing in children with and without autism? A test of natural pedagogy theory

**Tobias Schuwerk**[1] · **Johannes Bätz**[2] · **Birgit Träuble**[2] · **Beate Sodian**[1] · **Markus Paulus**[1]

Received: 17 October 2018 / Accepted: 2 July 2019

## Abstract

Theories suggest that the perception of others' actions and social cues leads to selective processing of object features. Most recently, natural pedagogy theory postulated that ostensive cues lead to a selective processing of an object's features at the expense of processing of its location. This study examined this hypothesis in 10-year-old children with and without autism spectrum condition (ASC) to better understand social information processing in ASC and the relevance of observing others in human object processing in general. Participants saw an agent either ostensively pointing to an object or non-ostensively grasping an object. Thereafter, the cued or uncued object changed either its location or identity. We assessed not only behavioral responses, but also participants' gaze behavior by means of eye tracking. In contrast to natural pedagogy theory, we found that in the non-ostensive grasping context, participants rather noticed an identity change than a location change. Moreover, location changes were more readily identified in the ostensive pointing context. Importantly, there was no difference between children with and without ASC. Our study shows that the perception of ostensively vs. non-ostensively framed actions leads to different processing of object features, indicating a close link between action perception, object processing, and social cues. Moreover, the lacking group difference in our study suggests that these basic perception–action processes are not impaired in autism.

## Action perception and object processing

Humans are a remarkable social species. It has been argued that the evolutionary success of our species depends crucially on the social transmission of knowledge and our ability to interact and cooperate with each other (Boyd & Richerson, 1988). Interestingly, the impact of others on our behavior is not restricted to verbal exchange and discursive practices, but the mere observation of others' behavior deeply affects our own actions and perception of the physical world (for reviews, see Becchio, Bertone, & Castiello, 2008; Kunde, Weller, & Pfister, 2018; Thill, Caligiore, Borghi, Ziemke, & Baldassarre, 2013). Notably, perceiving others' object-directed behavior affects object processing (Böckler, van der Wel, & Welsh, 2014, 2015). For example, Fagioli, Ferlazzo, and Hommel (2007) showed that watching a grasping action leads to a faster detection of object size changes than object location changes. Thus, it has been shown that others' actions entrain our behavior and perception of the world (Knoblich & Sebanz, 2008); and it has been proposed that this is possible, because executed and perceived action rely on the same event codes (cf., Dolk, Hommel, Prinz, & Liepelt, 2013; Paulus, 2012).

Psychology has experienced an increased interest in how precisely the processing of and learning from observed actions is affected by its concomitant social cues. Social cues can be defined as communicative signals to another person and encompass conventionalized and deliberately chosen behaviors (e.g., pointing), in some definitions also more implicit cues that are often not intentionally displayed (e.g., body orientation), but nevertheless informative for the observer (c.f.; Frith & Frith, 2008; Paulus, Murillo, & Sodian, 2016). These social signals are powerful tools to direct someone's attention and, therefore, to modulate cognitive processes. For example, abundant studies with children and adults have shown that observing others' eye gaze facilitates the processing of objects (Becchio et al., 2008), resolves ambiguities in conversation (Hanna & Brennan, 2007), and enables social learning, that is, the acquisition

---

Tobias Schuwerk
tobias.schuwerk@psy.lmu.de

[1] Department of Psychology, Ludwig-Maximilians-University, Leopoldstr. 13, 80802 Munich, Germany

[2] Department of Psychology, University of Cologne, Cologne, Germany

of novel action knowledge (Bandura, 1977). Thus, the processing of others' gaze and pointing cues is a psychological core mechanism of object processing and social learning (Moore, 2013).

## Natural pedagogy theory

One recent theoretical account that has received wide interest, natural pedagogy theory, suggests that ostensive cues (that is, social cues that signal another person that information is intentionally communicated) lead to a modulation of the processing of object features (Csibra, 2010; Csibra & Gergely, 2009). More concretely, ostensive signals such as establishing eye-contact or using addressee-directed speech are supposed to transmit the intention to communicate and tell the learner that everything that follows next is directed to her (Sperber & Wilson, 1986). Thus, ostensive signals initiate a learning context and bias the learner's information processing during the learning phase. For example, it has been proposed that, in an ostensive context, a learner expects to be taught something generalizable and thus focusses her attention to intrinsic, durable, and thus identity-relevant features of an object (Csibra & Gergely, 2009). Moreover, it has been suggested that this type of natural pedagogy is an evolutionary selected adaptive cognitive system that is already in place in young infants (Csibra & Gergely, 2011).

A study with neurotypical adults by Marno, Davelaar, & Csibra (2014) provided empirical support for this notion. In a multiple-trial change detection paradigm, an agent cued an object in two different ways. In the so-called ostensive context trials, the agent directly gazed at the participant, waved, and then pointed at one out of five different objects. In the non-ostensive context trials, the agent rubbed her chin and tried to grasp one out of these objects. In the end of each trial, a static image of the scene was presented in which one out of the five objects (sometimes the cued object, sometimes an uncued object) had changed in either its identity or location. The participants had to state which of the objects had changed. The change detection rates supported natural pedagogy's claim that in an ostensive context, people preferentially encoded the object's identity at the expense of its location (for similar findings with infants see Yoon, Johnson, & Csibra, 2008)

However, the conclusiveness of the empirical evidence for natural pedagogy theory is hotly debated. On one hand, further evidence has been provided by other studies with infants (e.g., Träuble & Bätz, 2014; Yoon et al., 2008) and adults (e.g., Marno et al., 2014). For example, Senju and Csibra (2008) reported that infants only follow another person's gaze when they were communicatively addressed by ostensive signals. Yet, on the other hand, some scholars have suggested that some of these findings could be more easily explained by domain-general attentional processes (e.g., de Bordes, Cox, Hasselman, & Cillessen, 2013; Gredebäck, Astor, & Fawcett, 2018; Heyes, 2016; Szufnarowska, Rohlfing, Fawcett, & Gredebäck, 2014). More precisely, it has been suggested that the findings of enhanced gaze following after ostensive cues can be explained by perceptual saliency (i.e., ostensive cues being more salient and, therefore, more effective) rather than a dedicated and evolutionarily specialized learning mechanism. Moreover, recent work has failed to replicate some of the evidence in favor of natural pedagogy theory (Silverstein, Gliga, Westermann, & Parise, 2019). Taken together, although strong theoretical claims have been put forward (e.g., Csibra & Gergely, 2011; Csibra & Shamsudheen, 2015), the empirical basis is disputed and inconclusive. Further empirical work is, therefore, required to probe the predictions made by natural pedagogy theory.

It was one aim of the current study to contribute to this debate by clarifying the empirical basis of the claim that ostensive cues affect object processing. Our experiment based on the paradigm of Marno et al., (2014) as well as Yoon et al. (2008) contributes to the field in two main aspects. First, a replication attempt of the infant paradigm (Yoon et al., 2008) was not able to find the same pattern (Silverstein et al., 2019). It would, therefore, be important to clarify the extent to which the finding by Marno et al. (2014) could be reliably found in a developing population. Given natural pedagogy's strong claim that the effect of ostensive cues is particularly important in development, we decided to focus our study on children. Given that the paradigm by Marno et al. relies on explicit verbal responses in a complex task setting, we decided to test children in middle childhood. Second, Marno and colleagues reported that, in an ostensive pointing context (but not in a non-ostensive grasping context), participants preferentially encoded the object's identity at the expense of its location. This seems to be in contrast to other studies that, based on the ideomotor theory, reported that either planning or observing a grasping action enhanced detection of changes in object features (Fagioli, Ferlazzo, et al. 2007; Fagioli, Hommel, et al. 2007; Wykowska, Schubö, & Hommel, 2009).

## Action perception in autism

A second question relates to the impact of social cues for people with autism spectrum condition (ASC). This condition is characterized by problems with communication and social interaction. To explain these two cardinal symptoms of ASC, researchers targeted characteristics of basic social cognitive processing in ASC (e.g., Frith, 2012; Hamilton, 2015). It is hypothesized that, compared to people without ASC, individuals with ASC process others' actions and related social stimuli, i.e., any informational input that stems from other human beings, differently. This difference in basic cognitive processing then causes the problems many autistic persons are facing in social situations, such as difficulties to initiate and maintain a reciprocal conversation and extracting what is meant to be communicated. Thus, these impaired processes affect social learning about actions and objects in ASC.

Consequently, research started to explore in greater detail whether and how autistic persons—relative to non-autistic persons—process social signals differently. Whereas some proposed a general deficit in the processing of social signals (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012; Shultz, Jones, & Klin, 2015), others challenged this claim (Cusack, Williams, & Neri, 2015; Jaswal & Akhtar, 2018). Recent studies suggest that the processing of social information might rather be more effortful and fragile (e.g., Aldaqre et al., 2016; Aldaqre, Paulus, & Sodian, 2015; Schuwerk, Sodian, & Paulus, 2016; for a review, see Guillon, Hajikhani, Baduel, & Rogé, 2014).

Given that autistic people have difficulties in social learning (Williams, Whiten, & Singh, 2004) and given the claim that social learning is based on enhanced processing of intrinsic object features in ostensive contexts (Csibra & Gergely, 2009), the question arises whether they would be less susceptible for the effects of ostensive cues. Taken together, studying the influence of ostensive context on cognitive processing in ASC is critical to better understand social information processing in ASC and the relevance of observing others in human object processing in general. Thus, a second aim of the study was to examine differences in the impact of ostensive cues on object processing between children with autism and without autism.

## The current study

The current study is an adaption of the multiple-trial change detection task employed by Marno et al. (2014) and Yoon et al. (2008). To make the task suitable for our age group, we used stimuli that allowed for adjusting the task difficulty to a medium level between Marno et al.'s paradigm for adults and Yoon et al.'s paradigm for infants. In this task, participants watched videos of a scene with four objects. An actress highlighted one of these objects either using ostensive or non-ostensive signals and actions. Note that our operationalization of ostensive and non-ostensive signals and accompanying actions was close to Marno et al. (2014). The actress either greeted the viewer by looking directly into the camera, smiling and waving (ostensive condition), or she kept her face low, looked down at the table and used her hand to go through her hair (non-ostensive condition). After a short distractor task, participants had to indicate which one out of these four objects had changed. In each trial, always one object had changed either its location (a short- or long-distance forwards or backwards) or identity (changed its color or shape). According to natural pedagogy theory (Marno et al., 2014), the ostensive context should facilitate identity change detection at the expense of location change detection. Based on Yoon et al., (2008), we further predicted a facilitated location change detection performance at the expense of identity change detection in the non-ostensive context.

In addition to the children's explicit responses, we assessed their gaze patterns during the critical test phase, in which one object had changed. We employed eye tracking to tap preferential processing of the object that had changed, indicated by relative fixation duration (cf., Corkum & Moore, 1998; Ferguson and Breheny, 2012). This would allow us to not only assess behavioral patterns, but also underlying processing strategies (cf., Karatekin, 2007). This is particularly relevant for the comparison between children with and without ASC, as it has been argued that similar behavioral patterns could be subserved by different processing strategies (e.g., Aldaqre et al., 2016). Following natural pedagogy theory, one would predict a higher preference for visual processing of objects that changed in identity as compared to objects that changed in location. In the non-ostensive context, we expected to find a higher preference for the object that changed in location as compared of the object that changed identity.

We were especially interested in assessing group differences between children with and without ASC given the recent debate on the nature of action processing in people with ASC (e.g., Cusack et al., 2015; Palmer, Paton, Kirkovski, Enticott, & Hohwy, 2015; Vivanti et al., 2011). If ostensive contexts do not (or only weakly) affect object processing in ASC, we would expect no (or a reduced) effect of this condition in the ASC group. Yet, if ostensive contexts do not differentially affect information processing in ASC, we would expect no difference between groups.

## Methods

The data of this study are available at https://osf.io/dwzn3/. To protect data privacy, we only provide the coded explicit responses and preprocessed gaze data. The demographic information is not shared as it cannot be guaranteed that it is impossible to identify individual data sets. Following best practice recommendations, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons, Nelson & Simonsohn, 2012).

## Participants

A total of 46 participants took part in this study: 23 children with ASC (ASC group; $M_{\text{age}} = 9.9$ years, $\text{SD} = 2.6$ years, 21 male) and 23 neurotypical comparison participants (NT group, $M_{\text{age}} = 9.4$ years, $\text{SD} = 1.7$ years, 19 male). Three additional participants had to be excluded, because they terminated the task prematurely (1 ASC, 1 NT), or because not enough gaze data could be sampled (1 NT). Children's caregivers gave informed written consent and received monetary compensation for travel expenses. Children received individualized gifts for their participation. The study was conducted in laboratories at LMU Munich, Germany. The local ethics committee approved the study. The sample size was determined based on a combination of the result of an a priori power analysis using G*Power [$N = 16$, for the predicted interaction effect in a repeated-measures analysis of variance (ANOVA), $f = 0.25$, power $(1 - \beta) = 0.80$, $\alpha = 0.05$, Faul et al. (2007)] and the sample size of the previous study ($N = 24$) by Marno et al. (2014).

Participants with ASC were recruited via local ASC associations, clinics, and practice-based physicians. The sample comprised 15 children with Asperger syndrome and eight children with childhood autism. To be included, they had to be diagnosed by qualified clinical psychologists or psychiatrists, meeting International Classification of Diseases-10th Revision criteria (World Health Organization, 1993). Sources of this information were individual medical reports. Of these, 16 reported a diagnosis based on expert evaluations supported by evidence-based assessment of ASC, including the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000). The other seven medical reports notified clinical diagnoses based on expert evaluations without detailing employed diagnostic tools.

Children from the NT group were recruited via birth records. They were matched by chronological age as well as verbal and non-verbal IQ [Wechsler-Intelligence Scale for Children (WISC-IV), subtests: "vocabulary", "similarities", "matrix reasoning", "picture completion" (Wechsler, 2003; German version by Petermann & Petermann, 2007), Table 1]. Children from the comparison group had no history of developmental, psychiatric, or neurological conditions. Note that one child from this sample was diagnosed with attention-deficit hyperactivity disorder and two with anxiety disorders. Due to the comorbidity of these disorders in ASC (in the current sample $n = 6$ with ADHD and $n = 1$ with an anxiety disorder), we refrained from excluding them but rather considered these individuals as closely matched (Schwartz & Susser, 2011).

To characterize levels of autistic traits in our samples, caregivers of all participants completed the Social Responsiveness Scale (SRS), a quantitative measure of autistic traits (Constantino & Gruber, 2005; German version by Bölte & Poustka, 2008), and the Social Communication Questionnaire (SQR; Rutter, Bailey, & Lord, 2003; German version by Bölte & Poustka, 2006). The SCQ assesses communication skills and social functioning, and consists of two forms, a *lifetime form*, referring to the entire development of the child, and a *current form*, focusing on the most recent 3 months. All participants from the ASC group scored above threshold on the SRS (T-score of $\geq 60$), and most of them scored above the discriminative cutoff in the SQR lifetime form (sum score of $\geq 15$ for comparisons between ASC and NT groups). Four individuals from the ASC group scored below that threshold. All participants from the NT group scored below the cut-off of the SQR lifetime form. Four participants from the NT group scored above the threshold of the SRS. All of these children remained included in their

**Table 1** Mean scores (standard deviation in brackets) of demographics and control measures, listed for each group

| | ASC ($n = 23$) | NT ($n = 23$) | Group comparison | | |
|---|---|---|---|---|---|
| | | | | *P* value | Effect size |
| Chronological age in years | 9.9 (2.6) | 9.4 (1.7) | $t(44) = 0.83$ | $p = 0.409$ | Cohen's $d = 0.25$ |
| Verbal IQ[a] | 110.9 (16.5) | 110.6 (13.9) | $t(44) = 0.05$ | $p = 0.962$ | Cohen's $d = 0.01$ |
| Non-verbal IQ[a] | 108.5 (14.8) | 108.6 (15.7) | $t(44) = -0.03$ | $p = 0.977$ | Cohen's $d = -0.01$ |
| SRS[b,*] T-score | 82.8 (10.4) | 49.1 (9.1) | $t(42) = 11.46$ | $p < 0.001$ | Cohen's $d = 3.54$ |
| SQR current form[c,*] sum score | 16.8 (8.2) | 4.9 (3.2) | $t(42) = 6.36$ | $p < 0.001$ | Cohen's $d = 1.96$ |
| SQR lifetime form[d,*] sum score | 23.0 (8.8) | 5.5 (3.3) | $t(42) = 8.75$ | $p < 0.001$ | Cohen's $d = 2.70$ |

For group comparison, results from independent-groups *t* tests are provided

*ASC* autism spectrum condition group, *NT* neurotypical comparison group

[a] Verbal and non-verbal IQ were assessed using the Wechsler-Intelligence Scale for Children (WISC-IV)

[b] Social Responsiveness Scale; discriminative cutoff: T-score $\geq 60$

[c] Social Communication Questionnaire current form

[d] Social Communication Questionnaire lifetime form; discriminative cut-off: sum score $\geq 10$

*For these measures, data from one individual with ASC and one NT control are missing respective group, because the criterion for group assignment was the presence/absence of an ASC diagnosis. Table 1 provides descriptive statistics and group comparisons for all measures.

## Material and design

The current study was based on the paradigms by Yoon et al. (2008) and Marno et al., (2014) and adapted for eye tracking. The within-participant factors were *context*, *change*, and *cue*. Each of these factors had two levels: The introductory sequence of the presented video clips was either ostensive and non-ostensive (context), the depicted target objects changed either in identity or location (change), and the target object was either cued by an actress or not (cue). We introduced group (ASC group vs. NT group) as between-participants factor. The interpretation of our results is based on the precondition that identity and location changes are equally difficult to detect for each participant. Any differences would constitute a confounding variable. Therefore, we followed the procedure of the original study by Marno et al. and included trials in which the participants had to detect the same types of changes, yet without the crucial context manipulation (four identity change trials and four location change trials). Performance in these trials was taken as a basis to compute the difference between the performances in location and object change trials. This baseline score was included as a covariate.

A set of unfamiliar, abstract three-dimensional objects was used as target objects. These objects consisted of a colored cube and a black T-shaped part that either pointed outward like a pin or pointed inward like an anchor (see also Träuble & Pauen, 2007; Träuble & Bätz, 2014). The scene in all video clips depicted a table with a chessboard-like surface and an occluder (20-cm high) on the far end of the table. An actress was sitting behind the occluder, so that only the head and parts of the shoulders were visible. Four target objects were placed on the table. Each object had a different color. The objects were evenly distributed on the chessboard-like grid of the table (left/right/front/back). Figure 1 displays the scene, the objects, and the sequence of events.

A trial started with the context phase. In the beginning of this phase, the four test objects were presented on the lower part of the scene. After approximately 3 s, the upper part of the scene with the actress was revealed. The actress either greeted the viewer by looking directly into the camera, smiling, and waving (ostensive condition), or she kept her face low, looked down at the table, and used

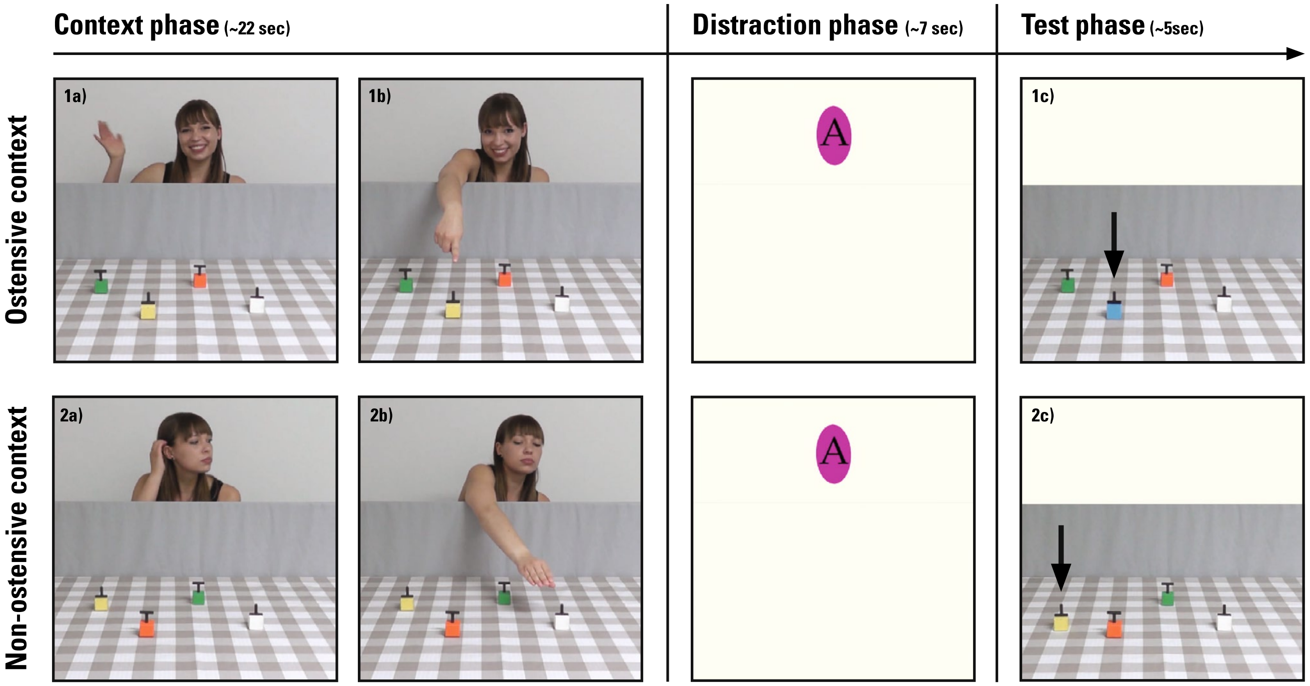

**Fig. 1** Task: examples of scene, trial sequence, and condition characteristics. (1a) Still frame from the ostensive context condition. The actress waved and smiled into the camera. (1b) Subsequently, she pointed to one of the far objects, looked into the camera and smiled. (2a) Still frame from the non-ostensive condition. The actress looked down at the table; afterwards, she used her hand to go through her hair. (2b) Subsequently, she grasped one of the objects (the sequence stopped before she would have touched it). (1c) Test phase in which gaze behavior was recorded. In this example, the cued object changed in identity (color). (2c) In this example of the test phase, the uncued object changed in location (moved to the front)

her hand to go through her hair (non-ostensive condition). Then, the actress cued one of the objects by either pointing at (ostensive condition) or trying to grasp it (non-ostensive condition). It is important to note that the ostensive and non-ostensive context differed in not only one, but several crucial signals/actions. We employed the present set of signals/action to closely follow Marno et al. (2014). A detailed empirical evaluation of the respective influence of each signal/action on task performance can be found in the original study (Marno et al., 2014).

The upper part of the scene was occluded again, so that, for another approximately 3 s, only the test objects on the lower part of the scene were visible. These events lasted approximately 22 s and were identical in all video clips. This sequence was followed by 7-s-long distraction phase showing a rotating pink oval on the top of a white screen, flashing a random letter for approximately 1 s. The participants had to name the presented letter as soon as they registered it. The purpose of this distractor task was (1) to increase the difficulty of the task by prolonging the time that the participants had to keep the presented objects in the working memory, and (2) to direct the participant's gaze to the upper part of the scene. This was necessary for an unbiased analysis of gaze patterns in the subsequent test phase, i.e., to introduce a reference/starting point for fixations just before the test objects are revealed in the lower part of the scene. In the subsequent test phase, the previous scene was revealed again. This time, the actress was absent and one of the objects had changed. The first type of change was an identity change: it was either swapped with a differently colored object (easy identity change) or with an object that had the same color, but the T-shaped part was inverted as compared to the previous one (difficult identity change). The second type of change was a location change: the object had either moved a short-distance backwards or forwards (difficult) or a long-distance backwards or forwards (easy). We employed two levels of difficulty, because the appropriate level of task complexity could not be determined in advance. The relevant previous studies either tested infants (Yoon et al., 2008) or adults (Marno et al., 2014) and hence used very simple or rather complex objects. Thus, we had to adapt the material for our age group of 10 years. The two levels of difficulty were introduced to increase variance in task performance and avoid floor or ceiling performance. In preliminary analyses, we found that the ASC and the comparison group did not differ in change detection performance in both levels of difficulty. Moreover, no systematic influence of task difficulty on the planned analyses was observed. Consequently, we collapsed the data across this factor. The test phase lasted for 5 s after which the screen turned black. These last 5 s of the video clips served as analysis window for the eye tracking analysis.

Subsequently, the experimenter asked which object had changed. Note that the cue was not informative with respect to which object had changed. In half of the trials, the cued object had changed; in the other half of the trials, another object had changed. The participants either responded verbally by naming the position of the object they thought had changed (1–4) or by pointing at the respective object. The experimenter started the next trial via button press, either after the participant's response, or after an interval of approximately 20 s. The test sessions were video-taped and the responses were coded manually.

In an additional second test question, the experimenter asked the participants what they thought had changed (either identity or location). We intended to use this test question for exploratory analyses. Because this variable adds little informational value, and because exploratory analyses revealed no additional insights, we do not further explicate on this variable.

A total of 40 videos were presented. Thirty two of them were from the experimental conditions (2 context × 2 change × 2 cue × 2 difficulty × 2 parallel versions = 32 videos). The remaining eight video clips were from the baseline condition (2 change × 2 difficulty × 2 parallel versions). The participants were randomly assigned to one of the four pseudorandomized sequences. The whole test session lasted about 40 min, with an optional brake after 20 trials.

## Apparatus and procedure

Participants sat on a chair with a distance of approximately 60 cm from a 17-inch TFT screen (1280 × 1024 pixel) integrated in a Tobii T60 eye tracker (60-Hz sampling rate; Tobii Technology, Stockholm, Sweden). A flexible monitor arm allowed for individual adjustment of the eye tracker's position. The stimuli were presented with Tobii Studio 3.1 (Tobii Technology), following a nine-point calibration procedure. A test session started with the eye tracking task, after a short instruction. Using pictures of objects different than the test objects, the experimenter explained all possible types of changes and the response format. Subsequently, verbal and non-verbal IQ tests were administered. Meanwhile, the children's caregiver(s) filled the SRS and SCQ. The present task was part of a larger test battery to investigate social interaction in ASC.

## Measures and data analysis

We calculated the percentage of correct responses for each condition. Explicit responses referring to the one out of the four objects that had changed in the test phase were coded as correct. Furthermore, gaze behavior during the test phase was analyzed. We were interested in the relative fixation duration on the correct object, indicating a preference for that object (cf., Corkum & Moore, 1998; Ferguson and Breheny, 2012). To this end, each object served as area of interest (AOI) for gaze data extraction. All four AOIs had the same size (ca. 160 × 180 pixel) and each covered 2.8% of the screen. We divided the total duration of all fixations on the correct object by the total duration of all fixations on all objects during the test period and multiplied this score by 100. The resulting score gives an estimate of the extent to which cognitive processing was biased towards the correct relative to the incorrect objects. The Tobii Studio standard fixation filter (velocity threshold of 35 pixels/window; distance threshold of 35 pixels) was used to separate fixations from saccades.

For the analysis of explicit responses and gaze behavior, we started with an omnibus analysis of covariance (ANCOVA). Following the procedure of Marno et al., (2014), we included the difference in performance between identity- and location-change baseline trials as a covariate in all analyses to account for individual differences in the sensitivity to detect identity or location changes. Within-participant factors were context (ostensive vs. non-ostensive), change (identity vs. location) and cue (cued vs. uncued). The between-participant factor was group (ASC vs. NT). This ANCOVA was performed to examine if there was a general cueing effect of ostensive and non-ostensive referential gestures to check whether explicit responses and visual attention allocation were influenced by this task manipulation. Furthermore, we were interested in any overall group differences in task performance.

Analogous to Marno et al. (2014), we specifically tested whether context had a differential influence on type of encoded object information. We performed a repeated-measures ANCOVA on explicit responses and gaze behavior only in trials in which the cued object had changed. Within-participant factors were context (ostensive vs. non-ostensive) and change (identity vs. location). In addition, within the ostensive context, we ran a repeated-measures ANCOVA with both measures with change (identity vs. location) and cue (cued vs. uncued) as within-participant factors. We were in the favorable situation to directly test whether our task, that can be considered as conceptual replication of Marno et al. (LeBel, McCarthy, Earp, Elson, & Vanpaemel, 2018; Schmidt, 2009), yields the same effects when running the same analyses. This confirmatory analysis, testing the prediction of a very specific pattern of results (especially the two predicted two-way interaction effects), substantially reduced multiple testing issues.

Data were analyzed using IBM SPSS Statistics 24 (SPSS Inc., Chicago, IL, USA) and JASP (JASP Team, Amsterdam, NL). The significance level for all frequentist analyses was $p \leq 0.05$. For Bayesian analyses, we used the default JZS prior (Wetzels, Grasman, & Wagenmakers, 2012). We chose this prior, because we did not have strong evidence for the expected effect size (Rouder, Speckman, Sun, & Morey, 2009; Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). In the evaluation of the resulting Bayes factors, we followed the convention that a Bayes factor (BF) > 10 yields strong evidence (Jeffreys, 1961).

## Results

### Explicit responses

Table 2 provides mean percentages of correct explicit responses, separately for each group. The repeated-measures omnibus ANCOVA revealed a significant main effect of cue, $F(1, 43) = 5.91$, $p = 0.019$, $\eta_p^2 = 0.12$, indicating the predicted cueing effect: Change detection performance was better in the cued ($M = 75.5\%$, SE = 2.7%) than in the uncued ($M = 65.9\%$, SE = 4.0%) conditions. Notably, the ASC and the NT group did not differ in detecting the object changes, $F(1, 43) = 0.81$, $p = 0.373$, $\eta_p^2 = 0.02$. The other main effects were also not significant [$Fs(1, 43) \leq 1.15$, $ps \geq 0.290$, $\eta_p^2 \leq 0.03$]. There was a significant two-way interaction between context and cue, $F(1, 43) = 6.98$, $p = 0.011$, $\eta_p^2 = 0.14$, indicating that depending on the context, cueing

**Table 2** Response phase: means (M) and standard errors (SE) of correctly detected object changes (in %), listed for each condition and group

| | ASC | | NT | |
|---|---|---|---|---|
| | *M* | SE | *M* | SE |
| Ostensive context | | | | |
| Identity change | | | | |
| Cued | 73.9 | 6.9 | 79.4 | 4.1 |
| Uncued | 64.1 | 5.4 | 62.0 | 6.8 |
| Non-ostensive context | | | | |
| Location change | | | | |
| Cued | 78.3 | 4.2 | 83.7 | 5.1 |
| Uncued | 60.9 | 7.3 | 71.7 | 6.7 |
| Identity change | | | | |
| Cued | 78.3 | 6.1 | 78.3 | 5.9 |
| Uncued | 62.0 | 6.3 | 60.9 | 7.2 |
| Location change | | | | |
| Cued | 60.9 | 6.3 | 70.7 | 5.1 |
| Uncued | 68.5 | 7.1 | 77.2 | 6.1 |
| Baseline | | | | |
| Identity change | 75.0 | 5.7 | 72.8 | 5.9 |
| Location change | 69.6 | 5.9 | 75.0 | 5.4 |

*ASC* autism spectrum condition group, *NT* neurotypical comparison group and non-cueing had a different influence on task performance. We also observed a significant interaction between change and cue, $F(1,43)=9.53$, $p=0.004$, $\eta_p^2=0.18$. This suggests that types of changes were differently detected in cued and uncued trials. No other two-way interaction was significant [$F$s$(1, 43) \leq 1.16$, $ps \geq 0.288$, $\eta_p^2 \leq 0.03$]. The three-way interaction between context, change, and cue was significant, $F(1, 43)=10.66$, $p=0.002$, $\eta_p^2=0.20$, showing that the significant context × cue interaction was different for change type, respectively, that the change × cue interaction was different for cue type. No other three-way interaction, nor the four-way interaction, were significant [$F$s$(1, 43) \leq 0.87$, $ps \geq 0.357$, $\eta_p^2 \leq 0.02$].

To break down these interactions and to test the specific influence of context on change detection, we performed a 2 (context: ostensive vs. non-ostensive) × 2 (change: identity vs. location) repeated-measures ANCOVA on trials in which the object that had changed was cued (see Fig. 2). The ASC and the NT group were collapsed due to lacking evidence for differential performance in the omnibus ANCOVA. This analysis revealed a significant main effect of context, $F(1, 44)=5.91$, $p=0.019$, $\eta_p^2=0.12$. Participants detected more changes in the ostensive ($M=78.8\%$, SE $=3.0\%$) than in the non-ostensive context ($M=72.0\%$, SE $=3.2\%$), which indicates that ostensive signaling improved change detection. The main effect of change was not significant, $F(1, 44)=1.10$, $p=0.299$, $\eta_p^2=0.02$. Overall, identity and location changes were equally well detected. The interaction between context and change was significant, $F(1, 44)=9.66$, $p=0.003$, $\eta_p^2=0.18$. Bonferroni-corrected post hoc $t$ tests (corrected significance level $p=0.013$) revealed that in the ostensive context, identity ($M=76.6\%$, SE $=3.7\%$) and location changes ($M=81.0\%$, SE $=3.3\%$) were equally well detected, $t(45)=-1.14$, $p=0.262$, Cohen's $d=-0.18$. In the non-ostensive context, the participants were better at detecting identity changes ($M=78.3\%$, SE $=4.2\%$) than location changes ($M=65.8\%$, $SE=4.1\%$), $t(45)=2.43$, $p=0.019$, Cohen's $d=0.44$. Identity change detection did not differ between contexts, $t(45)=-0.44$, $p=0.660$, Cohen's $d=-0.06$. However, location change detection was significantly worse in the non-ostensive as compared to ostensive context, $t(45)=3.58$, $p=0.001$, Cohen's $d=0.61$.

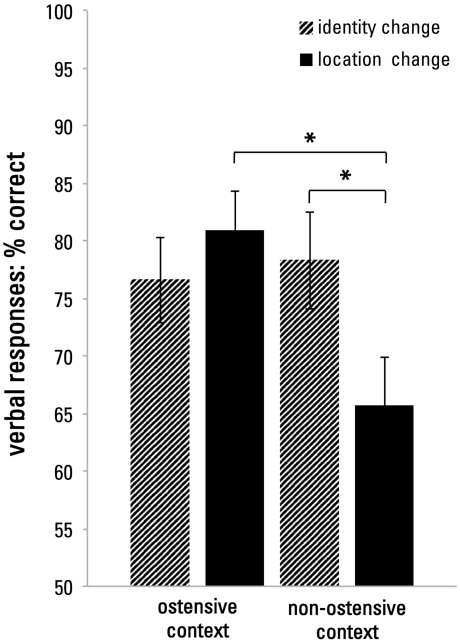

**Fig. 2** Explicit responses: mean percentage of correct responses (± SEM) for each condition. Responses referring to the one out of the four objects that had changed in the test phase were coded as correct. Post hoc $t$ test results indicate that the significant context × change within the cued objects only was mainly driven by relatively poor location change detection in the non-ostensive context

In the second planned ANCOVA, we considered only trials from the ostensive context condition and modeled change (identity vs. location) and cue (cued vs. uncued) as within-participant factors. The main effect of change was not significant, $F(1, 44)=1.48$, $p=0.231$, $\eta_p^2=0.03$. Within the ostensive context, identity and location changes were equally well detected. The significant main effect of cue, $F(1, 44)=11.52$, $p=0.001$, $\eta_p^2=0.21$, showed that changes of cued objects ($M=78.8\%$, SE $=3.0\%$) were better detected than changes of uncued objects ($M=64.7\%$, SE $=4.2\%$). Crucially, the change × cue interaction was not significant, $F(1, 44)=0.06$, $p=0.804$, $\eta_p^2<0.01$. This speaks against differential effect of cueing on the detection of the type of change within the ostensive pointing context.

Since we strived to interpret the null finding that children with and without ASC did not differ in their task performance, we examined this result with a complementary Bayesian repeated-measures ANOVA including the baseline performance differences between identity and location changes as a covariate using the default JZS prior. We were specifically interested in the interaction between context and group. The $BF_{01}$ of 19.88 provided strong evidence in favor of the null hypothesis that context had no differential influence on task performance depending on group.

## Relative fixation duration

Mean relative fixation durations for each condition and group are listed in Table 3. Consistent with the explicit responses, the repeated-measures omnibus ANCOVA of the distribution of visual attention during the test phase yielded a significant main effect of cue, $F(1, 43)=11.14$, $p=0.002$, $\eta_p^2=0.21$, attributable to higher relative fixation durations on the object that had changed when it was cued ($M=46.5\%$, SE $=1.6\%$) than when it was uncued ($M=38.8\%$, $SE=1.7\%$). However, baseline performance differences between identity and location changes, included as a covariate, interacted

**Table 3** Test phase: means (M) and standard errors (SE) of relative fixation durations (in%), listed for each condition and group

| | ASC | | NT | |
|---|---|---|---|---|
| | *M* | SE | *M* | SE |
| Ostensive context | | | | |
| Identity change | | | | |
| Cued | 42.8 | 3.5 | 42.3 | 3.0 |
| Uncued | 36.5 | 3.4 | 37.2 | 3.0 |
| Location change | | | | |
| Cued | 49.2 | 3.5 | 47.9 | 3.6 |
| Uncued | 39.4 | 2.8 | 44.9 | 3.1 |
| Non-ostensive context | | | | |
| Identity change | | | | |
| Cued | 49.0 | 3.8 | 48.6 | 3.0 |
| Uncued | 38.0 | 3.6 | 35.0 | 3.0 |
| Location change | | | | |
| Cued | 43.9 | 3.3 | 48.5 | 3.2 |
| Uncued | 37.8 | 3.8 | 41.3 | 3.1 |
| Baseline | | | | |
| Identity change | 43.2 | 2.9 | 42.6 | 2.2 |
| Location change | 39.6 | 3.1 | 43.8 | 3.0 |

The score was calculated by dividing total duration of all fixations on the correct object by the total duration of all fixations on all objects (times 100)

*ASC* autism spectrum condition group, *NT* neurotypical comparison group

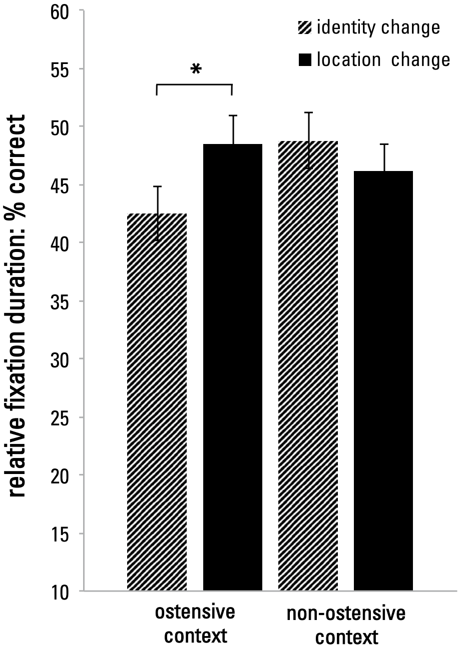

**Fig. 3** Gaze behavior during test phase: mean percentage (± SEM) of relative fixation duration on the object that had changed in the test phase. Post hoc *t* tests showed that in the ostensive context relatively more and/or longer fixations were directed to the object that had changed in location than to the object that had changed in identity

significantly with the factor cue, $F(1, 43)=4.36$, $p=0.043$, $\eta_p^2=0.09$. Thus, individual differences in change detection between identity and location changes have to be considered when interpreting this main effect. In contrast to the behavioral responses, also a significant main effect of change was observed, $F(1, 43)=5.50$, $p=0.024$, $\eta_p^2=0.11$, indicating a higher relative fixation duration on the object when it had changed in location ($M=44.1\%$, $SE=1.3\%$) as compared to when it had changed in identity ($M=41.2\%$, SE $=1.2\%$). The baseline performance significantly interacted with this factor, $F(1, 43)=8.28$, $p=0.006$, $\eta_p^2=0.16$, yielding that individual differences in the detection of the two types of changes account for some variance of this factor. Just as in the analysis of the behavioral responses, the main effect of context was not significant, $F(1, 43)=0.03$, $p=0.870$, $\eta_p^2<0.01$. Parallel to the explicit responses, the ASC and the NT group did not differ in their relative fixation duration on the object that had changed, $F(1, 43)=0.29$, $p=0.594$, $\eta_p^2=0.01$. Unlike in the explicit response, we found a significant context × change interaction, $F(1, 43)=5.93$, $p=0.019$, $\eta_p^2=0.12$, showing that depending on the context, types of changes resulted in differential relative fixation durations. No other two-way, three-way (which was not in line with the behavioral responses), nor the four-way interactions (which was parallel to the analysis of the behavioral response), were significant [$F$s$(1, 43)\leq 2.77$, $p$s $\geq 0.103$, $\eta_p^2\leq 0.06$].

Again, to follow up on the specific influence of context on gaze patterns for the two types of change, we performed a 2 (context: ostensive vs. non-ostensive) × 2 (change: identity vs. location) repeated-measures ANCOVA within cued object trials only (Fig. 3). Again, data were collapsed for the ASC and the NT group. In contrast to the explicit response findings, no significant main effect was observed. Relative fixation durations did neither differ between the ostensive and the non-ostensive context, $F(1, 44)=0.29$, $p=0.593$, $\eta_p^2=0.01$, nor between identity and location changes, $F(1, 44)=1.61$, $p=0.211$, $\eta_p^2=0.04$. Consistent with the explicit responses, also the analysis of relative fixation durations in the test phase revealed a significant interaction between context and change, $F(1, 44)=7.97$, $p=0.007$, $\eta_p^2=0.15$. Bonferroni-corrected post hoc $t$ tests (corrected significance level $p=0.013$) showed that in the ostensive context, objects which had changed in identity ($M=42.5\%$, SE $=2.3\%$) were significantly less fixated than objects that had changed in location ($M=48.5\%$, SE $=2.5\%$), $t(45)=-2.73$, $p=0.009$, Cohen's $d=-0.37$. In the non-ostensive context, objects that had changed in identity ($M=48.8\%$, SE $=2.4\%$) and location ($M=46.2\%$, SE $=2.3\%$) were equally long fixated. Across contexts, relative fixation durations differed neither for identity change, $t(45)=-1.87$, $p=0.068$, Cohen's $d=-0.39$, nor location change trials, $t(45) = 0.82$, $p = 0.416$, Cohen's $d = 0.12$.

The second ANCOVA addressing the influence of type of change and cueing in ostensive context trials only revealed a significant main effect of change, $F(1, 44) = 11.78$, $p = 0.001$, $\eta_p^2 = 0.21$, which was not observed in the analysis of the explicit responses. Within the ostensive context, objects that changed in location ($M = 45.3\%$, $SE = 1.4\%$) were longer fixated than objects that changed in identity ($M = 39.7\%$, $SE = 1.2\%$). Mirroring the analysis of explicit response, the main effect of cue was significant, $F(1, 44) = 4.15$, $p = 0.048$, $\eta_p^2 = 0.09$. In the ostensive context, cued objects ($M = 45.5\%$, $SE = 1.9\%$) were longer fixated that uncued objects ($M = 39.5\%$, $SE = 1.8\%$). As it was the case for the explicit responses, the change × cue interaction was not significant, $F(1, 44) = 0.03$, $p = 0.858$, $\eta_p^2 < 0.01$. Also the gaze behavior in the ostensive context revealed no differential effect of cueing on the detection of the type of change.

Also for gaze behavior, we followed up the lacking evidence for group effects with an analogous Bayesian repeated-measures ANOVA (baseline relative fixation duration for identity and location changes as a covariate, JZS prior). Consistent with the analysis of the explicit responses, the $BF_{01}$ of 45.50 provided very strong evidence in favor of the null hypothesis that context had no differential influence on relative fixation durations depending on group.

## Discussion

The current study investigated how the perception of others' actions and social cues affects object processing in children with and without ASC. In particular, the study had two aims. First, we tested a central prediction of natural pedagogy theory, that is, that ostensive cues lead to a selective processing of object features on the expense of object location. Second, we examined whether object processing in children with and without ASC is differently affected by ostensive cues. We expanded previous approaches by simultaneously assessing verbal responses and participants' gaze behavior. The current study yields two central findings. First, although we closely followed previous research designs (Marno et al., 2014; Yoon et al., 2008), and although we found an influence of ostensive vs. non-ostensive contexts on processing of object features, our results are not in line with the specific predictions derived from natural pedagogy theory. Rather, our results indicate that within the non-ostensive grasping context, participants rather noticed an identity change than a location change. Second, we did not find differences between children with and without ASC, neither on the behavioral nor on the gaze level. We will discuss the theoretical implications of our findings in the subsequent paragraphs.

### Action perception affects object processing

The first question focused on an empirical investigation of hypotheses derived from natural pedagogy theory (Csibra & Gergely, 2009). To this end, our experimental setup closely followed previous work (Marno et al., 2014; Yoon et al., 2008). It was predicted that in an ostensive context, people would preferentially encode the object's identity at the expense of its location (Marno et al., 2014). In a non-ostensive grasping context, though, there should either be no difference (Marno et al., 2014) or a preferential encoding of location information (Yoon et al., 2008).

Yet, our results are in strong contrast to the predictions as our pattern of results was completely reversed. There was no enhanced encoding of either object feature in the ostensive condition, whereas we found a preferential encoding of object identity in the non-ostensive grasping condition. This pattern of results in the two interactions that addressed our hypotheses was not only evident in the verbal change detection task, but also in participants' visual attention allocation as assessed by means of eye tracking (that is, the context x change interaction in cued trials only was significant in both measures, and the change x cue interaction in ostensive trials only was non-significant in both measures). Overall, this finding relates well to a recent set of studies that reported a failure to replicate an effect of ostensive cues on object encoding in infants (Silverstein et al., 2019). Moreover, it extends previous work by simultaneous assessment of explicit behavioral responses and looking times patterns. The finding that both measures converged on the same result render it unlikely that the explicit response format may have masked differential implicit processing or vice versa.

Our findings add to an intense debate on the empirical basis of natural pedagogy theory. That is, whereas some studies reported findings that are in line with natural pedagogy theory (e.g., Senju & Csibra, 2008; Träuble & Bätz, 2014; Yoon et al., 2008), did others report difficulties in replicating these findings (Silverstein et al., 2019) or provided more parsimonious explanations (de Bordes et al., 2013; Gredebäck et al., 2018). Overall, the current state of empirical evidence indicates that the empirical basis for natural pedagogy theory is unclear.

How else then to explain the difference between conditions? Interestingly, research on action–perception interrelations has shown that the intention to either grasp, reach, or point to an object affects the processing of action-related dimensions (e.g., shape and color) in visual search tasks (e.g., Bekkering & Neggers, 2002; Fagioli, Ferlazzo, et al. 2007; Fagioli, Hommel, et al. 2007; Wykowska, Schubö, & Hommel, 2009; for reviews see Schütz-Bosbach & Prinz, 2007; Witt, 2011). For example, Fagioli, Ferlazzo, et al. 2007; Fagioli, Hommel, et al. 2007) as well as Wykowska et al. (2009) demonstrated that either planning or observing a grasping action enhanced detection of changes in object features. The authors argued that preparing an action leads to the processing of action-related stimulus dimensions. Given that precise object features are relevant when preparing a grasp, these stimulus dimensions encoded in greater detail. This line of work relates to our study as (following Marno et al., 2014) the so-called ostensive context was characterized by the pointing action and the non-ostensive context by a grasping action. If we assume that the perception of another person's action leads to the activation of the same motor code in the observer (Paulus, 2012), our results can be related to the previously reported ones. Consistent with this line of research (Fagioli, Ferlazzo, et al. 2007; Fagioli, Hommel, et al. 2007; Wykowska et al., 2009), we found that, in the non-ostensive grasping condition, participants were better in detecting changes in object features than in object location. One should note, however, that detection of identity changes did not differ between conditions and that it, therefore, seems that the effect was driven by a reduced tendency to detect location changes in the non-ostensive condition. Yet, on the other hand, given the lack of a neutral baseline and given that both conditions differed with respect to a number of factors, it remains an open question to which extent the effect is due to facilitation or interference processes. We have to leave it to future research to address this issue.

### No difference between autistic and neurotypically developing children

The second main finding is that across several measures and analyses, we did not observe any difference between autistic and non-autistic participants. This is theoretically interesting as it is currently intensely debated to which extent specific differences in basic social information processes could lead to social interaction difficulties in ASC (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012; Shultz et al., 2015). Here, we tested the hypothesis that attentional guiding by ostensive cues and different action types might differ between the groups. Yet, this was not the case. Our results support the recent theoretical claims that action processing in ASC is far less impaired than previously thought (Cusack et al., 2015; Falck-Ytter, 2009; Marsh, Pearson, Ropar, & Hamilton, 2015; Sebanz, Knoblich, Stumpf, & Prinz, 2005). It is important to note that we did not only find no difference in explicit responses, but also for participants' gaze behavior. Given that the assessment of gaze behavior is often assumed to be indicative of processing strategies (cf., Frischen, Bayliss, & Tipper, 2007; Karatekin, 2007), this pattern of results also suggests a lacking difference in the attentional strategies underlying behavioral performance. It is possible, though, that while children with ASC have the same sensitivity to social cues as NT participants, they might require greater effort to process these cues (Aldaqre et al., 2016). Second, if we assume that our pattern of results could be explained by basic perception–action mechanisms (similar to Fagioli, Ferlazzo, et al. 2007; Fagioli, Hommel, et al. 2007; Wykowska et al., 2009), our findings suggest that the basic cognitive processes that connect the processing of (perceived) actions and object features are intact in children with ASC. Moreover, given that our task did not result in ceiling effects and was thus of a mid-level difficulty, it is unlikely that our task was merely too simple. Yet, given that this is a post hoc explanation of our results, further confirmatory work is required.

Overall, it remains an issue for future research to identify which cognitive characteristics contribute to the symptomatology observed in ASC. A currently discussed candidate mechanism is attenuated predictive processing of sensory information (Ganglmayer et al., 2019; Pellicano & Burr, 2012). Further, recent years have seen a shift away from attributing social interaction problems only to deficits of the autistic person towards focusing on the mismatch between interaction partners to explain social impairments (e.g., Bolis et al., 2017; Heasman & Gillespie, 2018).

## Limitations and conclusion

Although our study contributes to research on the impact of action perception, some limitations and open questions have to be considered. In the current design, our dependent variables were based on only four trials per condition. In order not to overburden our young participants, we had to restrict the overall number of trials. Notably, in this aspect, our method differs considerably from the original task by Marno et al. (2014), who assessed data from 36 trials per condition. Although this small number of trials is comparable to previous research (e.g., Yoon et al., 2008), a risk of increased noise has to be taken into account when interpreting the results. Furthermore, the present task was constructed in such a way that only very specific object features (shape, color, and location) played a role. It thus remains to be determined how generalizable our findings are to other situations in which object processing faces a less reduced set of relevant object features.

While the study informs theories on the nature of the difficulties associated with ASC, it should be noted that—like in most other studies—participants with typical verbal and non-verbal intelligence were examined. Our result can, therefore, not be generalized to autistic persons with more severe cognitive limitations. Future research is necessary to explore this question in greater detail. Moreover, like the vast majority of experimental research in this area, our paradigm—although presenting a social situation—is highly controlled and differs substantially from actual social interaction in a real-life context. It has been claimed that such setting fails to capture multifaceted aspects of autistic (and non-autistic) perception and cognition (De Jaegher, 2013). Future social cognitive research has to find ways to account for the nature and complexity of social interaction in naturalistic contexts.

Taken together, our study shows that while the perception of an ostensive pointing or a non-ostensive grasping action leads to different processing of object features, our results are exactly opposite to what was predicted by natural pedagogy theory. Moreover, no difference between the ASC and the comparison group suggests that these processes are not impaired in individuals with ASC.

**Acknowledgements** This study was funded by a grant from VolkswagenStiftung. We thank all children and their families for participating in this study. Furthermore, we are grateful to Tabea Schädel, Veronika Eisenschmid, Verena Rampeltshammer, and Teresa Wenhart for their help in data acquisition. We thank Martin Sobanski (Heckscher-Klinikum gGmbH) and Martina Schabert (autkom) for their help with recruiting participants.

## References

Aldaqre, I., Paulus, M., & Sodian, B. (2015). Referential gaze and word learning in adults with autism. *Autism, 19*(8), 944–955.

Aldaqre, I., Schuwerk, T., Daum, M. M., Sodian, B., & Paulus, M. (2016). Sensitivity to communicative and non-communicative gestures in adolescents and adults with autism spectrum disorder: saccadic and pupillary responses. *Experimental Brain Research, 234*(9), 2515–2527.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs: Prentice Hall.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences, 12*(7), 254–258.

Bekkering, H., & Neggers, S. F. (2002). Visual search is modulated by action intentions. *Psychological Science, 13*(4), 370–374.

Böckler, A., van der Wel, R. P., & Welsh, T. N. (2014). Catching eyes: effects of social and nonsocial cues on attention capture. *Psychological Science, 25*(3), 720–727.

Böckler, A., van der Wel, R. P., & Welsh, T. N. (2015). Eyes only? Perceiving eye contact is neither sufficient nor necessary for attentional capture by face direction. *Acta Psychologica, 160,* 134–140.

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., & Schilbach, L. (2017). Beyond autism: introducing the dialectical misattunement hypothesis and a bayesian account of intersubjectivity. *Psychopathology, 50*(6), 355–372.

Bölte, S., & Poustka, F. (2006). *Fragebogen zur sozialen Kommunikation (FSK) [Social Communication Questionnaire]*. Bern: Huber.

Bölte, S., & Poustka, F. (2008). *Skala zur Erfassung sozialer Reaktivität (SRS) [Social Responsiveness Scale]*. Bern: Huber.

Boyd, R., & Richerson, P. J. (1988). *Culture and the Evolutionary Process*. Chicago: University of Chicago press.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences, 16*(4), 231–239.

Constantino, J. N., & Gruber, C. P. (2005). *Social Responsiveness Scale (SRS)*. Los Angeles: Western Psychological Services.

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology, 34*(1), 28.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind and Language, 25,* 141–168.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences, 13*(4), 148–153.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1567), 1149–1157.

Csibra, G., & Shamsudheen, R. (2015). Nonverbal generics: Human infants interpret objects as symbols of object kinds. *Annual Review of Psychology, 66,* 689–710.

Cusack, J. P., Williams, J. H., & Neri, P. (2015). Action perception is intact in autism spectrum disorder. *Journal of Neuroscience, 35*(5), 1849–1857.

de Bordes, P. F., Cox, R. F., Hasselman, F., & Cillessen, A. H. (2013). Toddlers' gaze following through attention modulation: Intention is in the eye of the beholder. *Journal of Experimental Child Psychology, 116*(2), 443–452.

De Jaegher, H. (2013). Embodiment and sense-making in autism. *Frontier in Integrative Neuroscience, 7,* 15.

Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: a referential coding account. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1248.

Fagioli, S., Ferlazzo, F., & Hommel, B. (2007a). Controlling attention through action: Observing actions primes action-related stimulus dimensions. *Neuropsychologia, 45*(14), 3351–3355.

Fagioli, S., Hommel, B., & Schubotz, R. I. (2007b). Intentional control of attention: Action planning primes action-related stimulus dimensions. *Psychological Research, 71*(1), 22–29.

Falck-Ytter, T. (2009). Young children with autism spectrum disorder use predictive eye movements in action observation. *Biology Letters, 6*(3), 375–378.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology, 48*(1), 257–263.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694.

Frith, U. (2012). Why we need cognitive explanations of autism. *The Quarterly Journal of Experimental Psychology, 65*(11), 2073–2092.

Frith, C., & Frith, U. (2008). Implicit and Explicit Processes in Social Cognition. *Neuron, 60*(3), 503–510.

Ganglmayer, K., Schuwerk, T., Sodian, B., & Paulus, M. (2019). Do children and adults with autism spectrum condition anticipate others' actions as goal-directed? A predictive coding perspective. *Journal of Autism and Developmental Disorders, 8,* 1–13.

Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze following is not dependent on ostensive cues: a critical test of natural pedagogy. *Child Development, 89,* 2091–2098.

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience and Biobehavioral Reviews, 42,* 279–297.

Hamilton, A. F. D. C. (2015). Cognitive underpinnings of social interaction. *The Quarterly Journal of Experimental Psychology, 68*(3), 417–432.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*(4), 596–615.

Heasman, B., & Gillespie, A. (2018). Perspective-taking is two-sided: Misunderstandings between people with Asperger's syndrome and their family members. *Autism, 22*(6), 740–750.

Heyes, C. (2016). Born pupils? Natural pedagogy and cultural pedagogy. *Perspectives on Psychological Science, 11*(2), 280–295.

IBM SPSS Statistics, Version 24.0 [Computer software]. Armonk: IBM Corp.

JASP, Version 0.9 [Computer software]. Amsterdam, NL: JASP Team.

Jaswal, V. K., & Akhtar, N. (2018). Being vs. appearing socially uninterested: challenging assumptions about social motivation in autism. *The Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X18001826.

Jeffreys, H. (1961). *The Theory of Probability*. Oxford: Oxford University Press.

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review, 27*(3), 283–348.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1499), 2021–2031.

Kunde, W., Weller, L., & Pfister, R. (2018). Sociomotor action control. *Psychonomic Bulletin & Review, 25*(3), 917–931.

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science, 1*(3), 389–402.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

Marno, H., Davelaar, E. J., & Csibra, G. (2014). Nonverbal communicative signals modulate attention to object properties. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 752–762.

Marsh, L. E., Pearson, A., Ropar, D., & Hamilton, A. D. C. (2015). Predictive gaze during observation of irrational actions in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders, 45*(1), 245–261.

Moore, C. (2013). *The Development of Commonsense Psychology*. London: Psychology Press.

Palmer, C. J., Paton, B., Kirkovski, M., Enticott, P. G., & Hohwy, J. (2015). Context sensitivity in action decreases along the autism spectrum: a predictive processing perspective. *Proceedings of the Royal Society of London B: Biological Sciences, 282*(1802), 20141557.

Paulus, M. (2012). Action mirroring and action understanding: An ideomotor and attentional account. *Psychological Research, 76,* 760–767.

Paulus, M., Murillo, E., & Sodian, B. (2016). When the body reveals the mind: Children's use of others' body orientation to understand their focus of attention. *Journal of Experimental Child Psychology, 148,* 101–118.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504–510.

Petermann, F., & Petermann, U. (2007). *Hamburg-Wechsler Intelligenztest für Kinder—IV (HAWIK-IV)*. Bern: Huber.

Rouder, J. N., Speckman, P. J., Sun, D., & Morey, R. D. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100.

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods, 22*(2), 322.

Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: Action-induced modulation of perception. *Trends in Cognitive Sciences, 11*(8), 349–355.

Schuwerk, T., Sodian, B., & Paulus, M. (2016). Cognitive mechanisms underlying action prediction in children and adults with autism spectrum condition. *Journal of Autism and Developmental Disorders, 46*(12), 3623–3639.

Schwartz, S., & Susser, E. (2011). The use of well controls: an unhealthy practice in psychiatric research. *Psychological Medicine, 41*(6), 1127–1131.

Sebanz, N., Knoblich, G., Stumpf, L., & Prinz, W. (2005). Far from action-blind: Representation of others' actions in individuals with autism. *Cognitive Neuropsychology, 22*(3–4), 433–454.

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology, 18*(9), 668–671.

Shultz, S., Jones, W., & Klin, A. (2015). Early departures from normative processes of social engagement in infants with autism spectrum disorder. *The Many Faces of Social Attention* (pp. 157–177). Berlin: Springer International Publishing.

Silverstein, P., Gliga, T., Westermann, G., & Parise, E. (2019). Probing communication-induced memory biases in preverbal infants: Two replication attempts of Yoon, Johnson and Csibra (2008*). Infant Behavior and Development, 55,* 77–87.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology, 26,* 4–7.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition* (2nd ed.). Oxford: Blackwell.

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports, 4,* 5304.

Tobii Studio (3.1) [computer software]. Stockholm, Sweden: Tobii Technology.

Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., & Baldassarre, G. (2013). Theories and computational models of affordance and mirror systems: an integrative review. *Neuroscience and Biobehavioral Reviews, 37*(3), 491–521.

Träuble, B., & Bätz, J. (2014). Shared function knowledge: Infants' attention to function information in communicative contexts. *Journal of Experimental Child Psychology, 124,* 67–77.

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition, 105*(2), 362–379.

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

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-WISC-IV.* Psychological Corporation.

Wetzels, R., Grasman, R. P., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for ANOVA designs. *The American Statistician, 66*(2), 104–111.

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders, 34*(3), 285–299.

Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science, 20*(3), 201–206.

World Health Organization. (1993). *ICD-10: The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.

Wykowska, A., Schubö, A., & Hommel, B. (2009). How you move is what you see: action planning biases selection in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 35*(6), 1755–1769.

Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences, 105*(36), 13690–13695.