

What hinders child semantic computation: children's universal quantification and the development of cognitive control*

UTAKO MINAI

University of Kansas, Department of Linguistics and RIKEN Brain Science Institute, Laboratory for Language Development

NOBUYUKI JINCHO AND NAOTO YAMANE

RIKEN Brain Science Institute, Laboratory for Language Development

AND

REIKO MAZUKA

RIKEN Brain Science Institute, Laboratory for Language Development and Duke University, Department of Psychology & Neuroscience

*(Received 4 March 2009 – Revised 30 August 2010 – Accepted 16 August 2011 –
First published online 19 December 2011)*

ABSTRACT

Recent studies on the acquisition of semantics have argued that knowledge of the universal quantifier is adult-like throughout development. However, there ARE domains where children still exhibit non-adult-like universal quantification, and arguments for the early mastery of relevant semantic knowledge do not explain what causes such non-adult-like interpretations. The present study investigates Japanese four- and five-year-old children's atypical universal quantification in light of the development of cognitive control. We hypothesized that children's still-developing cognitive control contributes to their atypical universal quantification. Using a combined eye-tracking and interpretation task together with a non-linguistic measure of cognitive control, we revealed a link between the achievement of adult-like universal quantification and the development of flexible perspective-switch. We argue that the development of cognitive control is one of the factors that contribute to children's processing of semantics.

[*] We gratefully acknowledge Koji Sugisaki, Miwa Isobe, Takuya Goro, Edson Miyamoto and Robert Fiorentino for their fruitful feedback regarding the current study, and Kiwako Ito and Akitoshi Ogawa for their advice and comments on the design and analysis of the experiment. Address for correspondence: Utako Minai, Department of Linguistics, University of Kansas, Blake Hall Rm. 427, 1541 Lilac Lane, Lawrence, KS 66045, USA. e-mail: minai@ku.edu

INTRODUCTION

Atypical universal quantification by preschool children

Children's computation of the meanings of sentences involving the universal quantifier (e.g. *every* in English) has been a subject of debate. Initially, children were claimed to exhibit atypical universal quantification (Inhelder & Piaget, 1964). Following this original observation, later studies specifically showed that if children are asked 'Is every boy riding an elephant?' along with a picture showing some boys each riding an elephant and an EXTRA elephant nobody is riding (this is typically called the EXTRA OBJECT CONDITION), three-, four- and five-year-old children would answer 'No', pointing to the extra elephant as the reason, even though its presence does not falsify the premise *every boy is riding an elephant* (e.g. Philip, 1995). This atypical semantic interpretation involving the universal quantifier is called the SYMMETRICAL RESPONSE (henceforth SR); it seems that children reject these sentences by reasoning that the falsifier is the presence of the extra object which ruins the symmetrical one-to-one relation between boys and elephants in the picture. As opposed to SR discussed as children's atypical response, we will call the typical adult-like interpretation the 'logical reading' (henceforth LG, i.e. a 'True' response to pictures in which, for example, every boy is riding an elephant irrespective of the presence of extra elephants).

Some linguists have attempted to provide linguistic theory-based explanations for SRs. These approaches attribute the origin of children's SRs to their non-adult-like semantic representation of *every* (e.g. Drozd, 2001; Geurts, 2003; Philip, 1995). Other researchers have provided evidence suggesting that throughout their development, children's semantic knowledge and representation of the universal quantifier is not different from adults, but that SRs reflect the effect of some extralinguistic factor which blocks children's USE of their adult-like semantic knowledge (e.g. Crain, Thornton, Boster, Conway, Lillo-Martin & Woodams, 1996; Gualmini, 2004; Minai, 2006). In particular, Sugisaki and Isobe (2001) reported that modified extra-object pictures, in which multiple numbers of extra objects were depicted, elicited nearly perfect rates of LG responses from four- and five-year-old children. Gourou, Norita, Nakajima and Arijji (2001) also elicited higher rates of LG responses from children around the same age, by utilizing revised extra-object pictures in which different subtypes of agents within the same supertype are associated with the same objects in different colors (e.g. each of three different kinds of Pokémon is riding on a red pony, a blue pony and a yellow pony respectively, while there is a green pony nobody is riding, i.e. the extra pony). Taken together, these studies suggest that children interpret the universal quantifier the same way as adults under certain circumstances, and that

extralinguistic factors play a role in modulating children's level of success in universal quantification.

In search of the basis for symmetrical responses: language development and cognitive control

The fact that children do exhibit LG interpretation of the universal quantifier under certain circumstances shows that it is not a lack of semantic knowledge of universal quantification that causes them to show SRs in contexts such as the Extra Object Condition described above. This explanation *PER SE*, however, does not account for *WHY* there are domains in which children still exhibit SRs, nor does it offer an explanation of what allows them to grow out of it. The present article investigates what elicits the SR in child language from the perspective of cognitive development. Sugisaki and Isobe (2001) and Gouro *et al.* (2001) argued that the successful use of adult-like semantic knowledge by three-, four- and five-year-olds is easily hindered by an extralinguistic factor; they demonstrated that children's interpretation of the universal quantifier improved when the pictures were changed. Gouro *et al.* (2001) pointed out the visual *SALIENCE* of the extra object in the picture, defining it as an *EXCEPTIONALITY* in which the extra object uniquely lacks the property shared by the other objects, i.e. being paired with an agent, and argued that the salience of the extra object affects children's responses to sentences with the universal quantifier.

This raises the question of *WHY* salience matters for children. We propose that some cognitive factors may contribute significantly to children's SR for universal quantification. Specifically, we argue that a key factor is children's ability to flexibly switch between different perspectives, which is required in computing universal quantification in extra-object contexts, without being hindered by the undue focus on the salient but irrelevant information provided in the extra object.

Let us spell out the relevant issues regarding children's flexible perspective switch. Since Piaget's early observation (e.g. Piaget, 1954), children's cognitive skills have been assumed to be inflexible in various aspects, such as focusing on one dimension of stimuli (e.g. the height of water in a narrow glass) while unable to simultaneously consider another dimension (e.g. the shape of the glass). Recent research has examined such phenomena in light of their development of *COGNITIVE CONTROL*, the cognitive system that controls a set of skills required for a variety of cognitive processes such as planning, decision-making, abstract thinking, reasoning, rule acquisition and error correction or troubleshooting (e.g. Conboy, Sommerville & Kuhl, 2008; Davidson, Amso, Anderson & Diamond, 2006; among many others). We focus on one of the cognitive aspects characterized in terms of cognitive control, i.e. the ability to flexibly switch perspectives. This cognitive aspect

is commonly assessed in preschool children, whose age ranges from three to five, using cognitive tasks such as the DIMENSIONAL CHANGE CARD SORT (DCCS) (e.g. Frye, Zelazo & Palfai, 1995; Zelazo, 2006; Zelazo, Frye & Rapus, 1996; Zelazo, Müller, Frye & Marcovitch, 2003). Children are asked to sort cards which can be classed according to two competing dimensions (color and shape), first according to one dimension (e.g. shape) and then according to the other (e.g. color). The decrease of the rates of successful card sorting over the dimension-switch offers a measurement of how flexibly children switch perspectives. Previous research has demonstrated that three-, four- and five-year-olds can successfully sort the cards with respect to the first dimension; however, with respect to the second dimension, many three- and four-year-olds fail, whereas five-year-olds succeed approximately 75% of the time (Frye *et al.*, 1995). Some researchers attribute such DCCS performance patterns to children's difficulty in controlling their attention, which is called ATTENTIONAL INERTIA (e.g. Diamond, Carlson & Beck, 2005; Kirkham, Cruess & Diamond, 2003; cf. Frye *et al.*, 1995; Zelazo *et al.*, 2003; 1996; among others). According to this approach, children in this age range have a difficulty in perceiving an object from multiple perspectives, and thus are not able to flexibly switch perspectives. For example, if children perceive a white cat, they first focus on one aspect of it, e.g. color (its white-ness), but experience difficulty re-perceiving it while focusing on another aspect, e.g. shape (its cat-ness), failing to disengage from the first perspective (i.e. color).

The development of cognitive control has been discussed in light of its link with child language. These studies have been searching for the potential link among cognitive development and overall language development. Bialystok and her colleagues examined the relationship between child bilingualism and cognitive development (e.g. Bialystok, 1999; 2001). Whereas bilingual children typically show disadvantaged development in their verbal ability, compared with their monolingual peers (e.g. Macnamara, 1966; Myers & Goldstein, 1979), they outperform the monolinguals on a number of cognitive tasks including the DCCS (e.g. Bialystok, 1999; Bialystok & Martin, 2004).

Examining the relationship between language development and development of cognitive control from another perspective, Mazuka, Jincho and Oishi (2009) proposed that children's developing cognitive control may contribute to children's PARSING AND INTERPRETATION OF INDIVIDUAL SENTENCES. They note that research points out children's tendency to PERSEVERATE in sentence processing, in which once children were led to a wrong path in sentence processing, they exhibit difficulty recovering from it (e.g. Trueswell, Sekerina, Hill & Logrip, 1999). They claim that children's tendency to perseverate in interpreting individual sentences shares many characteristics with cognitive inflexibilities related to their still-developing

cognitive functions. They further imply that children's tendency to exhibit non-adult-like universal quantification might be linked to their still-developing cognitive control.

Universal quantification, visual salience and cognitive control

Recall that preschool children, who exhibit still-developing cognitive control, also exhibit the tendency to commit atypical universal quantification in extra-object contexts. This raises the question of whether there indeed is a link between these two domains. Our speculation that the extra object, as salient but irrelevant information, hinders children's successful universal quantification is consistent with the Attentional Inertia argument (e.g. Kirkham *et al.*, 2003). As shown by Sugisaki and Isobe (2001) and Gouro *et al.* (2001), children's judgments with universal quantification are non-adult-like when the extra object (which should be ignored) is salient in the visual scene. In order to evaluate the truth of the sentence with the universal quantifier, children must examine the contents of the picture as a truth condition. If their attention is captured by the extra object initially, they have to suppress the perspective from which they initially perceived the picture (i.e. the depiction 'about' a saliently remnant object) and re-evaluate it regarding whether or not all the agents have an object. Here the focus must shift from the presence of the extra object to the relation between the agents and the objects. Children who have difficulty in shifting perspectives in such a way may fail in suppressing their initial attention to the salient extra object, resulting in committing atypical universal quantification (i.e. the SR). On the other hand, children with more developed cognitive control would be able to suppress the extra object in interpreting the universal quantifier in the sentence. The present study aims at examining this hypothesis.

The support for our hypothesis provides converging evidence for recent claims regarding children's semantic computation. Our hypothesis, attributing children's non-adult-like universal quantification to their cognitive development, is consistent with Rakhlin's (2007) claim that children's atypical universal quantification may reflect their still-developing theory of mind (ToM). Assuming that one needs to appropriately restrict the domain over which quantification by *every* should range, in order to correctly understand sentences with *every*, Rakhlin argues that children's still-developing ToM poses a difficulty for incorporating others' perspectives into establishing an appropriate domain. Our hypothesis also converges with the Question-Answer Requirement (QAR) model, proposed by Gualmini, Husley, Hacquard and Fox (2008), which claims that sentences are interpreted as answers to a particular question. Since the question is determined by contextual cues, how children interpret sentences depends

on what cues children use to establish the questions they are answering. Salient information, such as an extra object in the visual scene, would attract children's attention, perhaps causing them to think that the question they should be answering is about the extra object.

Present study

Based on previous studies of both child universal quantification and the development of cognitive control, we selected four- and five-year-old children as our target age. Since they are at a transitional developmental stage in which their semantic interpretation of universal quantification can easily be hindered by extralinguistic factors, we can examine the link between the development of cognitive control and their interpretation of the universal quantifier. The experiment was conducted in Japan, and our subjects were Japanese-acquiring children.

The previous studies investigating children's interpretation of universal quantification utilized the TRUTH VALUE JUDGMENT (TVJ) Task (e.g. Crain & Thornton, 1998). In this task, children's semantic interpretation is measured by their response to questions regarding the contents of visually presented truth conditions. Note that children's online processing of the sentence is not assessed, because children's semantic judgments are reflected in the resultant output of their semantic computation. Thus, in order to investigate the online aspects of their computation of universal quantification, we monitored children's eye-movements during the TVJ task. Sentence comprehension studies utilizing the visual-world paradigm have demonstrated that, given a visual stimulus depicting the content of an auditorily presented sentence, the listener's eye-movements are closely linked with the online comprehension of the sentence (e.g. Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Trueswell & Gleitman, 2004). In our study, children were asked to judge the truth of universally quantified sentences based on the visual presentation of pictures containing extra objects, as in previous TVJ studies. Examining when and how often children look at various parts of the picture during the TVJ task could provide useful information about what children are paying attention to when they reach the SR, and how it may differ from those when they arrive at an LG interpretation.

In the present study, the same children also perform the DCCS, which, as we have discussed above, has been widely adopted to measure preschoolers' ability to switch perspectives between two competing dimensions that both serve as different standards for card sorting. Combining these two tasks allows us to examine whether there is a link between children's ability to interpret the universal quantifier and the flexibility in their switch of perspective over the dimensional change in card sorting.

Our hypothesis leads us to predict that there is a link between children's TVJ responses and DCCS performance; we expect children who show the SR in the TVJ task for the universal quantifier to exhibit poor performance in the DCCS, while children who do not show the SR to exhibit higher performance in the DCCS. We also predict that when children show the SR in the TVJ task, they look more frequently at the extra object, reflecting that the salience of the extra object drew more attention. In contrast, children would show fewer looks at the extra object in the picture when they were able to avoid the SR and show adult-like universal quantification.

METHOD

Participants

Seventy-four Japanese-acquiring four- and five-year-olds (4;00–5;11, Mean 5;01) participated in the experiment. We included data only from those who completed both TVJ and DCCS tasks and whose eye-movements were successfully recorded. Consequently, twenty-two additional children were tested but their data were not included in the analysis for the following reasons: (i) they could not complete both TVJ and DCCS tasks (6); (ii) they could not follow the instructions, e.g. they made responses before the sentence was completed in the TVJ task, or were unable to choose one of the boxes for sorting in the DCCS (5); (iii) they gave uninterpretable responses in TVJ, e.g. saying 'Right' or 'Wrong' to all items including the fillers (5); or (iv) there were technical problems with the eye-tracker, or coding errors occurred (6). As controls, we also analyzed the data collected from forty-eight adult native speakers of Japanese. An additional five adults participated, but their data were excluded from the analysis due to technical difficulties with the eye-tracker. Twenty of the adults also performed the DCCS as controls, though the task was clearly too easy for them. The study was approved by the RIKEN Ethics Review Board and Duke University Institutional Review Board. All the children participated in the experiment with their parental consent, and the adults participated in the experiment with their agreement based on informed consent. They were tested individually in a quiet experiment room.

Stimuli

TVJ task. The stimuli sentences were spoken naturally by a female native speaker of Japanese and digitally recorded. Seventeen stimulus sentences were prepared: eight target sentences, two warm-up sentences, and seven filler sentences. All sentences contain a universally quantified subject (including three negative sentences serving as fillers, which correspond to

English sentences containing *nobody*; all the stimulus sentences are listed in ‘Appendix 1’). A sample target sentence is given in (1).

- (1) *Dono-kame-mo kasa-o sashi-teruyo.*
 which-turtle-also umbrella-ACC hold-ing
 ‘Every turtle is holding an umbrella.’

The Japanese translation for *every*-NP in English is represented as a morphological combination of the WH-operator (traditionally called ‘indeterminate words’, e.g. Kuroda, 1965) *dono* (‘which’) and the particle *mo* (‘also’), with the NP in between. A recent theoretical issue is the association between *dono* and *mo*, regarding whether it is implemented via a movement (e.g. Nishigauchi, 1999; Takahashi, 2002; cf. Shimoyama, 2006). Crucially for the current article, the sentences with *dono*-NP-*mo*, as in (1), were used as the stimulus sentences in previous experiments on Japanese children’s universal quantification (Sugisaki & Isobe, 2001; Gouro *et al.*, 2001); as we focus on children’s logical interpretation of universal quantification, following these studies, the theoretical discussion on the representation of *dono*-NP-*mo* is beyond our scope.

Pictures whose depictions correspond to each of the stimulus sentences were generated. For each of the eight target items, two types of picture sets were created. The first type depicted three animals holding one object apiece and three remaining objects that no animals were holding (i.e. the extra objects); we call this set the ‘Multiple Object’ pictures (see Figure 1). The second set of pictures depicted only one extra object, but was otherwise identical to the Multiple Object pictures; we call this set the ‘Single Object’ pictures (see Figure 2). Thus, a total of twenty-five pictures were created, i.e. two corresponding pictures each (the Multiple Object version and the Single Object version) for the eight targets and one corresponding picture each for the seven fillers and two warm-ups. The displays were divided into two-by-two rectangular lattices (but no visible border lines). Three lattices contained the pairing of an agent and an object respectively, and the remaining one contained the extra object(s). The area of interest (AOI) for gaze-tracking was assigned to the grid containing the extra object(s). The position of the AOI lattice was counterbalanced across items.

The expected responses in the TVJ for each item were determined by match or mismatch between the sentence meaning and the depiction of the corresponding picture: matched sentence–picture pairs to elicit ‘True’ responses and non-matched sentence–picture pairs to elicit ‘False’ responses. For all the target items, the adult-like (LG) interpretation of the universal quantifier would yield a ‘True’ response, as the sentence and picture matched. Half of the filler items provided mismatched pictures in order to elicit ‘False’ responses, while the other half of the fillers provided matched pictures.

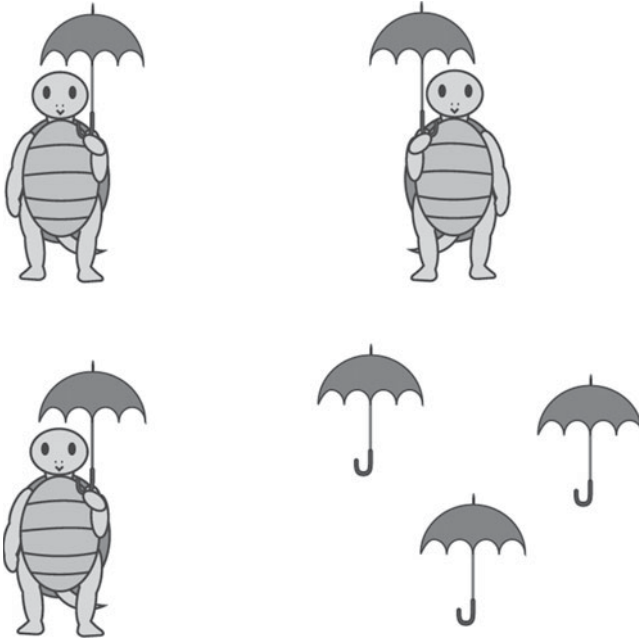


Fig. 1. Sample 'Multiple Object' picture for the sentence *Dono-kame-mo kasa-o sashi-teruyo* ('Every turtle is holding an umbrella').

In Sugisaki and Isobe (2001), the number of extra objects was much larger, six or seven, resulting in almost perfect LG performance by four- and five-year-old Japanese children. In the present study, we presented three extra objects, aiming to elicit both SRs and LG responses from within the same group of children. As a control, we also tested another group of children with pictures that contained a single extra object, in order to (i) confirm that the children we were testing were at a stage in which they would provide an overwhelming proportion of SRs if they were given a single extra object, and (ii) compare children's responses and eye-movement patterns across two groups of children.

DCCS. *DCCS* was administered on a computer screen, instead of using actual cards with pictures. Two sets of materials were prepared in terms of the category of the items depicted in the cards, i.e. animal and vehicle. The 'animal' set features two types of animal illustrations, a cat and a monkey, serving as the sorting standard in terms of shape. Each animal picture is painted in one of two colors, green or black, serving as the sorting standard in terms of color. The model pictures on the sorting boxes depicted a green cat and a black monkey respectively, and the sorting cards contained either a green monkey or a black cat each. The 'vehicle' set features a bus and an

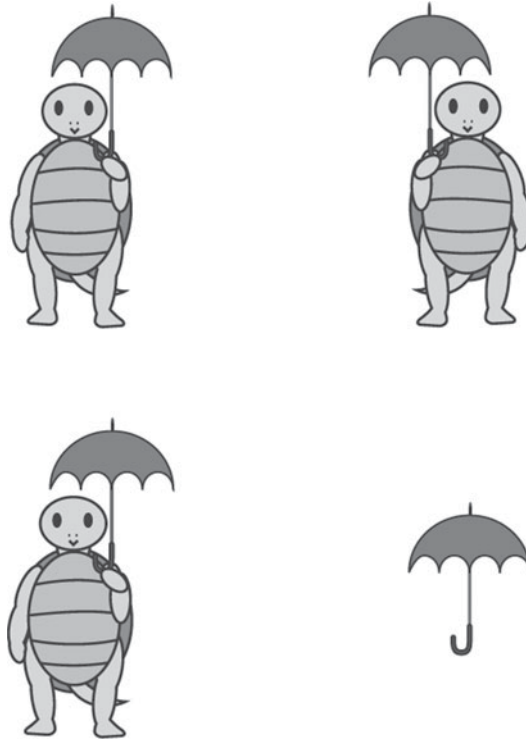


Fig. 2. Sample 'Single Object' picture for the sentence *Dono-kame-mo kasa-o sashi-teruyo* ('Every turtle is holding an umbrella').

airplane. Each vehicle picture is painted in either blue or white. Two model pictures depicted a white bus and a blue airplane respectively, and the sorting cards each contained a white airplane or a blue bus. The two sets were created so each child could participate in the card sorting session twice, first in one order of rule-switch (e.g. color to shape) using one of the sets of materials, and then in the reverse (e.g. shape to color) using the other set of materials. This also allows each subject to participate in the task twice to counterbalance the order of the rule presentation across subjects.

Design

For the TVJ task, Single Object pictures and Multiple Object pictures were presented in two blocks. One group of participants (TARGET GROUP) was given Multiple Object pictures in the first block (Block 1) and Single Object pictures in the second block (Block 2). Another group (CONTROL GROUP) was given Single Object pictures in Block 1 and Multiple Object pictures in

Block 2. Both adult and child participants were divided into the two groups described above. Based on a pilot study, we expect that children at this age will fall into two distinct groups when they receive the Multiple Object pictures first, while children's responses are largely uniform when they receive the Single Object pictures first. As discussed above, our main interest is children's responses to Multiple Object pictures. Forty-five children were tested with Multiple Object pictures first, then with Single Object pictures (TARGET group; 4;00–5;11, mean age 5;00). As a control, twenty-nine children were tested with the Single Object pictures first, and then with Multiple Object pictures (CONTROL group; 4;00–5;11, mean age 5;01). Among forty-eight adults who participated as a control, twenty-four were assigned to the TARGET group and the other twenty-four to the CONTROL group.

The experiment was designed this way to examine whether the preceding experience of watching different patterns of pictures would influence the TVJ response patterns in the latter half of the trials, allowing us to investigate whether children exhibit the tendency to persevere in their responses across blocks (e.g. Snedeker & Yuan, 2008). Each Block contained four target items in a Latin-square design; each child saw half of the test sentences with Multiple Object pictures, and the other sentences with Single Object pictures, but they never heard the same sentence twice. Note, however, our main data were obtained from children's responses in Block 1, as we anticipate children's responses in Block 1 to influence their responses in Block 2.

Procedure

TVJ/eye-tracking paradigm. Participants judged whether or not the sentence they heard and the picture on the display matched: if they thought the picture and the sentence matched, they were asked to say *Atari* ('Right') or *Atteru* ('Matched'); if they thought the picture and the sentence did not match, they were asked to say *Hazure* ('Wrong') or *Attenai* ('Not matched'). If a participant said *Hazure* ('Wrong'), the experimenter asked them why they said so, and recorded their answer.

Participants' eye-movements were monitored with a TOBII 1750 eye-tracking system (Tobii Technology AB) during the TVJ task, controlled by the experiment software E-Prime (Psychology Software Tools, Inc.). First, the subject's eye-movements were calibrated using the ClearView analysis software (TOBII Technology AB), with the 5-point calibration option. Each TVJ trial started with the presentation of a fixation point ('+') appearing in the center of the display, and the subject was asked to gaze at the fixation point. When the participant's gaze was fixated on the cross for 2 seconds, the cross disappeared and the stimulus picture

appeared. The picture was presented for 2500 milliseconds before the onset of the sentence. The eye-tracking continued until the participant responded to the TVJ task by saying 'Right' or 'Wrong'.

DCCS. All children who completed both blocks of the TVJ task, along with twenty of the adult controls, performed the DCCS. Participants were asked to sort testing cards that could be classified in terms of two competing dimensions, color and shape, first according to one dimension and then according to the other. The present study adopted the DCCS in the Standard Version, following Zelazo (2006). A session consisted of: (i) the training phase; (ii) the 'Pre-switch' Phase, in which children sorted cards based on the first sorting dimension; and (iii) the 'Post-switch' Phase, in which children sorted cards based on the second sorting dimension. The training phase with two warm-up trials immediately followed instructions telling the children to sort cards according to the first dimension. The experimenter corrected the children's mistakes during the training phase; no correction was made thereafter. The Pre-switch Phase followed the training phase. When a test card appeared, the child pointed to the sorting box he/she thought was correct. The Pre-switch Phase consisted of six sorting trials. The Post-switch Phase started immediately after a brief instruction telling children to sort cards now according to the second dimension. Children completed another six trials in this phase. Each child participated in two Sets (i.e. sorting sessions), in order to counterbalance the effect of the order of the two sorting dimensions. In Set 1, the first set of materials (e.g. 'animal' set) was used, and the other set of materials (e.g. 'vehicle' set) was used in Set 2. In Set 1, they first sorted cards according to one dimension and then according to the other; in Set 2, the order of the dimensions according to which they sorted cards was switched. The position of the two sorting boxes in each Set (e.g. the box with a black monkey on the right and the one with a green cat on the left, or vice versa) and the order of the sets (e.g. Set 1 with the 'animal' set and Set 2 with the 'vehicle' set, or vice versa) were counter-balanced across subjects.

RESULTS AND ANALYSES

We first report the overall results of the TVJ, mainly focusing on the comparison between children and adults. Then we provide a detailed report of analyses in which the participants were further divided into subgroups depending on their overall TVJ response patterns, searching for any link among children's TVJ, DCCS performance and eye-movements.

TVJ

Table 1 shows the mean proportions of LG responses for children and adults in each block. The results replicated previous findings for children's

TABLE 1. Mean proportions (and standard deviations) of LG responses in the TV \mathcal{J}

	TARGET group		CONTROL group	
	child ($N=45$)	adult ($N=24$)	child ($N=29$)	adult ($N=24$)
Block 1	0.4944 (0.4631)	0.8438 (0.2837)	0.1207 (0.3107)	0.5938 (0.4652)
Block 2	0.4833 (0.4809)	0.9167 (0.2823)	0.1648 (0.3157)	0.7188 (0.4318)

universal quantification: (i) four- and five-year-old children exhibited a strong tendency to commit SRs in both Single Object and Multiple Object conditions, compared to adults; (ii) children's tendency to exhibit the SR was lower in the Multiple Object condition.

The average LG answer rate for each participant was carried forward to a (2) age \times (2) group \times (2) block ANOVA, in which age (child vs. adult) and group (TARGET vs. CONTROL) were between-subject factors, while block (Block 1 vs. Block 2) was a within-subject factor. It revealed a significant main effect of group ($F(1, 118) = 17.209$, $p < 0.001$, $\eta_p^2 = 0.127$), age ($F(1, 118) = 37.959$, $p < 0.001$, $\eta_p^2 = 0.243$) and block ($F(1, 118) = 6.463$, $p = 0.012$, $\eta_p^2 = 0.052$). A significant interaction was reported only between block and age ($F(1, 118) = 5.062$, $p = 0.026$, $\eta_p^2 = 0.041$), and no other interactions were significant ($F(1, 118) = 2.226$, $p > 0.1$, $\eta_p^2 = 0.019$ between block and group; $F(1, 118) = 0.703$, $p > 0.4$, $\eta_p^2 = 0.006$ between group and age; $F(1, 118) = 0.002$, $p > 0.9$, $\eta_p^2 < 0.001$ between block and group and age). The post-hoc comparison of means (with Bonferroni adjustment) revealed that adults' performance was significantly better in Block 2 than in Block 1 ($p = 0.001$), while no effect of block was found in children ($p > 0.4$).

Looking at the children's data from another perspective, the lack of interaction between group and block demonstrates that TARGET children performed better in both blocks than CONTROL children. This suggests that the Multiple Object pictures presented in Block 1 (i.e. without providing children any prior experience) elicited significantly higher rates of LG responses than the Single Object pictures presented in Block 1. In addition, what children received in Block 1 significantly changed how children responded in Block 2; children's LG response rate was only 12% when Single Object pictures were given in Block 1, but it was 48% when Single Object pictures were given in Block 2. Likewise, children's LG response rate was 49% for Multiple Object pictures given in Block 1, but their LG responses decreased to 16% if Multiple Object pictures were given in Block 2. This is in contrast to adults, who gave more LG responses in Block 2, irrespective of the order of conditions. The tendency of children to perseverate in their responses across blocks will be discussed further in the

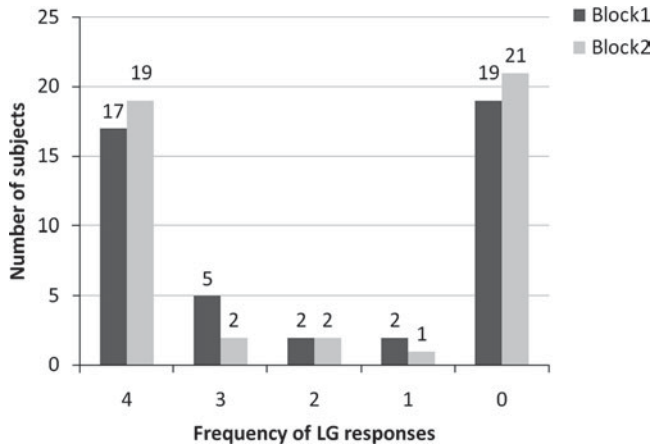


Fig. 3. Distribution of the TARGET children ($N=45$) based on the frequency of LG responses.

'Discussion' section. Note further that the Single Object pictures presented in Block 1 elicited higher percentages of SRs even from adults, suggesting that the impact of salience in the extra object is robust enough to distract adults as well. We will also return to this issue in the 'Discussion' section.

Clustering based on TVf responses. As reported above, the TARGET children as a group exhibited chance-level LG responses on average. This motivated us to further investigate whether every child uniformly gave LG responses at chance level, or whether their response patterns were distributed from child to child. Analysis of individual children's responses revealed a wide range of distribution in their response patterns; there is a clear distinction between children who almost always gave SRs and those who almost always gave LG responses (see Figure 3).

We thus conducted a hierarchical cluster analysis grouping by the Ward method, where the averaged rates of the LG responses for each Block were carried forward as the clustering parameters. There was a clear division between the two clusters of children, one that showed a high percentage of LG responses (we label this the LG-CLUSTER), and one that showed a high percentage of SRs (SR-CLUSTER, henceforth). Nearly half of the children in the TARGET group fell into the LG-cluster ($N=24$), while the other half fell into the SR-cluster ($N=21$). Note that, similar to the TARGET children, the CONTROL adults also exhibited chance level average rates of LG responses. Thus we conducted the same cluster analysis, which revealed a relatively well-balanced SR-cluster ($N=9$) and LG-cluster ($N=15$). We also attempted the same cluster analysis for the remaining groups of participants (i.e. TARGET adults, CONTROL children). Unlike the TARGET children

TABLE 2. Mean proportions (and standard deviations) of LG responses in the TVJ (block collapsed): TVJ-based cluster comparison

	TARGET group		CONTROL group	
	child (N=45)	adult (N=24)	child (N=29)	adult (N=24)
LG-cluster	0.9062 (0.1531) [24]	0.9545 (0.0908) [22]	0.7500 (0.1976) [5]	0.9667 (0.0742) [15]
SR-cluster	0.0119 (0.0376) [21]	0.0625 (0.0884) [2]	0.0162 (0.0440) [24]	0.1389 (0.2114) [9]

NOTE: Number of participants falling into each cluster is given in square brackets.

TABLE 3. Children's mean proportions (and standard deviations) of correct card sorting in the DCCS

	Set 1		Set 2	
	Pre-switch Phase	Post-switch Phase	Pre-switch Phase	Post-switch Phase
TARGET (N=45)	0.9185 (0.2553)	0.6444 (0.4545)	0.8704 (0.3235)	0.7074 (0.4161)
CONTROL (N=29)	0.9310 (0.2379)	0.5747 (0.4490)	0.8506 (0.3489)	0.8276 (0.3493)

and the CONTROL adults, the resultant division in these groups revealed a large imbalance in the number of subjects in each cluster. The majority of the CONTROL children fell into the SR-cluster (24 out of 29), and only five were in the LG-cluster. As for the TARGET adults, almost all the participants fell into the LG-cluster (22 out of 24), with only two exceptions falling into the SR-cluster. The mean proportions of LG responses in each cluster are illustrated in Table 2.

DCCS

An average accuracy of card sorting for each participant was calculated for the Pre- and Post-switch Phases. While adult controls ($N=20$) showed virtually perfect performance in card sorting throughout the entire task, children overall showed a noticeable decrease in the successful card sorting rate over the rule-switch, as shown in Table 3.

Children's data were carried forward to a (2) groups \times (2) sets \times (2) phases ANOVA, with group (TARGET vs. CONTROL) as a between-subject variable and set (Set 1 vs. Set 2) and phase (Pre-switch vs. Post-switch) as within-subject variables. The results revealed a significant main effect of phase ($F(1, 72) = 31.007$, $p < 0.001$, $\eta_p^2 = 0.301$), but no significant main effect of set ($F(1, 72) = 1.879$, $p > 0.1$, $\eta_p^2 = 0.025$) or group ($F(1, 72) = 0.054$,

TABLE 4. *TARGET* children's mean proportions (and standard deviations) of correct card sorting in the DCCS (set collapsed): DCCS-based cluster comparison

	Pre-switch Phase	Post-switch Phase
HIGH-cluster ($N=20$)	1.0000 (0.0000)	0.9628 (0.0783)
LOW-cluster ($N=25$)	0.8100 (0.2388)	0.4464 (0.2226)

$p > 0.8$, $\eta_p^2 = 0.001$). The interaction between set and phase was not significant ($F(1, 72) = 3.331$, $p = 0.072$, $\eta_p^2 = 0.044$). The post-hoc comparison of means (with Bonferroni adjustment) revealed that the Phase effect was significant in Set 1 ($p < 0.001$) but not in Set 2 ($p > 0.1$), suggesting that children's card sorting performance over the phase switch improved through sets. No other interactions were significant ($F(1, 72) = 1.897$, $p > 0.1$, $\eta_p^2 = 0.026$ between group and set; $F(1, 72) = 0.143$, $p > 0.7$, $\eta_p^2 = 0.002$ between group and phase; $F(1, 72) = 0.998$, $p > 0.3$, $\eta_p^2 = 0.014$ among set and phase and group). The results thus showed that children in the *TARGET* group did not significantly differ from those in the *CONTROL* group in their performance in DCCS; in both groups, children's performance was higher in the Pre-switch Phase than the Post-switch Phase in Set 1. Thus we can conclude that the difference in TVJ accuracy between *TARGET* children and *CONTROL* children reflects a genuine effect of salience (i.e. the number of extra objects in the pictures), confirming that children in either group were not biased in terms of their DCCS performance.

Across-task cluster comparison: relation between TVJ performance and DCCS performance

Recall that the *TARGET* children exhibited chance-level TVJ rates as a group, but yielded a well-balanced number of participants falling into two separate clusters with respect to their TVJ responses (24 in LG-cluster vs. 21 in SR-cluster). We conducted the same cluster analysis with the Ward method on their DCCS dataset, in particular, on the averaged rates of the correct card sorting both in Pre- and Post-switch Phases. Once again, there was a well-balanced division in participants in a HIGH-cluster (20 out of 45) who performed successfully both in Pre- and Post-switch Phases, and a LOW-cluster (25 out of 45) whose performance was poorer. The mean proportions of correct card sorting in each cluster are illustrated in Table 4.

In order to examine whether there is an across-task association in the children's performance in TVJ and DCCS, a chi-square test was conducted on the two-by-two clustering distribution (see Table 5), i.e. LG vs. SR based on TVJ performance during Block 1 and HIGH vs. LOW based on DCCS performance.

TABLE 5. *Number of children in each cluster in the TVJ-based (LG vs. SR) × DCCS-based (HIGH vs. LOW): results of clustering of the TARGET children and CONTROL children*

	TARGET children			CONTROL children		
	DCCS: HIGH-cluster	DCCS: LOW-cluster	Total	DCCS: HIGH-cluster	DCCS: LOW-cluster	Total
TVJ: LG-cluster	15	9	24	0	5	5
TVJ: SR-cluster	5	16	21	11	13	24
Total	20	25	45	11	18	29

TABLE 6. *TVJ-based cluster comparison of the TARGET children’s mean proportions (and standard deviations) of correct card sorting in the DCCS*

	Set 1		Set 2	
	Pre-switch Phase	Post-switch Phase	Pre-switch Phase	Post-switch Phase
LG-children (<i>N</i> = 24)	0.8958 (0.2858)	0.7778 (0.4102)	0.8611 (0.3393)	0.8194 (0.3437)
SR-children (<i>N</i> = 21)	0.9444 (0.2194)	0.4921 (0.4639)	0.8810 (0.3124)	0.5794 (0.4613)

It revealed a significant association between their TVJ performance and their DCCS performance ($\chi^2(1, N = 45) = 6.79, p = 0.016$). Having these results as a baseline implying that there is an across-task link in the performance measured in the two datasets, we further examined these cluster comparisons in more detail with respect to: (i) comparison of TVJ performance between DCCS-based HIGH/LOW clusters; and (ii) comparison of DCCS performance between TVJ-based LG/SR clusters.

Table 6 shows the comparison of the mean proportions of correct card sorting based on the TVJ clustering (LG vs. SR). We conducted a (2) sets × (2) phases × (2) clusters ANOVA, involving set (Set 1 vs. Set 2) and phase (Pre-switch vs. Post-switch) as within-subject variables and cluster (LG vs. SR) as a between-subject variable. It yielded a significant main effect of phase ($F(1, 43) = 22.544, p < 0.001, \eta_p^2 = 0.344$). The main effect of cluster was not significant ($F(1, 43) = 3.935, p = 0.054, \eta_p^2 = 0.084$), but the interaction between phase and cluster was significant ($F(1, 43) = 10.373, p = 0.002, \eta_p^2 = 0.194$). The post-hoc pair-wise comparisons of means (with Bonferroni adjustment) revealed a significant difference between LG- and SR-clusters in the Post-switch Phase ($p = 0.004$) but not in the Pre-switch Phase ($p > 0.5$). These results suggest that SR-children did worse in the rule-switch than LG-children. The main effect of set was not significant

($F(1, 43) = 0.045$, $p > 0.8$, $\eta_p^2 = 0.001$), and no interaction involving set was significant ($F(1, 43) = 0.014$, $p > 0.9$, $\eta_p^2 < 0.001$ between set and cluster; $F(1, 43) = 0.614$, $p > 0.4$, $\eta_p^2 = 0.014$ between set and phase; $F(1, 43) = 0.069$, $p > 0.7$, $\eta_p^2 = 0.002$ among set and phase and cluster).

We now turn to the comparison of the same children's TVJ responses based on the DCCS-based clusters. Whereas the mean rate of LG responses for HIGH-cluster was 0.6500 (standard deviation = 0.4148), those for LOW-cluster were 0.3600 (standard deviation = 0.4709). According to a one-way ANOVA with cluster as a between-subject variable, this difference between HIGH-cluster and LOW-cluster was significant ($F(1, 43) = 4.677$, $p = 0.036$, $\eta_p^2 = 0.098$), indicating that children who performed better in the DCCS exhibited higher rates of LG responses in the TVJ, and those who performed worse in the DCCS showed lower rates.

Taken together, the cluster comparisons across tasks conducted on the dataset from the TARGET children revealed a link between the success/failure of the logically correct TVJ in universal quantification with respect to extra-object pictures, and the success/failure of the flexible rule-switch in the DCCS.

We also attempted the same two-way clustering on the data from the CONTROL children, whose results are illustrated in Table 5. All 5 LG-children fell into the LOW-cluster, which does not apparently support the correlation between these two tasks, unlike the results from the TARGET children; but, due to the small size of the dataset, the interpretation of the result is not straightforward.

Eye-movements during TVJ

Let us now turn to the eye-movement data. The eye-tracking measurement was the participants' eye fixations to the AOI, i.e. the quadrant that contains the extra object(s). As was discussed above, it has been established that the participants' eye-movements are closely associated with the real-time comprehension of the sentence (e.g. Cooper, 1974; Tanenhaus *et al.*, 1995; Trueswell & Gleitman, 2004). Given this, measuring when and how often children look at the extra object(s) in the picture reflects how much attention children paid to the presence of the extra object(s) while computing the truth of the universally-quantified sentences.

Overall results: children vs. adults

We will first compare children's eye fixations with adults, in order to examine whether the eye-movement patterns between children and adults are similar or different during the TVJ. Figures 4 and 5 plot the trial-based mean proportions of the participants' eye fixations to the AOI both before

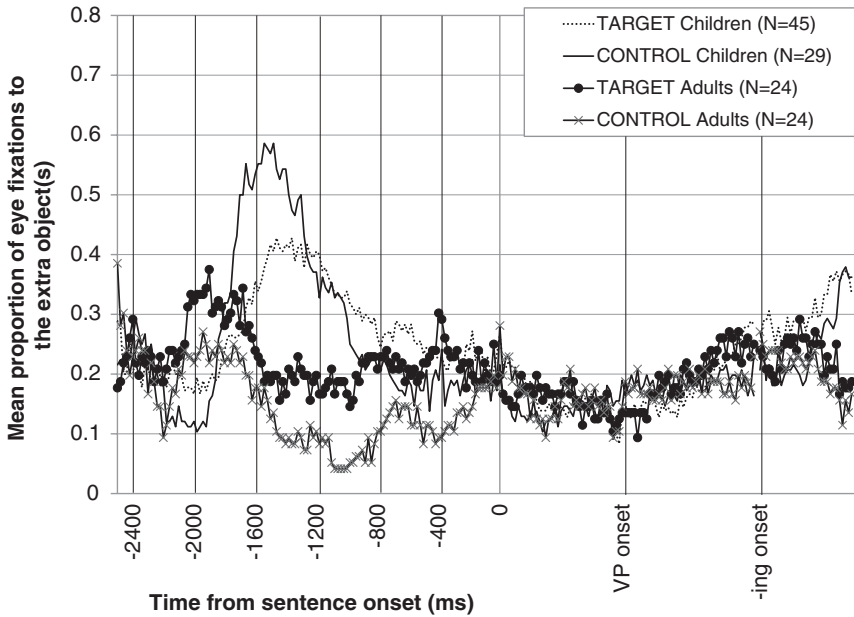


Fig. 4. Mean proportions of the eye fixations to the extra object(s) during Block 1.

and while they were hearing the target sentences presented along with the extra-object pictures.

The major difference in fixation pattern between children and adults was observed during the 2500 millisecond period before the sentence was auditorily presented in both groups; while children showed a large increase in fixations to the extra object(s), peaking around 1500 ms prior to the sentence onset, adults' increase was shallower and appeared slightly earlier. This tendency in the gaze patterns was robustly observed in Block 1 and was reduced in Block 2, but a similar pattern still held; the CONTROL adults were the exception, who exhibited an increase over the blocks. A (2) groups \times (2) ages \times (2) blocks ANOVA, with group (TARGET vs. CONTROL) and age (child vs. adult) as between-subject variables and block (Block 1 vs. Block 2) as a within-subject variable was conducted on participant-based mean proportions of fixation frequency to the extra object(s) during the 2500 ms pre-sentence period. It yielded a significant main effect of age ($F(1, 118) = 24.884, p < 0.001, \eta_p^2 = 0.174$) and of block ($F(1, 118) = 97.443, p < 0.001, \eta_p^2 = 0.452$). The interaction between age and block was also significant ($F(1, 118) = 8.11, p = 0.005, \eta_p^2 = 0.064$), and the post-hoc pair-wise comparison (with Bonferroni adjustment) suggested that: (i) children's fixation frequency was significantly higher than that

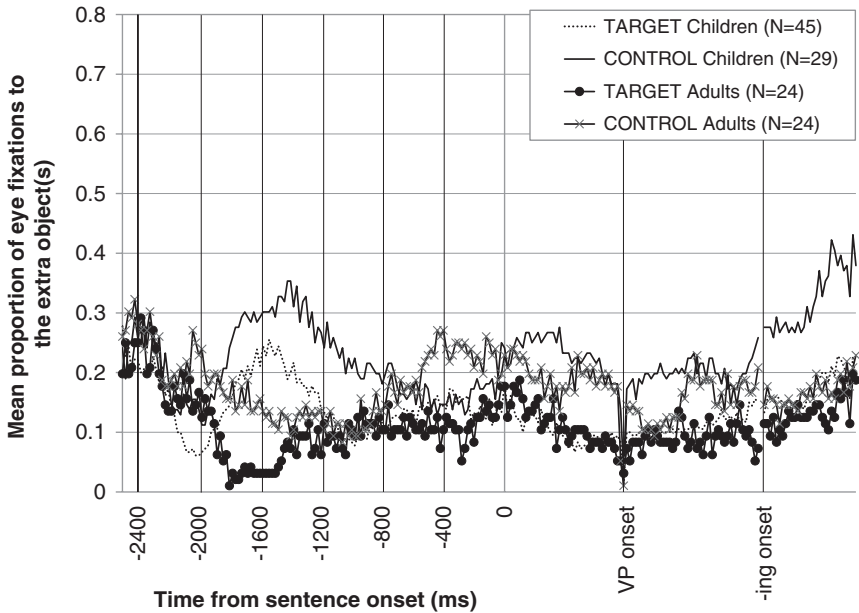


Fig. 5. Mean proportions of the eye fixations to the extra object(s) during Block 2.

of adults in both blocks ($p < 0.001$ for Block 1, $p = 0.013$ for Block 2); and (ii) the difference in fixation frequency both children and adults exhibited across the two blocks was significant ($p < 0.001$ for children, $p = 0.001$ for adults). The main effect of group was not significant ($F(1, 118) = 2.415$, $p > 0.1$, $\eta_p^2 = 0.02$), but the interaction between group and block was significant ($F(1, 118) = 40.762$, $p < 0.001$, $\eta_p^2 = 0.257$). The post-hoc pairwise comparison between group and block (with Bonferroni adjustment) suggested that the effect of block was significant in the TARGET group ($p < 0.001$) but not in the CONTROL group ($p > 0.2$). The interaction between group and age was not significant ($F(1, 118) = 2.657$, $p > 0.1$, $\eta_p^2 = 0.022$). Finally, the interaction among group, age and block was significant ($F(1, 118) = 7.066$, $p = 0.009$, $\eta_p^2 = 0.056$). The post-hoc multiple comparisons of means (with Bonferroni adjustment) revealed that: (i) the increase that the CONTROL adults exhibited over the blocks was not significant ($p = 0.06$), whereas the decrease that the other subject groups exhibited over the blocks was significant ($p < 0.001$ for all the other groups); (ii) the significant difference between children and adults within each group was observed in Block 1 ($p = 0.023$ for the TARGET group; $p < 0.001$ for the CONTROL group), but in Block 2, the difference was not significant for the TARGET group ($p = 0.051$) and for the CONTROL group ($p > 0.1$);

and (iii) the difference between the TARGET adults and the CONTROL adults was significant in both blocks ($p < 0.001$ for Block 1; $p = 0.002$ for Block 2), whereas the difference between the TARGET children and the CONTROL children was significant in Block 2 ($p < 0.001$) but not in Block 1 ($p > 0.9$).

Let us now sum up the findings. First, children exhibited a robust peak in fixation frequency around 1500 ms prior to the sentence, while adults' increase was much shallower. Second, participants generally showed a significant decrease in fixation frequency in Block 2 compared with Block 1, except for the CONTROL adults, who showed an increase in fixation in Block 2. Finally, whereas gaze patterns of the TARGET adults and the CONTROL adults significantly differed both in Block 1 and Block 2, the TARGET children and the CONTROL children exhibited the significant difference only in Block 2. Given the divergent TVJ response patterns between the TARGET children and the CONTROL children, this raises a question as to whether there is a link between the TVJ pattern and the eye-movement pattern. Recall that the TARGET children yielded numerically well-balanced clusters based on their TVJ response patterns (21 fell into SR-cluster while 24 fell into LG-cluster); as will be shown below, a closer look at the gaze data based on cluster comparison within group will reveal that the LG-children and the SR-children in the TARGET group distinguished themselves from each other in terms of eye-movement patterns, suggesting a link between TVJ performance and eye-movements.

TARGET group. We will now examine the data collected from the children and adults in the TARGET group, focusing first on the relation between TVJ response patterns and gaze patterns, in order to determine whether the age difference was observed within the same condition.¹ In order to investigate the genuine effect of each TVJ condition on participants' eye-movements without any preceding experience, we primarily focus on the eye-movement data from both groups during Block 1. Figure 6a plots the trial-based mean proportion of the participants' eye fixations to the AOI both before and while they were hearing the target sentences presented along with the extra-object pictures. Note that the data show a notable divergence in the eye-movement patterns among children depending on their TVJ response pattern during the pre-sentence period. The salient increase in the gaze at the extra object(s) around 1500 ms prior to the sentence onset was observed in SR-children; however, LG-children's increase around the same time point is much shallower.

[1] Because of the extremely small size of the SR-cluster for adults (2 adults out of 24), we collapsed the LG/SR distinctions for the gaze analysis and treated them all as a group of adult controls.

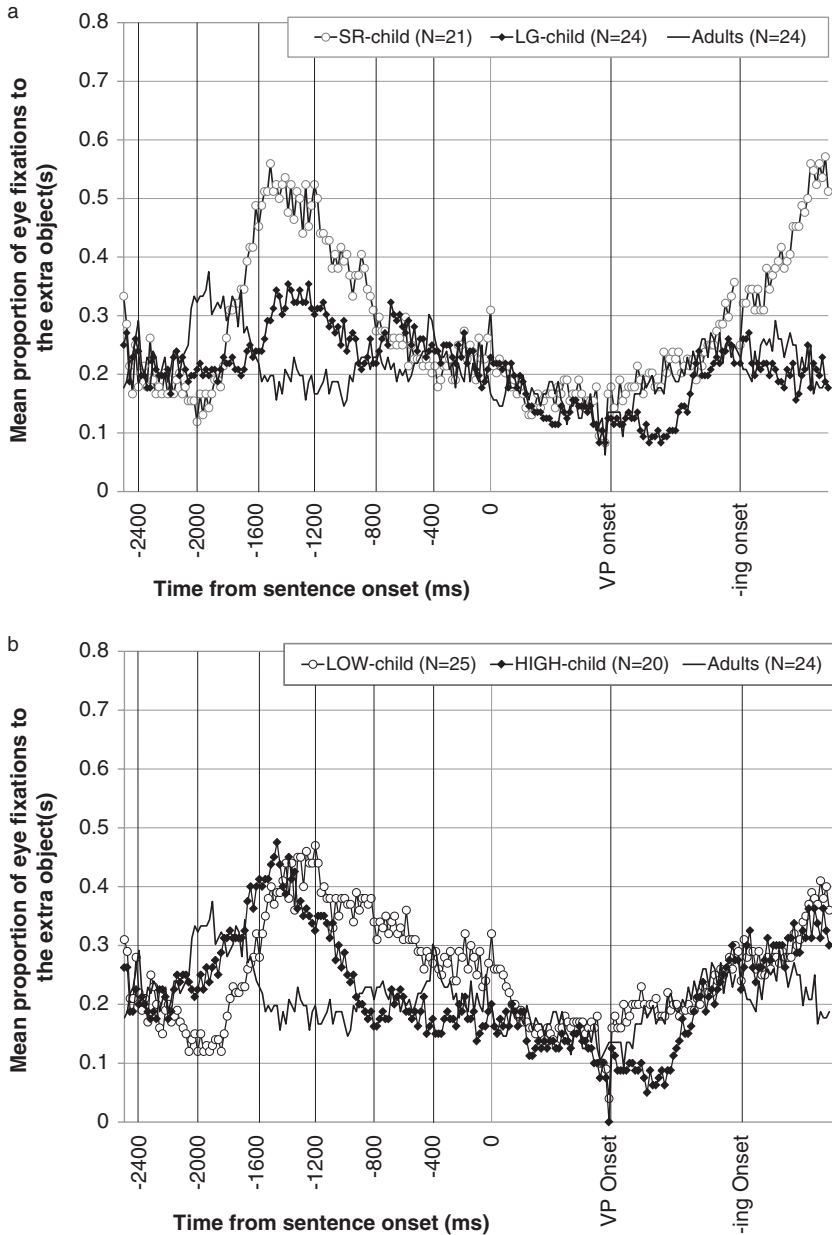


Fig. 6. a: Mean proportions of the eye fixations to the extra objects during Block 1 for the TARGET group: TVJ-based clustering. b: Mean proportions of the eye fixations to the extra objects during Block 1 for the TARGET group: DCCS-based clustering.

A one-way ANOVA involving cluster (SR-child vs. LG-child vs. adults) as a between-subject variable was conducted on the participant mean proportions of the eye fixations to the extra object(s) (i) during the 2500 ms pre-sentence period, (ii) on the successive 400 ms time windows (TWs, henceforth) starting from 2400 ms to the sentence onset, (iii) during the whole sentence presentation, and (iv) on the successive phrase-by-phrase TWs (Subject-NP, VP, *-ing*-phrase). The mean proportions of the fixation frequencies and the results of the ANOVAs are illustrated in Appendices 2a and 2b. It should be noted that the main effect of cluster became significant at the 1600–1200 TW and the 1200–800 TW. These time windows correspond to the point showing the divergence in the gaze plot, where the SR-children exhibited a robust peak in their fixation frequency. Note further that the post-hoc pair-wise comparison of means (with Bonferroni adjustment) revealed that, at the 1600–1200 TW, where SR-children exhibited a peak, there was a significant difference between the fixation frequency of SR-children on the one hand and LG-children and adults on the other ($p = 0.008$ for SR-children vs. LG-children; $p > 0.001$ for SR-children vs. adults). It should also be noted that the eye fixations of SR- and LG-children greatly diverged again during the sentence presentation, at around the *-ing*-phrase onset; while the fixation frequency of SR-children continued to increase largely, that of LG-children started to decrease. This may reflect the approximate point at which their decision-making in the TVJ was completed. We will turn to this divergence later in the ‘Discussion’ section.

Recall that, as was reported above, the TARGET children and the CONTROL children did not exhibit a significant difference in the frequency of their eye fixations to the extra object(s) during Block 1, even though their TVJ response patterns were significantly different. Interestingly, TVJ-based clustering of the TARGET children revealed that the divergent eye-movement patterns depended on their TVJ response patterns. The direct comparison of gaze data based on their TVJ responses thus demonstrated an across-task link, confirming the baseline interpretation that the increase in the eye fixations to the extra object(s) which indicates an increase in attention allocated to the extra object(s), i.e. the salient information in the pictures, results in more frequent SRs in the TVJ.

In sum, it was revealed that the eye-movement patterns of the TARGET children significantly diverged depending on their TVJ response patterns; while SR-children showed a significantly robust increase in fixation frequency to the extra object(s) during the pre-sentence period, at around 1500 ms prior to the sentence presentation, the LG-children’s increase was much shallower.

In order to determine whether there is a difference in eye-movement patterns between the HIGH-children ($N = 20$) and the LOW-children

($N=25$) within the TARGET group, we compared their eye fixations based on the DCCS-clustering conducted on the TARGET children (see Figure 6b). Although the plotted divergence in the fixation frequency is shallower than the divergence observed in the TVJ-based cluster comparison (see Figure 6a), a one-way ANOVA involving cluster (LOW-child vs. HIGH-child vs. adults) as a between-subject variable revealed significant differences at the 2000–1600 TW, 1600–1200 TW, 1200–800 TW and 800–400 TW during the 2500 ms pre-sentence period; the difference was also significant during the entire sentence presentation and the *-ing*-phrase. The mean proportions of the fixation frequencies and the results of the ANOVAs are illustrated in Appendices 3a and 3b.

In sum, the eye-movement patterns of the TARGET children diverged depending on their DCCS response patterns, as well as their TVJ response patterns. The LOW-children exhibited a significantly bigger increase in fixation frequency to the extra object during the pre-sentence period.

Recall that the TARGET children fell into four clusters according to the two-by-two clustering (TVJ-based and DCCS-based): SR-LOW cluster ($N=16$), SR-HIGH cluster ($N=5$), LG-LOW cluster ($N=9$) and LG-HIGH cluster ($N=15$), as is summarized in Table 5. In order to examine the eye-movement patterns across these clusters within TARGET children, we will now compare the TARGET children's eye fixations using the across-task clustering. As can be seen in Figure 7a, among the four clusters, SR-HIGH children exhibited the highest frequency of fixations to the extra objects during the pre-sentence period and during the sentence presentation; see Appendices 4a and 4b for the mean proportions of the fixation frequencies and the results of the ANOVAs. This finding apparently runs counter to our interpretation of the relationship among TVJ and DCCS, i.e. that SRs correlate with low performance in the DCCS. With the limited size of each resultant clustering, it is not straightforward to determine what is happening with the children who exhibited these response patterns. One possible speculation is that these children were not able to shift from the initial, extra-object-oriented perspective to the required, event-oriented perspective, because their relatively strong shifting abilities could not even override the initial perspective. We will turn to this issue in the 'Discussion' section.

CONTROL group. Finally, we will compare the eye-movement patterns of the children and adults in the CONTROL group, focusing on the data collected in Block 1, in order to determine whether the age difference was observed in the condition where Single Object pictures were presented. Figure 7b plots the trial-based mean proportions of the eye fixations to the extra object(s). It should be noted that, unlike the TARGET children, the CONTROL children in both clusters exhibited a similar gaze pattern during the pre-sentence period, with their fixation frequency increased in a similar

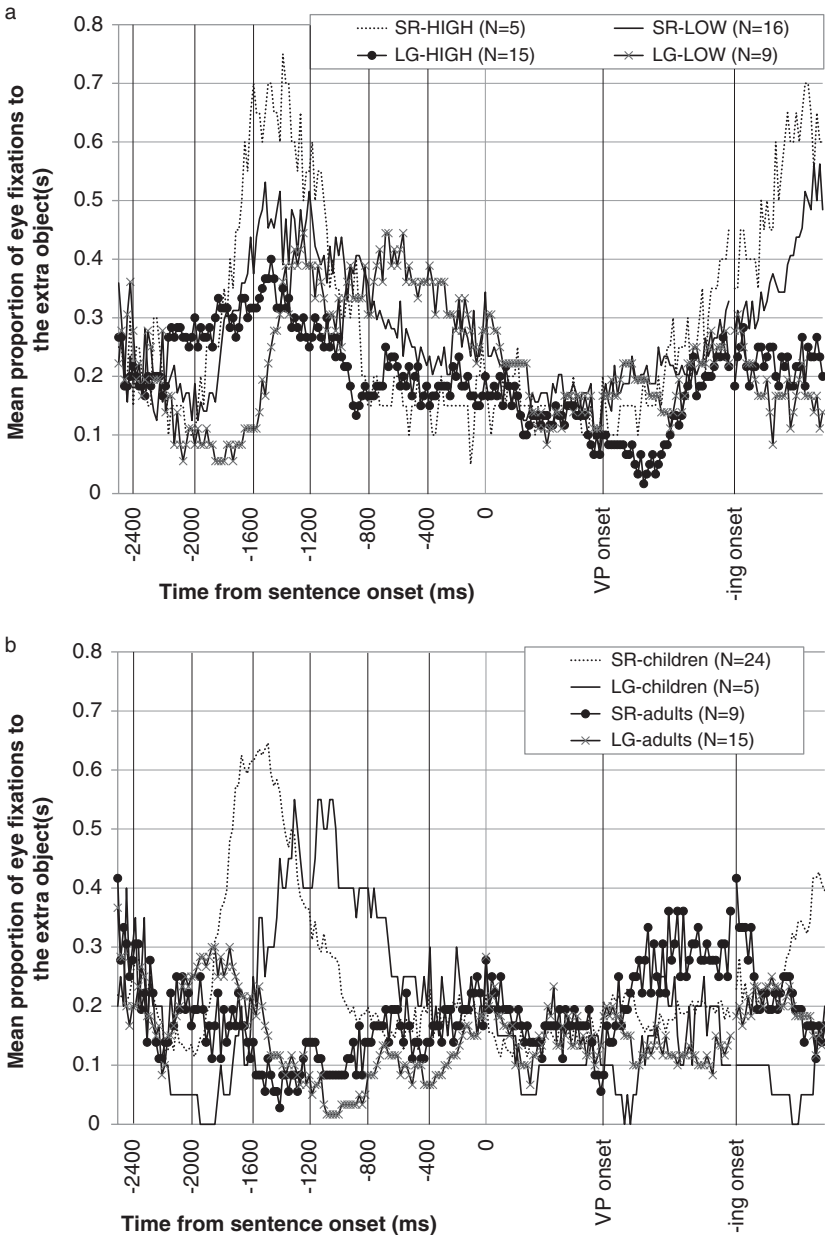


Fig. 7. a: Mean proportions of the eye fixations to the extra objects during Block 1 for the TARGET children: TVJ-based clustering x DCCS-based clustering. b: Mean proportions of the eye fixations to the extra object during Block 1 for the CONTROL group.

pattern in terms of magnitude but peaking at slightly different timings. However, considering the small number of LG-cluster CONTROL children (i.e. 5 out of 29), interpretation of the difference across clusters is not straightforward. A (2) ages \times (2) clusters ANOVA, with age (child vs. adult) and cluster (SR vs. LG) as between-subject variables, was conducted on the participant mean percentage of fixation frequency (i) during the 2500 ms pre-sentence period, (ii) on successive 400 ms TWs starting from 2400 ms to the sentence onset, (iii) during the whole sentence presentation, and (iv) on successive phrase-by-phrase TWs (Subject-NP, VP, *-ing*-phrase). The mean proportions of the fixation frequencies and the results of the ANOVAs are illustrated in Appendices 5a and 5b. The results suggest that the overall gaze pattern of children and adults diverged depending on their TVJ response patterns. However, during the 2500 ms pre-sentence period, their gaze patterns differed depending on age; the across-age divergence in the gaze pattern was robust, and the TVJ-based cluster subdivision in each age group did not reveal any significant difference by cluster. As we have noted above, however, the small size of the LG-cluster (5 out of 24) complicates the interpretation of this difference.

In addition, adults in both clusters in the CONTROL group showed similar gaze patterns, but unlike the TARGET adults, the CONTROL adults showed an increase in fixation frequency over the blocks. This may be associated with the improvements they exhibited in the TVJ response patterns over the blocks. These results may further imply that the relation between looks and correct TVJ is different for children and adults; children's looks might reflect their distraction by the extra object(s) leading to SR TVJ, while adults' looks might reflect a process of recovering from SR to LG TVJ patterns.

DISCUSSION

The major findings in the present study are summarized below: (i) in the TVJ Task, children showed the tendency to perseverate across blocks, whereas adults exhibited a significant increase in the rates of their LG responses in Block 2; (ii) children as a group exhibited chance level LG-responses when Multiple Object pictures were presented first, but the analysis of individual response patterns revealed a distinction between some children who uniformly exhibited the SRs and others who exhibited the LG responses most of the time; (iii) across-task cluster comparison within group revealed that children who mostly exhibited the SRs did significantly worse on the rule-switch measured by the DCCS than those who exhibited the LG responses; (iv) according to the analysis of the data from TARGET children, who exhibited the split TVJ response patterns, children who gave the SRs exhibited a remarkable increase in eye fixations to the extra

object(s) before the sentence was presented, whereas those who exhibited the LG responses did not show this tendency robustly; and (v) the impact of the single extra object in triggering SRs was robust even for some adults. Below we will discuss these findings from the perspective of the relation between semantic representations involved in universal quantification and the cognitive mechanisms required to construct and evaluate them in real time. In particular, we claim that the development of cognitive control is one of the factors that notably contribute to children's ability to perform adult-like universal quantification.²

Before proceeding to a detailed discussion of our hypothesis, let us first confirm two assumptions based on the findings: (i) children in our study show evidence of still-developing cognitive control; and (ii) they are at a stage in which they tend to make SRs in the universal quantifier interpretation in the extra-object context. Children in our study showed evidence of still-developing cognitive control in two ways. First, across-block comparison of the TVJ response patterns demonstrated that children showed a strong tendency to persevere in their responses across the two blocks; that is, children who showed SRs in the first block continued to show SRs in the second block irrespective of the picture type, while those who showed LG responses in the first block continued to show LG responses in the second block. In contrast, adults exhibited a significant increase in the rate of LG responses over the blocks, irrespective of the order of the type of pictures (Multiple Object vs. Single Object) presented in both blocks. These findings are consistent with findings in previous research reporting children's tendency to persevere in their responses across blocks (e.g. Snedeker & Yuan, 2008). Second, consistent with previous studies, children's performance in the DCCS on average was significantly lower than that of adults, and there are some children whose post rule-switch performance was at chance level. In addition, the results of the single-extra-object condition confirmed that children in our study are at the stage in which they show a strong tendency to show SRs in interpreting the universal quantifier when there is only one extra object.

Let us now discuss the relationship between the two tasks. The comparisons of the TARGET children's performance based on two-by-two clustering revealed an association between the success/failure in the TVJ and in the DCCS; those who showed SRs consistently in the TVJ exhibited worse performance in the rule-switch measured in the DCCS than those who showed LG responses. Furthermore, a robust divergence between

[2] An anonymous reviewer addressed an important issue regarding the order of development between linguistic knowledge and relevant cognitive ability. We would like to address that our dataset could not give us a concrete answer to this question; this issue will be one of the central topics to be discussed in future research.

SR-children and LG-children was also observed in their eye-movement patterns; SR-children showed a significantly higher rate of fixations to extra object(s) prior to sentence presentation, whereas LG-children did not exhibit this tendency robustly. Taken together, our findings show that four- and five-year-old children are at a developmental stage in which their semantic interpretation involving universal quantification is easily blocked by extralinguistic factors; due to their still-developing cognitive control, children's access to their knowledge about the universal quantifier tends to be hindered by the salience of the extra object(s) in the picture which is irrelevant information truth-conditionally.

Let us interpret our findings in terms of our hypothesis based on the Attentional Inertia account (e.g. Kirkham *et al.*, 2003). We hypothesized that attention is still developing. Such children tend to focus on the saliently remnant objects and have difficulty suppressing this initial perspective. The fact that children produced much fewer SRs when the extra objects were less salient in the Multiple Object condition shows that the salience of the remnant objects was an important factor affecting children's responses. Furthermore, the Multiple Object condition captured the transitional nature of children's ability to interpret the universal quantifier. We found that the TARGET children showed a divergent pattern of responses that can be captured by two distinct clusters, SR or LG, correlating with their performance in rule-switch in the DCCS. The implication is that the SR-cluster represent a subset of children whose cognitive control is less developed, who perceived the extra-object picture with a primary focus on the salient extra object, while the LG-cluster represents the other subset of children, whose cognitive control is more developed (though it is perhaps still developing).

Take sentence (1) and Figure 1 as a sample case. When the picture was presented in the beginning, SR-children's attention was highly attracted by the presence of the extra umbrellas in the picture. This was reflected in their gaze data, which revealed that they fixated to the extra objects significantly more frequently than the LG-children until the sentence started to play. Based on the presence of the extra objects, they established a perspective from which they perceived the picture as one about the saliently remnant umbrellas. In contrast, LG-children were not as strongly influenced by the presence of extra objects; consequently, they did not establish the same perspective.

When the sentence was provided as a description of the picture, and the children were asked to evaluate the match/mismatch of the sentence and the picture, focus had to be made on whether or not all the turtles in the picture were each holding an umbrella, in which the presence of the extra umbrellas was irrelevant information. Recall now that around the VP offset in the sentence presentation, the eye-movement pattern of SR- and LG-children

diverged again; while SR-children exhibited a continuing increase in fixation frequency to the extra objects, LG-children's fixation frequency did not show such an increase. The response patterns exhibited by SR-children are consistent with the Attentional Inertia account. The SR-children's initial perception of the picture was drawn by the salient extra objects, and they were unable to suppress the no-longer-relevant information about the extra umbrellas. Therefore, they held to their initial perspective, failing to recover and reorient to the new perspective of the picture perception as a depiction of an event of all the turtles holding an umbrella (which is required for the truth-value judgment of the sentence they are hearing). As a result, they evaluated the sentence 'incorrectly' on the basis of the presence of the extra objects in the picture. The increase of fixation frequency to the extra objects at the end of the sentence is also consistent with this account. The LG-children, in contrast, did not fixate disproportionately on the extra object either during the initial inspection of the picture or when they made their responses. This suggests that they were not unduly influenced by the remnant objects in interpreting the universal quantifier. The LG-children thus represent a subset of four- and five-year-old children whose performance in semantic computation involving universal quantification is on its way to becoming adult-like.

Let us turn to the question regarding why SRs take place. Philip (1995) attributed SRs to children's non-adult-like semantic representation of *every*-sentences, in which 'every turtle is holding a balloon' is semantically represented as 'all minimal events in which either a turtle or a balloon (or both) is a participant are events in which a turtle is holding a balloon'. Whereas his claim straightforwardly explains why children said 'No' in SRs, there could be other reasons for children to say 'No' in SRs; children might indeed have the adult-like semantic representation of the *every*-sentence, but they perceive the picture as a depiction of the extra object (based on their initial perspective, from which they could not disengage), resulting in 'No' responses on the basis of the MISMATCH BETWEEN THEIR ADULT-LIKE SEMANTIC INTERPRETATION OF THE *EVERY*-SENTENCE (which does not refer to the extra objects) AND THEIR INITIAL PERCEPTION OF THE SCENE (which they take to be 'about' the extra objects).³ Our findings, revealing

[3] Our explanation might lead to a prediction that children would say 'No' to sentences if a picture contains an extra object with or without *every*. We would like to thank the anonymous reviewer for raising this issue. There is evidence against such a prediction, showing that children overall interpreted non-*every*-sentences like 'boys are running' as true when the accompanying picture contains three running boys and a girl reading a book while sitting on a chair (Philip, 1995). Note that, in this case, the extra object is not mentioned in the sentence; in our experiment, the object was also mentioned in the sentence, but the presence of the object is not deterministic to the truth value of the sentence. We speculate that the object was mentioned not in a way that children might have expected. Questions remain regarding whether children can interpret sentences in

the across-cluster divergence in the gaze pattern according to children's TVJ response pattern, might provide support for the second possibility. Namely, children might have said 'No' based on the mismatch between the adult-like semantic representation of the *every*-sentence and the perception of the scene as 'about' the extra objects, exhibiting the robust increase in eye fixations to the extra objects. Additionally, note that Philip's claim would predict that children say 'No' in the Extra Object Condition, regardless of the salience of the extra object, e.g. the number of the extra object(s). However, as Sugisaki and Isobe (2001) and Gouro *et al.* (2001) discussed, children do not say 'No' when the salience of the extra object is modified in the pictures; our findings serve as converging evidence, showing that children's 'No' responses are linked to how much attention they paid to the presence of extra object(s). These two pieces of evidence together suggest that children's SR reflects the degree to which children commit to take the visual scene as 'about' the extra object(s).

Recall that a detailed analysis of our eye-movement data based on the two-by-two clustering of the TARGET children (TVJ-based and DCCS-based clustering) reveals that, among the four clusters (i.e. SR-HIGH, SR-LOW, LG-HIGH and LG-LOW), the SR-HIGH cluster exhibited the most eye fixations to the extra object(s). Given that the SR-HIGH group exhibited the response patterns which appear to run counter to our claim that children commit SRs due to their poor ability to shift perspectives, reflected in their low DCCS performance, this finding complicates the interpretation of our results, bringing up another possibility, i.e. that some children who pay more attention to the extra object(s) commit more SRs for reasons other than their limited flexibility in perspective switching. On the basis of the limited size of the resultant four sub-clusters in the current study, we leave this issue as an open question. Further research is needed in order to examine how the two possibilities are related.

Finally, note that a substantial number of adult participants (9 out of 24) exhibited SRs when they saw the Single Object pictures first. Given that adults presumably possess an appropriate linguistic knowledge of universal quantifier and fully developed cognitive control, this demonstrates that the impact of the salience of the single extra object was robust enough to occasionally affect even adults' judgments of universal quantifier interpretation. Even for adults, the judgment of the truth value of a universal quantifier requires the 'performance' of evaluating the visual stimuli in light of their universal quantifier knowledge; it is not enough to possess the

an adult-like way, when (i) there is an extra object in the picture, (ii) the extra object is not mentioned in the sentence that does not contain *every*, and (iii) the presence of the extra object is not deterministic to the truth of the sentence. We would like to underscore the need for further research to test such a scenario.

knowledge of the universal quantifier. It is consistent with our findings that children whose performance is hindered by their still-developing cognitive control are more likely to show SRs.

CONCLUSION

The results revealed a link between children's successful universal quantification with respect to extra-object pictures and a shallower decrease in card sorting accuracy over the rule-switch in the DCCS. Furthermore, the eye-tracking data acquired during the TVJ task revealed that children committing SRs exhibited significantly increased eye fixations to the extra object(s) prior to the auditory presentation of the sentence; such a tendency was not robustly observed when children made LG interpretations. Taken together, the present findings suggest that children's non-adult-like universal quantification with respect to extra-object pictures is considerably affected by their extralinguistic difficulty in switching perspectives using successful cognitive control in picture recognition. On the basis of these findings, we conclude that cognitive control is a factor that influences semantic processing involving universal quantification, and in children aged four to five, this is still developing. Our findings underscore the need to fully examine the relation between language acquisition and cognitive development for the ultimate understanding of the nature of children's linguistic representations and the process of child language development.

REFERENCES

- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development* **70**, 636–44.
- Bialystok, E. (2001). *Bilingualism in development: language, literacy, and cognition*. New York: Cambridge University Press.
- Bialystok, E. & Martin, M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional card sort task. *Developmental Science* **7**, 325–39.
- Conboy, B. T., Sommerville, J. A. & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology* **44**, 1505–512.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* **6**, 84–107.
- Crain, S. & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.
- Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D. & Woodams, E. (1996). Quantification without qualification. *Language Acquisition* **5**, 83–153.
- Davidson, M. C., Amso, D., Anderson, L. C. & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* **44**, 2037–78.
- Diamond, A., Carlson, M. & Beck, D. M. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology* **28**, 689–729.

- Drozdz, K. (2001). Children's weak interpretations of universally quantified questions. In M. Bowerman & S. Levinson (eds), *Conceptual development and language acquisition*, 340–76. New York: Cambridge University Press.
- Frye, D., Zelazo, P. D. & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Child Development* **10**, 483–527.
- Gourts, B. (2003). Quantifying kids. *Language Acquisition* **11**, 197–218.
- Gouro, T., Norita, H., Nakajima, M. & Ariji, K. (2001). Children's interpretation of universal quantifier and pragmatic interference. In Y. Otsu (ed.), *The proceedings of the second Tokyo Conference on Psycholinguistics*, 61–78. Tokyo: Hituzi Shobo Publishing Company.
- Gualmini, A. (2004). *The ups and downs in child language: Experimental studies on children's knowledge of entailment relationships and polarity phenomena*. New York: Routledge.
- Gualmini, A., Husley, S., Hacquard, V. & Fox, D. (2008). The Question–Answer Requirement for scope assignment. *Natural Language Semantics* **16**, 205–237.
- Inhelder, B. & Piaget, J. (1964). *The early growth of logic in the child*. New York: Routledge.
- Kirkham, N., Cruess, L. & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science* **6**, 449–67.
- Kuroda, S.-Y. (1965). Generative grammatical studies in the Japanese language. Unpublished doctoral dissertation, MIT.
- Macnamara, J. (1966). *Bilingualism and primary education: A study of Irish experience*. Edinburgh: Edinburgh University Press.
- Mazuka, R., Jincho, N. & Oishi, H. (2009). Development of executive control and language processing. *Language and Linguistic Compass* **3**(1), 59–89.
- Minai, U. (2006). Everybody knows, therefore every child knows: An investigation of logico-semantic competence in child language. Unpublished doctoral dissertation, University of Maryland.
- Myers, B. & Goldstein, D. (1979). Cognitive development in bilingual and monolingual lower-class children. *Psychology in the Schools* **16**, 137–42.
- Nishigauchi, T. (1999). Quantification and *wh*-constructions. In N. Tsujimura (ed.), *The handbook of Japanese linguistics*, 269–96. Oxford: Blackwell.
- Philip, W. (1995). Event quantification in the acquisition of universal quantification. Unpublished doctoral dissertation, University of Massachusetts.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Routledge.
- Rakhlin, N. (2007). Semantic manifestations of the developing theory of mind. Unpublished doctoral dissertation, University of Connecticut.
- Shimoyama, J. (2006). Indeterminate phrase quantification in Japanese. *Natural Language Semantics* **14**, 139–73.
- Snedeker, J. & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language* **58**, 574–608.
- Sugisaki, K. & Isobe, M. (2001). Quantification without qualification without plausible dissent. In J.-Y. Kim & A. Werle (eds), *The proceedings of the first Conference on the Semantics of Under-represented Languages in the Americas*, 97–100. Amherst, MA: GLSA.
- Takahashi, D. (2002). Determiner raising and scope shift. *Linguistic Inquiry* **33**, 575–615.
- 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**, 1632–34.
- Trueswell, J. C. & Gleitman, L. (2004). Children's eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. In J. M. Henderson & F. Ferreira (eds), *The interface of language, vision, and action: Eye movements and the visual world*, 319–46. New York: Psychology Press.
- Trueswell, J. C., Sekerina, I., Hill, N. M. & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* **73**, 89–134.

- Zelazo, P. D. (2006). The dimensional change card sort (DCCS): Aa method of assessing executive function in children. *Nature Protocols* **1**, 297–301.
- Zelazo, P. D., Frye, D. & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development* **11**, 37–63.
- Zelazo, P. D., Müller, U., Frye, D. & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development* **68** (3, Serial No. 274).

APPENDIX 1: COMPLETE LIST OF STIMULUS SENTENCES

Test sentences

1. Dono-kame-mo kasa-o sashi-teruyo.
which-turtle-also umbrella-ACC hold-ing
'Every turtle is holding an umbrella.'
2. Dono-kuma-mo terebi-o hakon-deruyo.
which-bear-also TV-ACC carry-ing
'Every bear is carrying a TV.'
3. Dono-risu-mo botan-o mot-teruyo.
which-chipmunk-also button-ACC hold-ing
'Every chipmunk is holding a button.'
4. Dono-panda-mo keeki-o tabe-teruyo.
which-panda bear-also cake-ACC eat-ing
'Every panda bear is eating a piece of cake.'
5. Dono-saru-mo kyabetsu-o kajit-teruyo.
which-monkey-also cabbage-ACC nibble-ing
'Every monkey is nibbling a cabbage.'
6. Dono-zou-mo kan-o ket-teruyo.
which-elephant-also can-ACC kick-ing
'Every elephant is kicking a can.'
7. Dono-tanuki-mo hon-o yon-deruyo.
which-raccoon-also book-ACC read-ing
'Every raccoon is reading a book.'
8. Dono-neko-mo jyuusu-o non-deruyo.
which-cat-also juice-ACC drink-ing
'Every cat is drinking juice.'

Warm-up sentences

9. Dono-buta-mo ne-teruyo.
which-pig-also sleep-ing
'Every pig is sleeping.'

10. *Dono-kitsune-mo hashit-teruyo.*
 which-fox-also run-ing
 'Every fox is running.'

Filler sentences

11. *Dono-niwatori-mo nai-teruyo.*
 which-chicken-also cry-ing
 'Every chicken is crying.'
12. *Dono-ushi-mo nabagutsu-o hai-teruyo.*
 which-cow-also boots-ACC wear-ing
 'Every cow is wearing a pair of boots.'
13. *Dono-raion-mo aoi taiko-o tatai-teruyo.*
 which-lion-also blue drum-ACC hit-ing
 'Every lion is hitting a blue drum.'
14. *Dono-usagi-mo pinku-no ame-o name-teruyo.*
 which-rabbit-also pink lollipop-ACC lick-ing
 'Every rabbit is licking a pink lollipop.'
15. *Dare-mo booshi-o kabut-te-naiyo.*
 who-also cap-ACC wear-ing-NEG
 'Nobody is wearing a cap.'
16. *Dare-mo hane-te-naiyo.*
 who-also jump-ing-NEG
 'Nobody is jumping.'
17. *Dare-mo hana-o kuwae-te-naiyo.*
 who-also flower-ACC hold in mouth-ing-NEG
 'Nobody is holding a flower in his mouth.'

APPENDIX 2A

Mean proportions (and standard deviations) of eye fixations during the TVJ for the TARGET group: TVJ-based clustering

Time window	SR-child ($N=21$)	LG-child ($N=24$)	Adult ($N=24$)
pre-sentence 2500–0 ms	0.3133 (0.0682)	0.2581 (0.0902)	0.2382 (0.0972)
pre-sentence 2400–2000 ms	0.1833 (0.0929)	0.2042 (0.1057)	0.2354 (0.0986)
pre-sentence 2000–1600 ms	0.2554 (0.2214)	0.2141 (0.2169)	0.3130 (0.1571)
pre-sentence 1600–1200 ms	0.4994 (0.2503)	0.3083 (0.1883)	0.1969 (0.1761)
pre-sentence 1200–800 ms	0.4024 (0.1974)	0.2693 (0.2028)	0.1906 (0.1731)
pre-sentence 800–400 ms	0.2530 (0.1662)	0.2641 (0.1982)	0.2161 (0.2055)
pre-sentence 400–0 ms	0.2238 (0.1884)	0.2286 (0.1914)	0.2234 (0.1524)
entire sentence presentation	0.2661 (0.1171)	0.1712 (0.0748)	0.1938 (0.0828)
Subject-NP	0.1688 (0.1016)	0.1531 (0.1221)	0.1581 (0.1010)
VP	0.2148 (0.1831)	0.1507 (0.1041)	0.1953 (0.1227)
-ing-phrase	0.4147 (0.2125)	0.2097 (0.1247)	0.2281 (0.1445)

APPENDIX 2B

Results of the one-way ANOVAs on the mean proportions of eye fixations during the TVJ for the TARGET Group: TVJ-clustering

Time window	$F(2, 66)$	p	η_p^2
pre-sentence 2500–0 ms	4.428*	0.016	0.118
pre-sentence 2400–2000 ms	1.573	0.215	0.046
pre-sentence 2000–1600 ms	1.487	0.233	0.043
pre-sentence 1600–1200 ms	12.326***	<0.001	0.272
pre-sentence 1200–800 ms	6.952**	0.002	0.174
pre-sentence 800–400 ms	0.408	0.667	0.012
pre-sentence 400–0 ms	0.006	0.994	<0.001
entire sentence presentation	6.372**	0.003	0.162
Subject-NP	0.119	0.888	0.004
VP	1.290	0.282	0.038
-ing-phrase	10.688***	<0.001	0.245

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

APPENDIX 3A

Mean proportions (and standard deviations) of eye fixations during the TVJ for the TARGET Group: DCCS-based clustering

Time window	LOW-child ($N = 25$)	HIGH-child ($N = 20$)	Adult ($N = 24$)
pre-sentence 2500–0 ms	0.2982 (0.0830)	0.2659 (0.0851)	0.2382 (0.0972)
pre-sentence 2400–2000 ms	0.1800 (0.1040)	0.2125 (0.0926)	0.2354 (0.0986)
pre-sentence 2000–1600 ms	0.1825 (0.2077)	0.2969 (0.2177)	0.3130 (0.1571)
pre-sentence 1600–1200 ms	0.3935 (0.2337)	0.4025 (0.2476)	0.1969 (0.1761)
pre-sentence 1200–800 ms	0.3810 (0.2304)	0.2694 (0.1641)	0.1906 (0.1731)
pre-sentence 800–400 ms	0.3170 (0.1703)	0.1863 (0.1734)	0.2161 (0.2055)
pre-sentence 400–0 ms	0.2700 (0.1769)	0.1719 (0.1913)	0.2234 (0.1524)
entire sentence presentation	0.2661 (0.1171)	0.1712 (0.0748)	0.1938 (0.0828)
Subject-NP	0.1757 (0.1126)	0.1413 (0.1110)	0.1581 (0.1010)
VP	0.2077 (0.1710)	0.1468 (0.1082)	0.1953 (0.1227)
-ing-phrase	0.3093 (0.1923)	0.3004 (0.2108)	0.2281 (0.1445)

APPENDIX 3B

Results of the one-way ANOVAs on the mean proportions of eye fixations during the TVJ for the TARGET Group: DCCS-clustering

Time window	$F(2, 66)$	p	η_p^2
pre-sentence 2500–0 ms	2.801	0.068	0.078
pre-sentence 2400–2000 ms	1.941	0.152	0.056
pre-sentence 2000–1600 ms	3.227*	0.046	0.089
pre-sentence 1600–1200 ms	6.528**	0.003	0.165

APPENDIX 3B (Cont.)

Time window	$F(2, 66)$	p	η_p^2
pre-sentence 1200–800 ms	5.983*	0.004	0.153
pre-sentence 800–400 ms	3.222*	0.046	0.089
pre-sentence 400–0 ms	1.786	0.176	0.051
entire sentence presentation	1.057	0.353	0.031
Subject-NP	0.568	0.569	0.017
VP	1.156	0.321	0.034
-ing-phrase	1.407	0.252	0.041

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

APPENDIX 4A

Mean proportions (and standard deviations) of eye fixations during the TVJ for the TARGET children: TVJ-clustering \times DCCS-clustering

Time window	DCCS: LOW		DCCS: HIGH	
	SR ($N=16$)	LG ($N=9$)	SR ($N=5$)	LG ($N=15$)
pre-sentence 2500–0 ms	0.3124 (0.0725)	0.2729 (0.0984)	0.3162 (0.0592)	0.2492 (0.0873)
pre-sentence 2400–2000 ms	0.1836 (0.0947)	0.1736 (0.1248)	0.1825 (0.0975)	0.2225 (0.0921)
pre-sentence 2000–1600 ms	0.2383 (0.2318)	0.0833 (0.1066)	0.3100 (0.1967)	0.2925 (0.2306)
pre-sentence 1600–1200 ms	0.4555 (0.2528)	0.2833 (0.1505)	0.6400 (0.2026)	0.3233 (0.2113)
pre-sentence 1200–800 ms	0.4055 (0.2219)	0.3375 (0.2521)	0.3925 (0.0998)	0.2283 (0.1627)
pre-sentence 800–400 ms	0.2813 (0.1603)	0.3806 (0.1781)	0.1625 (0.1682)	0.1942 (0.1802)
pre-sentence 400–0 ms	0.2469 (0.1888)	0.3111 (0.1549)	0.1500 (0.1865)	0.1792 (0.1987)
entire sentence presentation	0.2562 (0.1231)	0.1860 (0.1666)	0.2979 (0.1006)	0.1623 (0.0801)
Subject-NP	0.1746 (0.1102)	0.1778 (0.1205)	0.1500 (0.0628)	0.1383 (0.1248)
VP	0.2137 (0.1923)	0.1970 (0.1350)	0.2186 (0.1699)	0.1229 (0.0720)
-ing-phrase	0.3802 (0.2015)	0.1833 (0.0828)	0.5250 (0.2313)	0.2256 (0.1447)

APPENDIX 4B

Results of the 2 × 2 ANOVAs on the mean proportions of eye fixations during the TVJ for the TARGET children: TVJ-clustering × DCCS-clustering

Time window	TVJ main effect			DCCS main effect			TVJ × DCCS interaction		
	<i>F</i> (1, 41)	<i>p</i>	η_p^2	<i>F</i> (1, 41)	<i>p</i>	η_p^2	<i>F</i> (1, 41)	<i>p</i>	η_p^2
pre-sentence 2500–0 ms	3.817	0.058	0.085	0.133	0.717	0.003	0.257	0.615	0.006
pre-sentence 2400–2000 ms	0.202	0.656	0.005	0.511	0.479	0.012	0.559	0.459	0.013
pre-sentence 2000–1600 ms	1.542	0.221	0.036	4.090	0.050	0.094	0.979	0.328	0.023
pre-sentence 1600–1200 ms	11.533**	0.002	0.220	2.434	0.126	0.056	1.008	0.321	0.024
pre-sentence 1200–800 ms	3.027	0.089	0.069	0.838	0.365	0.020	0.520	0.475	0.013
pre-sentence 800–400 ms	1.324	0.257	0.031	7.186*	0.011	0.149	0.353	0.556	0.009
pre-sentence 400–0 ms	0.573	0.454	0.014	3.437	0.071	0.077	0.081	0.778	0.002
entire sentence	10.027**	0.003	0.197	0.076	0.784	0.002	1.016	0.319	0.024
presentation									
Subject-NP	0.012	0.912	<0.001	0.710	0.404	0.017	0.038	0.846	0.001
VP	1.324	0.257	0.031	0.502	0.483	0.012	0.656	0.423	0.016
-ing-phrase	19.568***	<0.001	0.323	2.781	0.103	0.064	0.837	0.366	0.020

p* < 0.05; *p* < 0.01; ****p* < 0.001.

APPENDIX 5A

Mean proportions (and standard deviations) of eye fixations during the TVJ for the CONTROL Group

Time window	Children		Adults	
	SR (<i>N</i> = 24)	LG (<i>N</i> = 5)	SR (<i>N</i> = 9)	LG (<i>N</i> = 15)
pre-sentence 2500–0 ms	0.2890 (0.0707)	0.2646 (0.0703)	0.1646 (0.0590)	0.1510 (0.0430)
pre-sentence 2400–2000 ms	0.1880 (0.1272)	0.1600 (0.0903)	0.2042 (0.1269)	0.1875 (0.0983)
pre-sentence 2000–1600 ms	0.3495 (0.2102)	0.0625 (0.0685)	0.1736 (0.0983)	0.2592 (0.1560)
pre-sentence 1600–1200 ms	0.5359 (0.1919)	0.3625 (0.3820)	0.0833 (0.1021)	0.1325 (0.1363)
pre-sentence 1200–800 ms	0.2547 (0.2299)	0.4450 (0.2032)	0.1069 (0.1201)	0.0400 (0.0753)
pre-sentence 800–400 ms	0.1672 (0.1662)	0.2850 (0.1599)	0.1583 (0.1132)	0.1000 (0.1154)

APPENDIX 5A (Cont.)

Time window	Children		Adults	
	SR (N=24)	LG (N=5)	SR (N=9)	LG (N=15)
pre-sentence 400–0 ms	0.1849 (0.1695)	0.2075 (0.2168)	0.1819 (0.1823)	0.1217 (0.1480)
entire sentence presentation	0.2063 (0.1172)	0.1103 (0.0455)	0.2167 (0.0723)	0.1600 (0.0578)
Subject-NP	0.1466 (0.1114)	0.1162 (0.1303)	0.1681 (0.1189)	0.1517 (0.1089)
VP	0.1919 (0.1475)	0.1395 (0.1354)	0.2571 (0.1577)	0.1318 (0.0724)
-ing-phrase	0.2806 (0.2202)	0.0750 (0.0589)	0.2250 (0.1706)	0.1967 (0.1451)

APPENDIX 5B

Results of the 2 × 2 ANOVAs on the mean proportions of eye fixations during the TV \ddot{y} for the CONTROL Group

Time window	Age main effect			Cluster main effect			Age × Cluster interaction		
	F(1, 49)	p	η_p^2	F(1, 49)	p	η_p^2	F(1, 49)	p	η_p^2
pre-sentence 2500–0 ms	38.829***	<0.001	0.442	0.829	0.364	0.017	0.072	0.790	0.001
pre-sentence 2400–2000 ms	0.310	0.580	0.006	0.330	0.568	0.007	0.023	0.881	<0.001
pre-sentence 2000–1600 ms	0.561	0.458	0.011	1.724	0.195	0.034	11.163**	0.002	0.186
pre-sentence 1600–1200 ms	39.243***	<0.001	0.445	0.550	0.462	0.011	3.262	0.077	0.062
pre-sentence 1200–800 ms	18.662***	<0.001	0.276	0.538	0.467	0.011	4.906*	0.031	0.091
pre-sentence 800–400 ms	2.620	0.112	0.051	0.124	0.726	0.003	3.532	0.066	0.067
pre-sentence 400–0 ms	0.390	0.535	0.008	0.213	0.647	0.004	0.564	0.456	0.011
entire sentence presentation	0.747	0.392	0.015	6.242*	0.016	0.113	0.439	0.511	0.009
Subject-NP	0.562	0.457	0.011	0.377	0.542	0.008	0.036	0.850	0.001
VP	0.891	0.350	0.018	5.036*	0.029	0.093	0.736	0.395	0.015
-ing-phrase	0.025	0.876	0.001	3.090	0.085	0.059	2.214	0.143	0.043

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.