Development of online use of theory of mind during adolescence: An eye-tracking study

Irene Symeonidou [a,*], Iroise Dumontheil [b], Wing-Yee Chow [a], Richard Breheny [a]

[a] Division of Psychology and Language Sciences, University College London, London WC1N 1PF, UK
[b] Department of Psychological Sciences, Birkbeck, University of London, London WC1E 7HX, UK

ARTICLE INFO

ABSTRACT

We investigated the development of theory of mind use through eye-tracking in children (9–13 years old, $n = 14$), adolescents (14–17.9 years old, $n = 28$), and adults (19–29 years old, $n = 23$). Participants performed a computerized task in which a director instructed them to move objects placed on a set of shelves. Some of the objects were blocked off from the director's point of view; therefore, participants needed to take into consideration the director's ignorance of these objects when following the director's instructions. In a control condition, participants performed the same task in the absence of the director and were told that the instructions would refer only to items in slots without a back panel, controlling for general cognitive demands of the task. Participants also performed two inhibitory control tasks. We replicated previous findings, namely that in the director-present condition, but not in the control condition, children and adolescents made more errors than adults, suggesting that theory of mind use improves between adolescence and adulthood. Inhibitory control partly accounted for errors on the director task, indicating that it is a factor of developmental change in perspective taking. Eye-tracking data revealed early eye gaze differences between trials where the director's perspective was taken into account and those where it was not. Once differences in accuracy rates were considered, all age groups engaged in the same kind of online processing during perspective taking but differed in how often they engaged

* Corresponding author.
  E-mail address: i.symeonidou@ucl.ac.uk (I. Symeonidou).

in perspective taking. When perspective is correctly taken, all age groups' gaze data point to an early influence of perspective information.

Introduction

Over the last couple of decades, cognitive neuroscience research has shown that brain areas involved in theory of mind (the "social brain")—our ability to attribute the beliefs, thoughts, desires, intentions, and feelings of others—undergo changes not only during childhood but also during adolescence (Burnett, Sebastian, Cohen Kadosh, & Blakemore, 2011). A substantial number of studies have now provided evidence for structural and functional changes in the social brain during childhood and adolescence. In addition, there is a body of evidence that theory of mind (ToM) is applied more robustly or accurately with age through middle childhood (Devine & Hughes, 2013; Epley, Morewedge, & Keysar, 2004; Lecce, Bianco, Devine, Hughes, & Banerjee, 2014; Surtees & Apperly, 2012; Wang, Ali, Frisson, & Apperly, in press) and adolescence (Dumontheil, Apperly, & Blakemore, 2010; Vetter, Altgassen, Phillips, Mahy, & Kliegel, 2013). Previous studies have shown that children's performance on paradigms such as false-belief tasks reaches ceiling at around the age of 5 years (Surian, Caldi, & Sperber, 2007; Wellman, Cross, & Watson, 2001). The question arises as to what factors affect older children and adolescents' successful application of such abilities.

The Director paradigm has been used to investigate the ability to take the perspective of another individual into account in a communicative context (Apperly, Back, Samson, & France, 2008; Brown-Schmidt & Hanna, 2011; Fett et al., 2014; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). In these studies, the participant interacts with another agent (a "director") to act on a set of objects (Director paradigm; Fig. 1). Crucially, some of the objects are blocked off from the director's point of view and are visible only to the participant. Thus, when the director talks about an object (e.g., "the large ball"; Fig. 2), the participant should ignore any object that is not visible to the director and instead select a referent from what is in the "common ground," that is, what is visible to both the participant and the director. This paradigm requires the participant to infer the speaker's referential intention (a mental state) based on beliefs that differ from his or her own due to the speaker's ignorance of the presence of an object that would be a potential referent for the instruction given. For example, in the setup shown in Fig. 2, the participant, but not the director, sees a third ball that best fits the description "the large ball" (the basketball) and needs to discount it as the intended referent because the director does not know about that ball.

The Director paradigm is useful for the study of the development of the social brain because it can be used to measure the application of aspects of ToM without asking participants to make an explicit
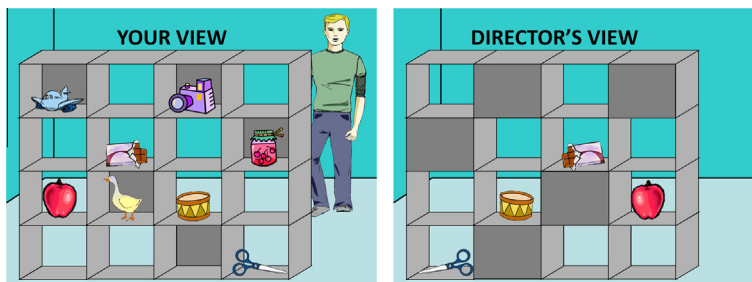

Fig. 1. Instruction screens used to explain the task to participants. The set of shelves on the left is shown from the participant's point of view, and the set of shelves on the right is shown from the director's point of view.
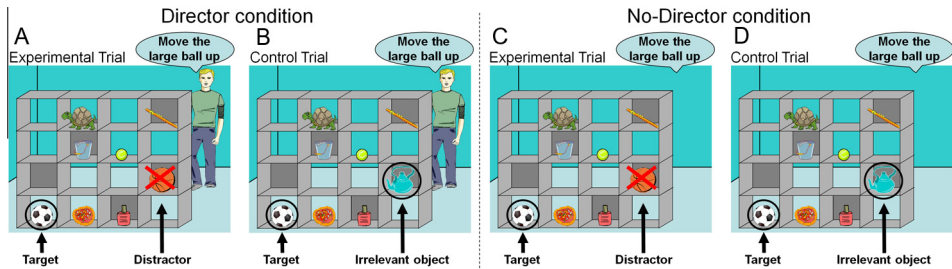
Fig. 2. Stimuli of the perspective-taking task. (A and B) Director condition. The participant heard the instruction "Move the large ball up" from the director. In an Experimental trial (A), if the participant did not take the director's perspective into account, he or she would move the basketball instead of the football, which is the second largest and can be seen by both the participant and the director. In a Control trial (B), the distractor is replaced by an irrelevant object. (C and D) No-Director condition. The participant is told that instructions do not refer to items in slots with a gray background; therefore, the correct responses are the same as in the Director condition.

judgment about their own or someone else's perspective or thoughts. Given that even adults manifest less than ceiling performance on the Director paradigm (Brown-Schmidt & Hanna, 2011; Keysar et al., 2000), it is well suited for exploring how ToM abilities develop across childhood and adolescence.

Early visual world eye-tracking studies using the Director paradigm demonstrated that adult participants are not able to ignore objects to which they have privileged visual access and that they are less accurate in choosing the object that is visually available to both participants and the director compared with control conditions (Epley et al., 2004; Keysar et al., 2000, 2003). Keysar and colleagues explained these results by suggesting that participants have an initial bias to take an egocentric perspective and that their initial interpretation of the instruction is later adjusted according to the speaker's knowledge state by a second process that corrects any manifest errors.

In contrast, studies using very similar paradigms have found evidence that participants are able to integrate information about the speaker's beliefs immediately (Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008). This is comparable to findings that individuals may rapidly and automatically compute what other people see (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). It seems clear that adults are able to integrate information to some extent about what their interlocutor knows from the earliest stages of referential language processing even though they are liable to attend to objects that are not known to the speaker (Brown-Schmidt & Hanna, 2011). Among adults, performance on perspective-taking tasks has been shown to be affected by factors such as the nature of the verbal stimulus (Barr, 2008), the extent to which the interaction offers information about the interlocutor's mental state (Brown-Schmidt, 2009a), cultural background (Wu & Keysar, 2007), inhibitory control (Brown-Schmidt, 2009b; Nilsen & Graham, 2009), and mood (Converse, Lin, Keysar, & Epley, 2008).

Dumontheil and colleagues (2010) tested 7- to 27-year-old female participants' performance using a computerized version of Keysar and colleagues' (2000) Director task. Participants were presented with a 4 × 4 grid that contained various objects in different slots. A director (an avatar) gave participants instructions about which objects to move and where to move them. As in the original design, some of the slots on the shelves were occluded so that the director could not see what was present on those shelves. The study also included a No-Director condition where participants performed the same task in the absence of a director. In this condition, they were told that the instructions would not refer to items in slots with a gray background. According to Dumontheil and colleagues (2010), the difference between the two conditions is that whereas only executive function processes are required for participants to perform well on the task in the No-Director condition, both ToM processes and executive function processes are required in the Director condition because participants need to take someone else's perspective into account in order to perform well.

Dumontheil and colleagues (2010) found that error rates were higher in the late adolescent group (14- to 17-year-olds) than in the adult group in the Director condition but that they did not differ

between the late adolescent and adult groups in the No-Director condition, which relies on executive functions only. This suggests that late adolescents differ from adults in their application of ToM inference when controlling for certain executive function demands.

In contrast to the error rates, Dumontheil and colleagues (2010) reported that reaction times on the critical trials in the Director condition remained the same across all age groups. Furthermore, they found that reaction times were longer in the No-Director condition than in the Director condition. According to the authors, this difference suggests that participants approach the Director task in a way that is more efficient than just applying an arbitrary rule. They proposed that the difference in accuracy between the late adolescent and adult groups might stem not from how efficiently or rapidly they can make referential inferences but rather from their propensity to integrate information about the speaker's beliefs in making those inferences. These suggestions point to a response to the question raised above as to what factors are responsible for the better application of ToM abilities during this stage of development. The current study used visual world eye-tracking measures on a variant of the Director task employed by Dumontheil and colleagues (2010).

It is widely accepted that eye fixations can indicate how we process information during spoken language comprehension (Altmann & Kamide, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). By measuring when participants fixate their gaze on an object, we can identify which object they are considering as a possible referent at a given point in time. Although eye-tracking has previously been used in variants of the Director task, samples of these studies have been limited to either adults (Keysar et al., 2003; Lin, Keysar, & Epley, 2010; Wu & Keysar, 2007) or younger children (5- and 6-year-olds in Nadig & Sedivy, 2002; 4- to 12-year-olds in Epley et al., 2004). The current study afforded an opportunity to examine and compare how children, adolescents, and adults apply ToM in real time through monitoring their eye gaze while they were performing the task. Specifically, the aim was to explore the time course of the decision process that yields both correct and incorrect responses in participants.

Furthermore, the current study also aimed to understand whether ongoing development of executive function processes, specifically inhibition, plays a role in perspective taking. Previous studies have found a correlation between inhibition and the application of ToM inference in children (3- and 4-year-olds in Carlson, Moses, & Claxton, 2004; 4- to 10-year-olds in Hansen Lagattuta, Sayfan, & Harvey, 2014; 4- to 9-year-olds in Lagattuta, Sayfan, & Blattman, 2010; 2- to 5-year-olds in Nilsen & Graham, 2009) and in adults (Brown-Schmidt, 2009b). A meta-analysis of 100 studies with 3- to 6-year-olds by Devine and Hughes (2014) highlighted the association between executive function and ToM. Inhibitory control, like other executive skills, is still maturing during adolescence (Leon-Carrion, García-Orza, & Pérez-Santamaría, 2004; Luna, Padmanabhan, & Hearn, 2011). A study by Vetter and colleagues (2013) found that inhibitory control, as measured with an anti-saccade task, predicted age-related variance in affective ToM during adolescence. No other study, however, has looked at the role of inhibitory control in adolescents' ability to inhibit one's perspective and consider someone else's perspective. Therefore, we measured participants' inhibitory control through two variants of the Go–NoGo task in order to investigate whether individual differences in inhibitory control can account for their performance on the Director task.

Based on the evidence of ongoing functional development of neural processes related to ToM and the results of Dumontheil and colleagues (2010), we expected that age-related differences in participants' accuracy would be greater in the Director condition than in the No-Director condition. We also expected to see age-related differences in inhibitory control, which may partially account for participants' performance on the Director task. We measured the extent to which participants considered the target object relative to the distractor by computing the target advantage score, which is the average probability of looks at the target object minus the average probability of looks to the distractor (Brown-Schmidt, 2009b; Kronmüller & Barr, 2015). This allowed us to examine and compare how each age group used ToM to identify the referent.

According to the perspective adjustment model proposed by Keysar and colleagues, participants have an initial egocentric bias, which is then adjusted, during a second process, to the speaker's perspective. Therefore, this theory predicts that participants will fixate on the distractor first and, if the second process of adjusting is successful (i.e., on correct trials), then participants will fixate their gaze on the target object and give a correct response. However, if the second process fails and perspective

taking does not occur, then we would expect participants' gaze to remain on the distractor, not to fixate on the target, and participants to eventually give an incorrect response. As such, it is predicted that although both adolescents and adults would fixate on the distractor object initially, demonstrating an initial egocentric bias, adults would then be more likely to switch to fixating their gaze on the target object and do so more quickly than adolescents. In other words, both adolescents and adults would have a smaller target advantage initially, but adults would have a bigger target advantage score than adolescents in earlier time regions.

Other models of perspective taking found in Hanna and colleagues (2003) and Brown-Schmidt, Gunlogson, and Tanenhaus (2008) suggest that perspective information is available from the outset but that the extent to which it is integrated online depends on how sensitive participants are to such information (Brown-Schmidt, 2012) or on participants' ability to integrate that information with linguistic and visual contextual information (Hanna et al., 2003). According to these proposals, an incorrect response may reflect a simple failure to integrate perspective from the start of a trial. If this is the case, then patterns of eye gaze may differ from the onset of the critical phase of the instruction between correct and incorrect trials.

Method

Participants

A total of 65 participants took part in this study, constituting three age groups: child (n = 14, 9–13 years old, M = 11.2 years, SD = 1.2), adolescent (n = 28, 14–17.9 years old, M = 16.2 years, SD = 1.1), and adult (n = 23, 19–29 years old, M = 23.0, SD = 2.8). The age groups were divided to match the age groups of Dumontheil and colleagues (2010). Adult participants were recruited from the University College London (UCL) Psychology participant pool, whereas children and adolescents were recruited from London schools. All participants were native English speakers. Data from three adults were excluded due to technical errors. We also excluded data from two adolescents who did not understand the instructions of the No-Director condition. Verbal ability was measured in all participants using the vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). Average verbal IQ scores were 124 (SD = 8.0) for adults, 118 (SD = 6.9) for adolescents, and 123 (SD = 10.2) for children. A one-way analysis of variance (ANOVA) showed a significant difference in verbal IQ scores between groups, $F(2,57)$ = 5.97, p = .03. Bonferroni-corrected post hoc comparisons showed that adolescents performed worse than adults (p = .04), but verbal IQ did not significantly differ between children and adolescents (p = .22) or between children and adults (p = 1.00).

Parents/guardians of all child and adolescent participants as well as the adult participants were given information sheets prior to the study. Informed consent was obtained from the parents/guardians for all child/adolescent participants and from all adult participants. This study was approved by the UCL research ethics committee.

Design

Director task

The current experiment had two within-participant factors: condition (Director or No Director) and trial type (Control or Experimental). It had one between-participants factor: age group (child, adolescent, or adult). The task design and stimuli were based on a previous study by Dumontheil and colleagues (2010; see also Apperly et al., 2010). Participants were presented with a visual scene of a 4 × 4 set of shelves containing eight different objects and were asked to move one of the objects in each trial.

In the Director condition, a "director" was shown standing behind the shelves, viewing the shelves from behind (Fig. 1). Participants were asked to listen to the director's instructions to move one of the objects (e.g., "Move the large ball up") and respond. Participants were told that they should take the director's viewpoint into account when following the director's instructions. They were told that objects in slots with a gray background were visible only to them, whereas the other objects could

be seen from either side of the shelves. In Experimental trials, the instruction referred to one object ("target") given the director's point of view but would refer to another object ("distractor") if one assumed participants' perspective (Fig. 2A). As such, participants needed to take the director's perspective into account in order to respond correctly in an Experimental trial. In Control trials, the distractor object was replaced by an irrelevant object and the instruction referred to an object that was visible to both participants and the director (Fig. 2B). In Filler trials, instructions referred to single objects that were visible to both participants and the director (e.g., the turtle in Fig. 2).

In the No-Director condition, the only difference was that the director was removed from the stimuli (Fig. 2C and D). Participants were told that the auditory instructions would refer only to items in clear slots and not items that were in slots with a gray background.

A total of 48 pairs of shelf configurations were created as Experimental and Control trials. All shelf configurations depicted eight objects and included either three (Experimental trials) or two (Control trials) exemplars of the same object that differed in position (top/bottom) or size (large/small). In Experimental trials, the exemplars were distributed such that the distractor object (the top-most, bottom-most, smallest, or largest object) was in a slot with a gray background, whereas the target object (the second top-most, bottom-most, smallest, or largest object) and the third object ("Object 2") were in a clear slot (Fig. 2A and C). In the Control trials, the distractor was replaced by an irrelevant object (Fig. 2B and D). The rest of the objects were unique objects distributed among three slots with a gray background and two clear slots. Another 48 pairs of shelf configurations were created for the Filler trials, in which the instructions referred to single objects that were in a clear slot. Taken together, the stimuli were divided such that half were presented in the Director condition and the other half were presented in the No-Director condition (counterbalanced across participants), and each participant saw 12 Experimental trials, 12 Control trials, and 24 Filler trials in the Director and No-Director conditions. The order of stimulus presentation was counterbalanced between participants.

The materials and design of the current study differed from those Dumontheil and colleagues' (2010) study in three ways. First, we included a greater number of Experimental and Control trials to increase statistical power and allow eye-tracking data analysis. Second, whereas each shelf configuration was used with three successive instructions in Dumontheil and colleagues' study, a different shelf configuration was used in each trial in the current study to minimize participants' learning strategies. Third, whereas participants needed to pretend to drag the object in the previous study, they were able to click and drag any object on the grid in the current study.

Inhibitory control tasks

To measure participants' inhibitory control, we used two different Go–NoGo tasks, both of which had one within-participant factor trial type (Go or NoGo). The Simple Go–NoGo task was based on the standard Go–NoGo paradigm (Simmonds, Pekar, & Mostofsky, 2008). A colored square was presented on the left or right side of the screen in each trial. Participants needed to indicate which side of the screen it appeared on if the square was green (a Go trial) and to inhibit their response if the square was red (a NoGo trial). The Complex Go–NoGo task was identical except that it used yellow and blue squares and included a 1-back working memory (WM) requirement (see Simmonds et al., 2008), such that participants needed to indicate on which side of the screen the square was shown (Go trials) except when a blue square was preceded by a yellow square (NoGo trials).

Procedure

Participants were tested individually in one session lasting approximately 45 min. They completed the various tasks in the following order: (1) the Director task (Director condition and then No-Director condition), (2) the inhibitory control tasks (Simple Go–NoGo task and then Complex Go–NoGo task), and (3) the vocabulary subtest of the WASI. Eye movements during the Director task were recorded using a Tobii TX300 eye-tracker at a sampling rate of 60 Hz. Stimuli for the Director task were presented using E-Prime 2, and the Go–NoGo tasks were programmed in Cogent (http://www.vislab. ucl.ac.uk/Cogent/index.html) running in Matlab 7.0 (MathWorks).

Standardized instructions were read to participants prior to the Director task. For the Director condition, participants were told that the director had a different view of the shelves (Fig. 1) and that the

director's point of view must be considered when following the director's instructions to move objects. We asked participants to give an example of an object that both they and the director could see as well as an object that the director could not see to ensure that participants understood the task. For the No-Director condition, participants were told that the director was no longer present and that instructions would refer only to items in clear slots, such that they should ignore items in slots with a gray background when performing the task. The Director condition was always presented prior to the No-Director condition in order to prevent participants from applying the rule provided in the No-Director condition (Dumontheil et al., 2010). Participants were presented with two Filler practice trials before each condition.

In each trial, the visual stimulus and auditory instructions were presented over a period of 2.2 s. The visual stimulus remained on the screen for another 3.8 s, during which participants responded by clicking an object and dragging it to a different position. A response was considered correct if the target object had been selected. Response times (RTs) were measured from the onset of the display.

In the Go–NoGo tasks, each square was presented for 400 ms, with a 600- to 800-ms jittered fixation cross intertrial interval (ITI). Participants responded by pressing the left or right key using their right index or middle finger, respectively (adapted from Watanabe et al., 2002). Participants performed two practice blocks prior to each task; the first practice block presented 10 Go trials to establish a habitual response, and the second one presented 6 Go trials and 4 NoGo trials. Practice was repeated if participants made two or more errors. Each participant performed 80 test trials on each task (25% NoGo).

Data analysis

Behavioral data analyses

All analyses were processed in SPSS and R. Statistical significance was set at $p < .05$. Bonferroni-corrected post hoc *t*-tests were performed to explore significant main effects and interactions further.

Director task. Mixed repeated measures ANOVAs with two within-participant factors (condition and trial type) and one between-participant factor (age group) were performed on participants' mean accuracy (percentage errors) and reaction times (RTs). Data for Filler trials were not analyzed; participants made fewer than 2% errors on these trials on average. Because verbal IQ differed among groups, we conducted an additional mixed repeated measures ANOVA that included verbal IQ as a covariate.

Inhibitory control tasks. For both the Simple and Complex Go–NoGo tasks, mean accuracy (percentage errors) was calculated for each participant in each trial type (Go or NoGo) and median RT was calculated for correct Go trials. A 2×3 mixed ANOVA with a within-participant factor trial type (Go or NoGo) and a between-participant factor age group (child, adolescent, or adult) was performed for each task. One-way ANOVAs were performed to examine the effect of age group on participants' RTs in Go trials in each task.

Association between Director and inhibitory control tasks. Regression analyses were performed to investigate whether age-related changes in Director task performance were associated with performance on the inhibitory control tasks. The difference in percentage errors between Director and No-Director Experimental trials, which is the critical measure of interest, was entered as the dependent variable. In a first step, age was entered as a continuous variable. In a second step, the four percentage error measures of the Simple and Complex Go–NoGo trials were entered in a stepwise regression. Finally, in a third step, IQ was entered to assess whether it accounted for variance associated with age in this model.

Eye-tracking data analyses

Eye movements were analyzed by computing a target advantage score, which is the average probability of looks at the target object minus the average probability of looks at the distractor (Kronmüller & Barr, 2015). A look to the target was defined as a look to the object or the slot in which the object

was located. Likewise, a look to the distractor (or the irrelevant object in Control trials) was defined as a look to the object or the slot in which the object was located. Note that because the distractor was an irrelevant object in the Control trials, target advantage scores in the Control trials represented the difference between looks to the target and looks to the irrelevant object.

Target advantage scores were calculated over 50-ms time bins for the figures where 0 ms was the noun onset. For statistical analyses, the target advantage scores were calculated for five different time windows (regions) during a trial. The five regions were time-locked to the onset of words in the auditory stimuli. Region 1 was the verb ("move"), Region 2 was the article ("the"), Region 3 was the modifier ("large"), Region 4 was the critical noun region ("ball"), and Region 5 was the directional preposition ("up"). Analyses of data in these regions were offset by 200 ms to account for the time required for planning and launching an eye movement (Hallett, 1978). Data in each region were analyzed for the Director and No-Director tasks separately. Because we did not have predictions regarding participants' eye gaze prior to the modifier (e.g., "large"), we focused our analyses on data in Regions 3 to 5 only.

First, 2 (Condition) × 2 (Trial Type) × 3 (Age Group) mixed ANOVAs were performed on correct trials for each region. Data from 2 participants in the child group were excluded from this analysis because they did not have enough correct trials. Second, 2 (Accuracy) × 3 (Age Group) mixed ANOVAs were performed separately for correct and incorrect Director Experimental trials, the key trials of interest, for each region. Because children and adolescents did not have many correct trials (in some cases only one or two), and some adults did not have any incorrect trials, some participants (5 children, 5 adolescents, and 9 adults) needed to be omitted from this analysis due to the lack of eye gaze data. Given the clear differences in accuracy between groups, we believe that it was not suitable to combine correct and incorrect trials in the analysis. However, interested readers can find the results of such analyses in the online supplementary material.

Results

Behavioral results

Director task
Accuracy. Fig. 3 shows participants' percentage error as a function of age group, and Table 1 shows the results of the statistical analysis. There were significant main effects of condition ($\eta_p^2 = .591$), trial type ($\eta_p^2 = .675$), and age group ($\eta_p^2 = .220$), which were qualified by significant Age Group × Trial Type and Age Group × Condition two-way interactions and a significant three-way interaction

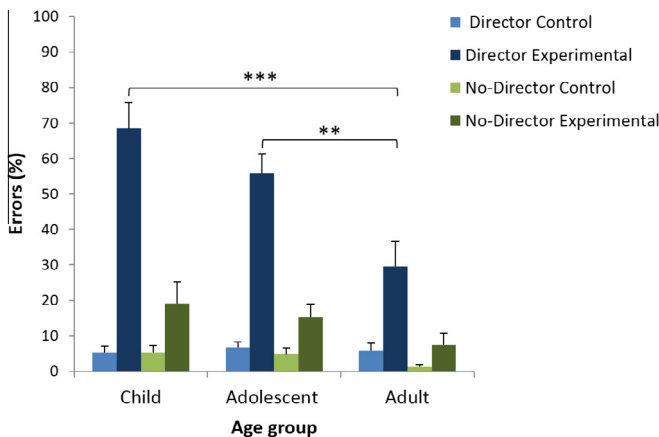**Fig. 3.** Percentage errors (means + standard errors) in the Director task for each age group. $^{**}p < .01$; $^{***}p \leqslant .001$.

Table 1
Omnibus repeated measures ANOVA of percentage errors in Director task.

	df	F
Omnibus ANOVA		
Age group	2, 57	8.06[**]
Condition	1, 57	82.27[**]
Trial type	1, 57	118.49[**]
Age Group × Condition	2, 57	2.27
Age Group × Trial Type	2, 57	6.65[**]
Condition × Trial Type	1, 57	70.73[**]
Age Group × Condition × Trial Type	2, 57	4.73[*]

[*] $p < .05$.
[**] $p < .01$.

($p = .013$, $\eta_p^2 = .142$). We explored this interaction further by analyzing the Experimental and Control trials separately. No significant effects were found in the Control trials (all $ps > .08$). In contrast, in Experimental trials there was a main effect of condition, $F(1, 57) = 83.57$, $p < .001$, $\eta_p^2 = .595$, a main effect of age group, $F(2, 57) = 8.10$, $p = .001$, $\eta_p^2 = .221$, and a significant interaction between condition and age group, $F(2, 57) = 3.67$, $p = .032$, $\eta_p^2 = .114$. Follow-up analysis examining Experimental trials in the Director and No-Director tasks separately revealed a significant effect of age group in Director Experimental trials, $F(2, 57) = 8.10$, $p = .001$, $\eta_p^2 = .221$, but not in No-Director Experimental trials, $F(2, 57) = 1.85$, $p = .167$. Follow-up t-tests for the Director Experimental trials showed that both the child and adolescent groups made significantly more errors than the adult group (child: $t(32) = 3.73$, $p = .001$, $d = 1.321$; adolescent: $t(44) = 2.94$, $p = .005$, $d = 0.887$), but there was no difference between the child and adolescent groups, $t(38) = 1.36$, $p = .183$.

Because the age groups significantly differed in terms of their verbal IQ, we repeated the $2 \times 2 \times 3$ mixed repeated measures ANOVA with standardized IQ (z-score) as a covariate. The same main effects and interactions were found.

To summarize, children and adolescents made more errors than adults only in the Director Experimental trials, which required taking the perspective of the director into account, in contrast to the rule-based control condition (No-Director Experimental trials). This replicated the behavioral results reported in Dumontheil and colleagues (2010).

Reaction time. Analysis of RT showed that children responded more slowly than the adolescents and adults in all trial types except for the Director Experimental trials, which showed no difference among age groups (see supplementary material for analyses). These results are also in line with those reported in Dumontheil and colleagues (2010).

Inhibitory control tasks
Simple Go–NoGo. The repeated measures ANOVA revealed significant main effects of trial type, $F(1, 57) = 50.34$, $p < .001$, $\eta_p^2 = .469$, and age group, $F(2, 57) = 5.29$, $p = .008$, $\eta_p^2 = .156$ (Fig. 4). More errors were made in NoGo trials than in Go trials and the child group made more errors than the adolescents ($p = .027$) and adults ($ps < .03$), who did not differ ($p = 1.00$). The interaction between age group and trial type was not significant ($p > .20$).

Complex Go–NoGo. We found significant main effects of trial type, $F(1, 57) = 108.87$, $p < .001$, $\eta_p^2 = .656$, and age group, $F(2, 57) = 11.44$, $p < .001$, $\eta_p^2 = .286$, in the Complex Go–NoGo task and no interaction between these factors ($p > .10$) (Fig. 4). Here, the child and adolescent groups did not differ ($p > .10$), but they both made more errors than the adult group ($ps < .01$).

Simple and Complex Go RT. Analyses of median RT in correct Go trials revealed that, in both the Simple and Complex Go–NoGo tasks, the child group was significantly slower than the adolescent and adult groups, who did not differ from each other (see supplementary material for RT analyses).
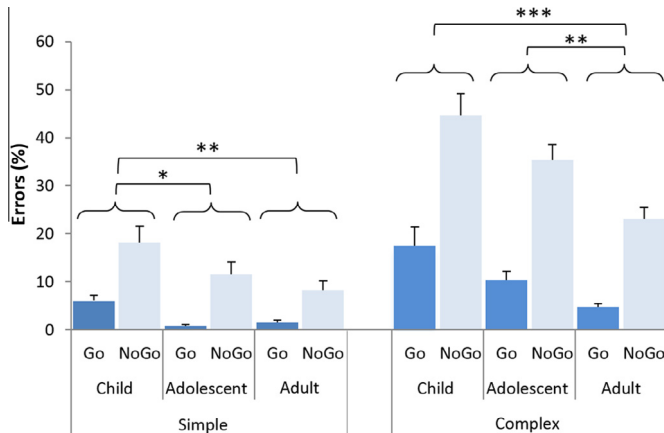
Fig. 4. Percentage errors (means + standard errors) in Go and NoGo trials in the Simple and Complex inhibitory control tasks for each age group. $^{*}p < .05$; $^{**}p < .01$; $^{***}p \leqslant .001$.

Association between inhibitory control tasks and Director task. Regression analyses showed that age significantly accounted for 14.4% of variance in the difference in percentage error between Director and No-Director Experimental trials (Table 2). Inhibitory control, as measured by the Simple NoGo percentage error, accounted for an additional 9.2% of variance, with more errors on Simple NoGo trials predicting more errors in Director versus No-Director Experimental trials. The other three measures of Go–NoGo percentage error were not significant predictors (all βs < .124 and ps > .370). The effect of age was still significant in the second model, suggesting partially independent effects of age and inhibitory control. Finally, entering verbal IQ as an additional regressor did not improve the model, and both the independent effects of age and Simple NoGo percentage error remained significant (Table 2).

Eye-tracking results

Table 3 shows mean target advantage scores across conditions, and Table 4 shows the results of the statistical analyses. Fig. 5A–D show plots of average target advantage over time. No significant effects were found in Region 3 ("large") (all ps > .20). Main effects of condition and trial type were marginal in Region 4 ("ball") (ps < .10) and were significant in Region 5 ("up") (condition: $\eta_p^2 = .042$; trial type: $\eta_p^2 = .234$). The Condition × Trial Type interaction was significant in both Region 4 ($\eta_p^2 = .031$) and

Table 2
Multiple regression analyses of Director task performance.

		β	t	p
Model 1 $R^2 = 14.4\%$; $F(1,57) = 9.60$, $p = .003$	(Constant)		5.421	.000
	Age	−.380	−3.099	.003
Model 2 $R^2 = 23.6\%$; $F(2,56) = 8.64$, $p = .001$	(Constant)		5.071	.000
$\Delta R^2 = 9.2\%$; $F(1,56) = 6.72$, $p = .012$	Age	−.298	−2.464	.017
	Simple NoGo	.314	2.591	.012
Model 3 $R^2 = 25\%$; $F(3,55) = 6.11$, $p = .001$	(Constant)		3.358	.001
$\Delta R^2 = 1.5\%$; $F(1,55) = 1.09$, $p = .302$	Age	−.287	−2.360	.022
	Simple NoGo	.311	2.575	.013
	Verbal IQ	−.122	−1.043	.302

Note. The difference in percentage error between Director and No-Director Experimental trials was entered as the dependent variable.

Table 3
Mean target advantage scores (and standard errors) for correct trials in each condition for each region.

Condition and trial type	Region 3 "large"	Region 4 "ball"	Region 5 "up"
Child			
Director Control	−0.08 (.02)	−0.12 (.02)	−0.02 (.04)
Director Experimental	0.00 (.03)	−0.04 (.03)	−0.07 (.07)
No-Director Control	−0.04 (.02)	−0.08 (.02)	−0.05 (.03)
No-Director Experimental	−0.03 (.04)	−0.15 (.04)	−0.28 (.06)
Adolescent			
Director Control	−0.03 (.02)	−0.09 (.02)	0.00 (.02)
Director Experimental	−0.03 (.02)	−0.07 (.04)	−0.11 (.05)
No-Director Control	−0.03 (.01)	−0.10 (.02)	0.01 (.02)
No-Director Experimental	−0.05 (.02)	−0.16 (.02)	−0.24 (.03)
Adult			
Director Control	−0.05 (.02)	−0.09 (.02)	0.03 (.02)
Director Experimental	−0.01 (.03)	−0.12 (.03)	−0.15 (.05)
No-Director Control	−0.02 (.02)	−0.07 (.02)	0.01 (.03)
No-Director Experimental	−0.01 (.03)	−0.13 (.04)	−0.24 (.05)

Table 4
Omnibus repeated measures ANOVA of target advantage for correct trials in each region.

	df	Region 3 "large"	Region 4 "ball"	Region 5 "up"
Omnibus ANOVA				
Age group	2, 55	0.44	0.13	0.20
Condition	1, 55	0.00	3.73^	8.52**
Trial type	1, 55	0.77	3.29^	81.58**
Age Group × Condition	2, 55	0.74	1.51	0.64
Age Group × Trial Type	2, 55	1.14	1.07	1.05
Condition × Trial Type	1, 55	1.43	11.45**	8.12**
Age Group × Condition × Trial Type	2, 55	0.24	1.96	0.42

^ $p < .1$.
** $p < .01$.

Region 5 ($\eta_p^2 = .035$). Follow-up analyses revealed no significant effects in Control trials in either region ($ps > .30$), whereas the target advantage in Experimental trials was significantly smaller in the No-Director task than in the Director task in both Region 4, $F(1,57) = 8.22$, $p < .01$, $\eta_p^2 = .055$, and Region 5, $F(1,57) = 10.84$, $p = .002$, $\eta_p^2 = .090$. No significant effects involving age group were observed in any of the regions (all $ps > .15$).

To summarize, in correct trials, participants showed a smaller target advantage in No-Director Experimental trials than in Director Experimental trials on hearing the noun. Crucially, their eye movement patterns did not vary across age groups in any of the regions.

We conducted additional analyses to examine the differences in eye gaze between correct and incorrect Director Experimental trials, the key trials of interest in the Director paradigm (Tables 5 and 6). We included an additional region (pre-response region) to check whether participants looked at the object they chose. This region was defined as the 200 ms before response time in a given trial. Graphs of target advantage again indicate a similar pattern across age groups (Figs. 5B and 6). Mixed 2 (Accuracy) × 3 (Age Group) ANOVAs revealed a main effect of accuracy across Regions 3 to 5 and the pre-response region (Region 3: $\eta_p^2 = .240$; Region 4: $\eta_p^2 = .271$; Region 5: $\eta_p^2 = .224$; pre-response: $\eta_p^2 = .118$). Participants had a bigger target advantage in incorrect trials than in correct trials in Regions 3 to 5; conversely, they had a bigger target advantage in correct trials than in incorrect trials in the pre-response region. No significant effects involving age group were observed (all $ps > .30$).

To summarize, analyses of eye movement data in the Director Experimental trials revealed clear differences between correct and incorrect trials, but there was no significant effects involving age group in any of the regions. In other words, there were no significant age-related differences in par-
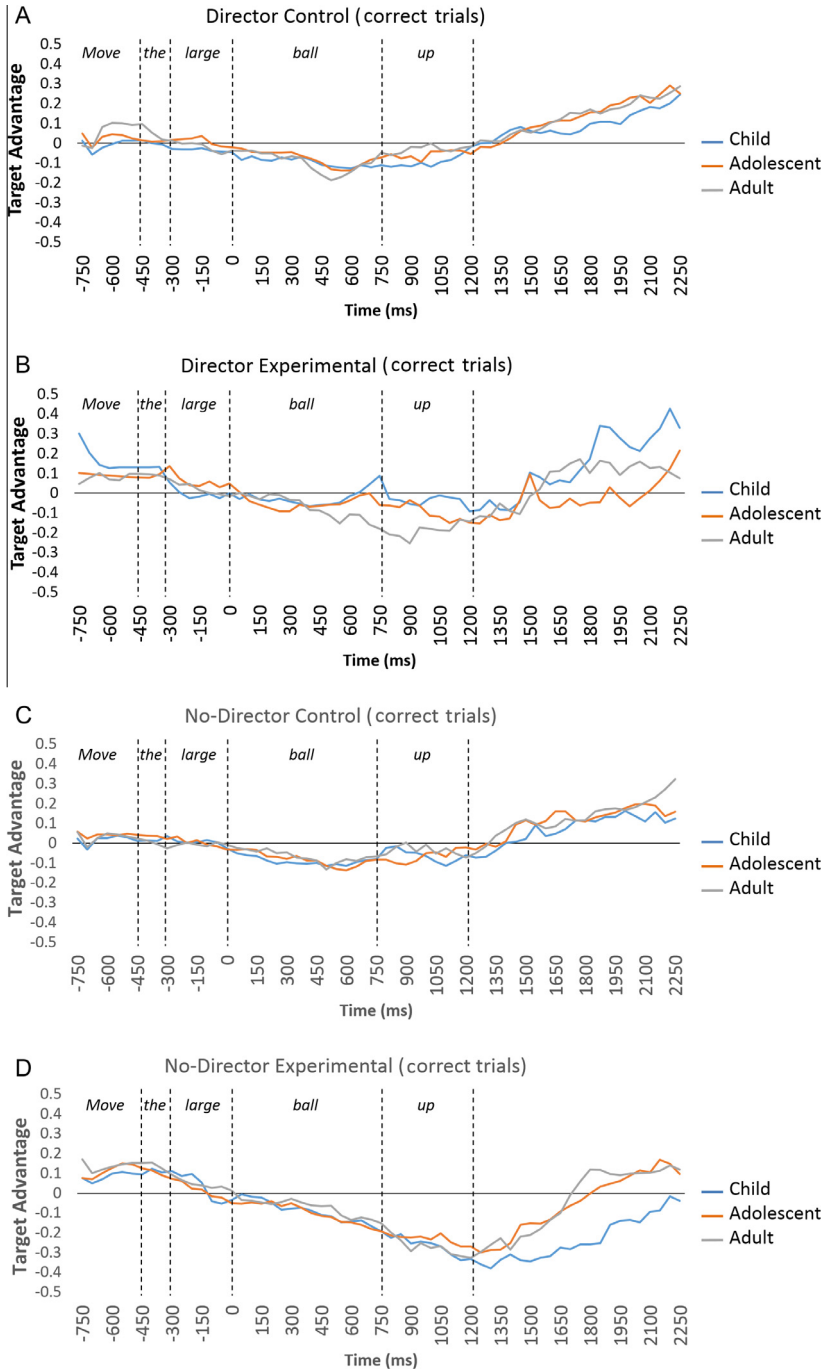
Fig. 5. Target advantage scores (looks to the target minus looks to the distractor [or to the irrelevant object for control trials]) per condition over 50-ms time bins: (A) Director Control; (B) Director Experimental; (C) No-Director Control; (D) No-Director Experimental.

Table 5
Mean target advantage scores (and standard errors) for correct and incorrect Director Experimental trials for each region.

Accuracy	Region 3 "large"	Region 4 "ball"	Region 5 "up"	Pre-response region
Child				
Correct	0.00 (.04)	−0.03 (.03)	−0.01 (.06)	0.32 (.15)
Incorrect	0.10 (.04)	0.11 (.03)	0.13 (.09)	−0.16 (.12)
Adolescent				
Correct	−0.03 (.02)	−0.09 (.04)	−0.14 (.05)	0.11 (.09)
Incorrect	0.17 (.05)	0.17 (.05)	0.19 (.06)	−0.12 (.11)
Adult				
Correct	0.00 (.04)	−0.09 (.03)	−0.07 (.08)	0.29 (.11)
Incorrect	0.27 (.06)	0.08 (.05)	0.18 (.11)	−0.01 (.12)

Table 6
Omnibus repeated measures ANOVA F-values of target advantage for correct and incorrect Director Experimental trials in each region.

	df	Region 3	Region 4	Region 5	Pre-response
Omnibus ANOVA					
Age group	2, 38	1.74	0.54	0.12	0.80
Accuracy	1, 38	23.68**	30.23**	22.06**	12.53**
Age Group × Accuracy	2, 38	1.03	1.02	0.87	0.70

** $p < .01$.
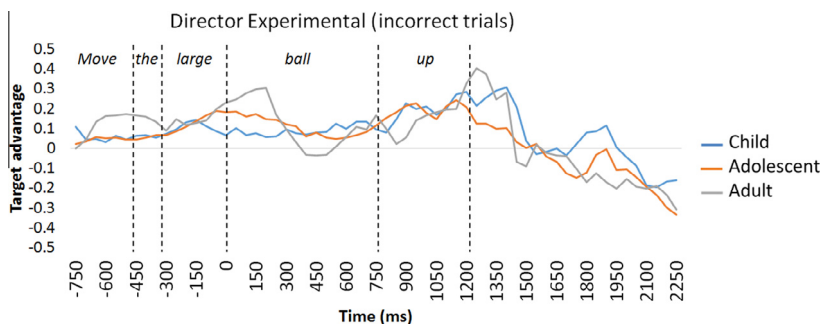
Fig. 6. Target advantage scores per condition over 50-ms time bins for incorrect Director Experimental trials.

ticipants' eye gaze patterns in both correct and incorrect Director Experimental trials. Furthermore, the effect of accuracy observed in Regions 3 to 5 was reversed in the pre-response region, such that participants initially showed a greater target advantage in the incorrect trials than in the correct trials, and it was only right before they responded that they showed a greater target advantage in correct trials than in incorrect trials.

Discussion

In this study, we collected behavioral and eye-tracking data to investigate online use of ToM during perspective taking in children, adolescents, and adults. Experimental trials of the Director task required participants to take into account the director's perspective to determine the intended referent in the instructions (e.g., "large ball"). The No-Director condition required participants to follow an explicit avoidance rule to determine the correct referent. Whereas both the Director and No-Director

conditions involved executive function, in particular inhibition, only the Director condition required an inference about the speaker's intentions given the speaker's ignorance of certain objects (Dumontheil et al., 2010).

The current study had three main objectives: (a) to replicate the behavioral findings of Dumontheil and colleagues (2010), (b) to assess the role of inhibitory control in this task across age groups, and (c) to compare the time course of information processing among children, adolescents, and adults through eye-tracking in order to determine in what ways the deployment of ToM differs across age groups.

Behavioral data

We found that in the Director Experimental condition, the child and adolescent groups performed worse than the adult group, but participants' accuracy did not differ across groups in the No-Director Experimental condition. These results are in line with Dumontheil and colleagues (2010) findings and provide further evidence for age-related differences between adolescents' and adults' tendency to take someone else's perspective into account. Conversely, response times decreased with age in all conditions except in Director Experimental trials, which did not vary with age, and are again similar to those observed by Dumontheil and colleagues.

Inhibitory control

We observed age-related differences in participants' performance on both inhibitory control tasks. In the Simple Go–NoGo task, the child group made significantly more errors and was significantly slower than the adolescent and adult groups. These results are in line with previous studies suggesting that with a low-level cognitive load inhibitory control performance reaches adult level performance by approximately 14 years of age (Lamm, Zelazo, & Lewis, 2006; Leon-Carrion et al., 2004; Luna, Garver, Urban, Lazar, & Sweeney, 2004). In the Complex Go–NoGo task, the child and adolescent groups performed worse than the adult group. Because the Complex Go–NoGo task included a 1-back WM requirement, these results are also in line with studies investigating WM that found age-related differences between adolescents and young adults (Conklin, Luciana, Hooper, & Yarger, 2007; Luna et al., 2004).

Association between inhibitory control tasks and Director task

What was novel in this study was the investigation of the relationship between the inhibitory control tasks and the Director task. The results show that inhibitory control as measured by Simple NoGo accuracy accounted for some of the variance in accuracy difference between Director and No-Director Experimental trials. Critically, inhibitory control accounted for only some of the variance due to age given that age remained a significant predictor, suggesting that some additional factors are behind age-related changes. These results are in line with a previous study by Vetter and colleagues (2013), who found that 15% of the variance of affective ToM performance was uniquely explained by age, indicating independent effects of age and inhibition.

Interestingly, no relationship was found between participants' performance on the Complex Go–NoGo task and the Director task. This is surprising because performance on the Complex Go–NoGo task showed age-related differences between the adolescent and adult groups. This could be because the Complex Go–NoGo task placed greater WM demands than the Director task. A study that explored the relationship between WM and perspective taking (Lin et al., 2010) found that participants with lower WM capacity performed more poorly on the Director task than participants with greater WM capacity and that participants' performance on the Director task was worse during high-WM load trials than during low-WM load trials. However, Lin and colleagues' (2010) study did not include a No-Director condition with matched general executive function demands, whereas the current study used the difference in accuracy between Director Experimental and No-Director Experimental trials as the measure of interest. Future research might shed light on these differences in results by using separate measures of WM and inhibition (Vetter et al., 2013).

A possible limitation for the interpretation of the behavioral results is the difference in verbal IQ among age groups. However, the significant interactions with age observed in the Director task were

still present when verbal IQ was added as a covariate. Moreover, adding verbal IQ as a predictor in the multiple regression did not affect the results given that both the independent effects of age and Simple NoGo accuracy remained significant.

Eye-tracking data

Through eye-tracking, we were able to investigate additional underlying factors in Director task performance. Examining correct and incorrect trials separately indicated that adults, adolescents, and children did not differ in their online processing of the task (Fig. 5). Analyses of Director Experimental trials data showed opposite effects of accuracy in the earlier regions (Regions 3–5) and the pre-response region, such that participants initially showed a greater target advantage in the incorrect trials than in the correct trials. They showed a greater target advantage in correct trails than in incorrect trials only right before they responded. These results seem inconsistent with the perspective adjustment model (Keysar et al., 2000, 2003), according to which one would expect that on both the correct and incorrect trials participants first attend to the distractor, the best-fit referent from an egocentric perspective. On correct trials participants would adjust to considering the director's perspective and attend to the target, whereas on incorrect trials the second adjustment process would fail because it is costly and participants would remain focused on the distractor. However, as is evident from the incorrect trials for all age groups, participants appear to consider the target early on in the trial and then, before responding, their eye gaze shifts to the distractor.

Based on the eye-tracking results, it seems that at the point where the director indicates which object should be moved (Region 4), participants on correct trials are already on a path to correctly considering the director's perspective. However, participants' strategy seems to be to first look at the objects that participants are not going to choose (or the objects they should not pick) as a process of elimination before focusing on the object that they will ultimately choose. This pattern has not been reported before. Other studies, such as Hanna and colleagues (2003), Heller and colleagues (2008), and Barr (2008), showed that bias in participants' eye gaze builds steadily toward the target after initial interference from the distractor. What sets our study apart from these eye-tracking studies is the fact that the distractor is the best fit for the description, making the task of ignoring the privileged object particularly challenging. In addition, the description itself contains a relational modifier (e.g., "big," "bottom") that implicitly refers to a contrast set. Given these two factors, it should not be surprising that participants adopt a strategy of checking all objects of the same type (e.g., all balls on the display) to ensure that they choose the correct one. Our eye gaze data of Object 2 (e.g., the second commonly viewable ball) suggests this also. It shows that prior to focusing on the object they chose, participants paid similar attention to each of the other two objects they eliminated (see Figs. S3–S5 in supplementary material). The only other studies that used a setup similar to ours are those reported in Keysar and colleagues (2000) and Wu and Keysar (2007). Neither of those articles reported eye gaze data in full, but their results are consistent with ours in that they found more looks to the distractor overall and first looks to the distractor were earlier than first looks to the target.

Adults, adolescents, and children in the current study seemed to follow a process of elimination strategy in control trials as well, where only the two commonly viewable objects denoted by the noun (e.g., "ball") were present and only one fit the full description ("large ball"). We propose that participants focus first on the second object to exclude it before looking at the target (see Figs. S3–S6 in supplementary material). Therefore, there may be a consistent strategy across all conditions, and all age groups, of looking at the objects that are not going to be chosen prior to focusing on the object to be chosen.

If we accept that participants adopt a process of elimination strategy, then our results suggest that the eye gaze pattern on correct trials is influenced by the actual beliefs of the director early on, as we see the pattern emerge in Region 4 when participants process the modifying adjective ("large"). The idea that eye gaze data reflect an influence of the speaker's perspective at an early stage is in line with a number of previous perspective-taking studies mentioned above (Brown-Schmidt et al., 2008; Hanna et al., 2003; Heller et al., 2008). The results are also consistent with Nadig and Sedivy's (2002) observation that 5- and 6-year-old children showed early sensitivity to a speaker's perspective in a much simplified director task. These articles proposed an alternative to the perspective adjust-

ment model, claiming that mental state information, like any other relevant information, is potentially available to be integrated into referential decision processes from the outset. On this constraint-based view, the extent to which mental perspective information is used depends on the extent to which other constraints are conflicting and how salient or available the perspective information is (Brown-Schmidt & Hanna, 2011; Samson et al., 2010). Hanna and colleagues (2003) argued that the Director paradigm used in Keysar and colleagues (2000) and in this article makes conflicting cues related to the linguistic form particularly strong because the occluded object (the distractor) is in fact the best fit for the description. Other studies such as Brown-Schmidt (2009b) have suggested that varying the strength (or quality) of cues to the speaker's mental perspective can affect online referential processes. If varying cues to mental perspective can have an impact on ToM integration, it seems plausible from this constraint-based perspective that individuals may differ in the extent to which their referential processes exploit a given cue. Thus, an explanation of the age-related differences that are not accounted for by inhibitory control may lie in differences in participants' sensitivity of online processes to mental perspective information. This is a hypothesis that requires further exploration.

If we assume that accuracy differences in our Director task are a product of varying abilities to integrate perspective information in incremental referential processes, rather than at a later corrective stage, then we can make sense of the reaction time results reported here and in Dumontheil and colleagues (2010). These results showed shorter RTs in the Director task compared with the No-Director task and also no RT difference in the Director condition between age groups. The observation made about these results is that the Director task engages a more efficient or rapid process than simple explicit rule following and does so to the same extent across age groups.

Taken together, the current results provide evidence for age-related differences between adolescents and adults in their online use of ToM. Contrary to perspective adjustment accounts of the Director task, our results suggest that all age groups appear to engage in the same kind of online processes during perspective taking but differ in how often mental state information informs incremental decision processes. Taking our results in the wider context of research into online use of ToM, we see one possible source of the difference between age groups as being their sensitivity to available cues to mental state information.

Acknowledgment

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jecp.2015.11.007.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73, 247–264.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. Cognition, 106, 1093–1108.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. Cognition, 109, 18–40.

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. Journal of Memory and Language, 61, 171–190.

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. Psychonomic Bulletin & Review, 16, 893–900.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. Language and Cognitive Processes, 27, 62–89.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. Cognition, 107, 1122–1134.

Brown-Schmidt, S., & Hanna, J. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. Dialogue & Discourse, 2(1), 11–33. http://dx.doi.org/10.5087/dad.2011.102.

Burnett, S., Sebastian, C., Cohen Kadosh, K., & Blakemore, S.-J. (2011). The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies. *Neuroscience and Biobehavioural Reviews, 35*, 1654–1664.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*, 299–319.

Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental Neuropsychology, 31*, 103–128.

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion, 8*, 725–730.

Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development, 84*, 989–1003.

Devine, R. T., & Hughes, C. (2014). Relations between false-belief understanding and executive function in early childhood: A meta-analysis. *Child Development, 85*, 1777–1794.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science, 13*, 331–338.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*, 760–768.

Fett, A. K. J., Shergill, S. S., Gromann, P. M., Dumontheil, I., Blakemore, S. J., Yakub, F., et al (2014). Trust and social reciprocity in adolescence: A matter of perspective-taking. *Journal of Adolescence, 37*, 175–184.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18*, 1279–1296.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43–61.

Hansen Lagattuta, K., Sayfan, L., & Harvey, C. (2014). Beliefs about thought probability: Evidence for persistent errors in mindreading and links to executive control. *Child Development, 85*, 659–674.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition, 108*, 831–836.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32–38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language, 83*, 1–19.

Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology, 46*, 1417–1432.

Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia, 44*, 2139–2148.

Lecce, S., Bianco, F., Devine, R. T., Hughes, C., & Banerjee, R. (2014). Promoting theory of mind during middle childhood: A training program. *Journal of Experimental Child Psychology, 126*, 52–67.

Leon-Carrion, J., García-Orza, J., & Pérez-Santamaría, F. J. (2004). Development of the inhibitory component of the executive functions in children and adolescents. *International Journal of Neuroscience, 114*, 1291–1311.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46*, 551–556.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development, 75*, 1357–1372.

Luna, B., Padmanabhan, A., & Hearn, K. O. (2011). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition, 72*, 101–113.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329–336.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58*, 220–249.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1255–1266.

Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of Go/No-Go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia, 46*, 224–232.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13- month-old infants. *Psychological Science, 18*, 580–586.

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development, 83*, 452–460.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Vetter, N. C., Altgassen, M., Phillips, L., Mahy, C. E. V., & Kliegel, M. (2013). Development of affective theory of mind across adolescence: Disentangling the role of executive functions. *Developmental Neuropsychology, 38*, 114–125.

Wang, J. J., Ali, M., Frisson, S., & Apperly, I. A. (in press). Language complexity modulates 8- and 10-year-olds' success at using their theory of mind abilities in a communication task. *Journal of Experimental Child Psychology*. http://dx.doi.org/10.1016/j.jecp.2015.09.006.

Watanabe, J., Sugiura, M., Sato, K., Sato, Y., Maeda, Y., Matsue, Y., ... Kawashima, R. (2002). The human prefrontal and parietal association cortices are involved in NO-GO performances: An event-related fMRI study. *NeuroImage, 17*, 1207–1216.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.

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

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18*, 600–606.