Preschoolers' comprehension of pronouns and reflexives: the impact of the task

CHRISTINA BERGMANN, MARKUS PAULUS and PAULA FIKKERT

Journal of Child Language / Volume 39 / Issue 04 / September 2012, pp 777 - 803
DOI: 10.1017/S0305000911000298, Published online: 14 September 2011

J. Child Lang. **39** (2012), 777–803. © Cambridge University Press 2011
doi:10.1017/S0305000911000298

Preschoolers' comprehension of pronouns and reflexives: the impact of the task*

CHRISTINA BERGMANN
Faculty of Arts, Radboud University Nijmegen, the Netherlands

MARKUS PAULUS
Ludwig Maximilian University, Munich, Germany

AND

PAULA FIKKERT
Faculty of Arts, Radboud University Nijmegen, The Netherlands

(*Received* 27 *October* 2009 – *Revised* 24 *December* 2010 – *Accepted* 27 *July* 2011 – *First published online* 14 *September* 2011)

ABSTRACT

Pronouns seem to be acquired in an asymmetrical way, where children confuse the meaning of pronouns with reflexives up to the age of six, but not vice versa. Children's production of the same referential expressions is appropriate at the age of four. However, response-based tasks, the usual means to investigate child language comprehension, are very demanding given children's limited cognitive resources. Therefore, they might affect performance. To assess the impact of the task, we investigated learners of Dutch (three- and four-year-olds) using both eye-tracking, a non-demanding on-line method, and a typical response-based task. Eye-tracking results show an emerging ability to correctly comprehend pronouns at the age of four. A response-based task fails to indicate this ability across age groups, replicating results of earlier studies. Additionally, biases seem to influence the outcome of the response-based task. These results add new evidence to the ongoing debate of the asymmetrical acquisition of pronouns and reflexives and suggest that there is less of an asymmetry than previously assumed.

[*] The authors wish to thank the team at the Baby Research Centre Nijmegen, the Netherlands, which greatly supported this study, in particular Angela Khadar. Furthermore, we are thankful to the two anonymous reviewers and the editors who provided insightful recommendations on a previous version of this article. Address for correspondence: Christina Bergmann, Faculty of Arts, Radboud University, P. O. Box 9103, 6500 HD Nijmegen, The Netherlands. e-mail: C.Bergmann@let.ru.nl

INTRODUCTION

Pronouns, such as *he* and *her*, are among the most frequent expressions in English, Dutch and a variety of other languages (Francis & Kučera, 1982; Baayen, Piepenbrock & van Rijn, 1993). Their interpretation, however, is more complex than one would assume based on their high frequency. The pronoun itself provides only vague information that can guide interpretation, such as the gender, number and animacy of the intended referent. Nonetheless, a link between the pronoun and the correct referent is often established easily. Additionally, the distribution and interpretation of pronouns seems to be closely related to reflexives, such as *herself*, due to their complementary occurrence, which is illustrated in sentences (1) and (2).

(1) Is Mama Bear touching herself?
(2) Is Mama Bear touching her?

The correct referent for the reflexive in (1) is found within the same sentence, whereas the pronoun in (2) cannot be linked to *Mama Bear*, and the referent must thus be located outside the sentence.

The acquisition of such referential expressions has been subject to investigation and continues to puzzle researchers since early studies by Jakubowicz in 1984. First, children start to comprehend reflexives as early as four years of age, whereas their ability to correctly interpret pronouns emerges only later, at six and a half years (6;6). More precisely, children were found to confuse the meaning of pronouns with reflexives, but not vice versa (e.g. Chien & Wexler, 1990; Jakubowicz, 1984; for an overview see, e.g., Conroy, Takahashi, Lidz & Phillips, 2009). Second, children correctly produce both pronouns and reflexives in the appropriate contexts from the age of four years onwards (Bloom, Barss, Nicol & Conway, 1994; de Villiers, Cahillane & Altreuter, 2006; Matthews, Lieven, Theakston & Tomasello, 2009).

A classical investigation was conducted by Chien and Wexler (1990), where the authors used a truth-value judgement task to assess pronoun comprehension. When confronted with a picture showing a bear character touching herself, rather than the girl depicted next to her, children at the age of up to 6;6, when hearing phrase (2), incorrectly answered 'yes' in about 50% of the trials. Whereas, upon hearing sentence (1) and seeing a picture of the bear character touching the girl, the same children correctly rejected the question in 84% of the trials. These results indicate knowledge about reflexives, which receive the correct interpretation within the sentence, since the reflexive refers to the subject of sentence (1). At the same time, children seem to be unaware of the fact that a pronoun has to refer to a different entity than the subject. The correct referent of the pronoun has

not been mentioned within sentence (2), hence the discourse context has to provide the correct meaning.

These findings are surprising, as their conjunction implies the inability to comprehend what can be produced by the same grammar. A number of theories have been put forward to capture the complementary distribution of pronouns and reflexives (nativist accounts based on considerations by Chomsky, 1981: Reinhart, 2011; Conroy *et al.*, 2009; de Villiers *et al.*, 2006; Elbourne, 2005; Optimality Theory: Hendriks & Spenader, 2005/2006; usage-based: Matthews *et al.*, 2009). The most demanding objective of these accounts, however, is to explain the gap of several years that separates the acquisition of these seemingly closely related expressions and the observation that the production of pronouns precedes comprehension.

Theoretical considerations aside, a number of studies have attempted to replicate this phenomenon of a delay in pronoun comprehension in various languages, such as English (e.g. Matthews *et al.*, 2009) and Dutch, Spanish and Italian (for a comparative study of the three aforementioned languages see Ruigendijk, Baauw, Zuckerman, Vasic, de Lange & Avrutin, 2011). Interestingly, not every study investigating pronouns has reproduced the early findings. For example, Spenader, Smits and Hendriks (2009) found for Dutch that children from the age of four onwards perform at adult level when confronted with pronouns in a more natural pragmatic environment. In everyday discourse, as given in (3), the pronoun refers to a previously introduced and established topic. In this context children correctly map the pronoun onto the extra-sentential antecedent.

(3) This is Goldilocks. Mama Bear is touching her.

Across studies, the group average of correct answers when confronted with a pronoun varies between 16% and 82% (as summarised by Conroy *et al.*, 2009). This high variance is, according to Conroy and colleagues, not in accord with the assumption that children resort to guessing and thus perform on chance level as predicted by Chien and Wexler (1990) and Grodzinsky and Reinhart (1993), among others. Rather, a number of researchers (Bloom *et al.*, 1994; Conroy *et al.*, 2009; Elbourne, 2005; Hendriks & Koster, 2010; Matthews *et al.*, 2009) suggested that these highly unstable outcomes are due to experimental artefacts. One source of varying performance might be located in cues to the correct referent, caused by factors such as an unbalanced discourse or a salient depiction of only one potential referent (cf. Elbourne, 2005; Spenader *et al.*, 2009). In consequence, a more carefully balanced design is necessary that makes the grammatical antecedent as pragmatically available as the ungrammatical antecedent. Therefore, both the introduction and number of previous

occurrences should be balanced across potential referents and images used in the experiments should show both figures with equal size and colouration.

Another possible cause for diverse results may be that the experiment itself has an impact on performance (as suggested by Bloom *et al.*, 1994; Conroy *et al.*, 2009; Matthews *et al.*, 2009; Reinhart, 2011). Studies investigating pronoun comprehension usually rely on response-based methods, where a child is instructed to give a verbal response or perform a predefined action. The assumption is that the response (directly or indirectly) reflects the comprehension of pronouns. However, more factors than mere comprehension affect the generation of an appropriate answer. The child has to remember the actual task, attend to the stimuli, compare auditory and visual information, select the correct and expected reply and actually give an answer. All these processes, in addition to language comprehension, are measured during the administration of a response-based method. Furthermore, the memory capacity and attention system of four- to six-year-olds is limited in comparison to adults' cognitive capacities (e.g. Baddeley, 1992; Cowan, Nugent, Elliott, Ponomarev & Saults, 1999).

Despite the ongoing discussion about the limits of a child's cognitive resources affecting pronoun comprehension (e.g. Hendriks & Koster, 2010), a possible impact of the task itself on children's performance during experiments has not yet been subject to investigation. Given that children's performance level in response-based experiments is potentially influenced by task demands and cognitive load, another form of measure is called for. This method has to be less demanding than a response-based task and should impose a minimum of additional processing cost. An on-line measure of response that tracks comprehension as it unfolds would be ideal.

Eye-tracking can be used to measure on-line processing of linguistic input in adults (e.g. Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Allopenna, Magnuson & Tanenhaus, 1998) and in children (e.g. Trueswell & Gleitman, 2004; Trueswell, Sekerina, Hill & Logrip, 1999). Participants usually inspect images while linguistic input is presented auditorily. At the same time eye-movements reveal underlying comprehension processes. For example, when naming an object, both adults and children will direct their attention and thus their gaze automatically at the visual representation within 200 to 400 ms after the naming (e.g. Arnold, Brown-Schmidt & Trueswell, 2007). Hence, eye-tracking is a very suitable method to test comprehension processes with the benefit of relatively low task demands since the allocation of attention and the resulting eye-movements are largely involuntary and automatic. Experiments using this method are thus less demanding in terms of attention and memory, since there is no mandatory secondary task.

A first study contrasting off-line and on-line behaviour during the comprehension of referential expressions was conducted by Sekerina, Stromswold and Hestvik (2004), where the on-line fixation behaviour was

compared to the outcome of a pointing task conducted in parallel to assess the resolution of ambiguous pronouns. This study showed a discrepancy between implicit knowledge revealed in fixation behaviour and explicit performance during the pointing task. More precisely, the eye-movements showed awareness of the ambiguity, while children's pointing behaviour consistently showed preference of one referent, without consideration of the other possibility. However, this study did not reduce processing cost for two reasons: first, the pronouns could refer to two possible referents; and second, there was a response-based task superimposed onto the eye-movement measurements, where the former possibly influenced the latter (cf. Yarbus, 1967).

The present experiment aims at disentangling the potential effects of the task itself and children's linguistic abilities. To this end, we contrasted eye-tracking with a separate, response-based off-line task, more precisely, a picture selection task. During eye-tracking, the participants simply saw pictures of toys and listened to sentences, which included either a pronoun or a reflexive in object position, while their eye-movements were recorded. In the response-based task, two images depicting either the pronominal or reflexive interpretation of an action were presented alongside the respective sentence, which had previously occurred during eye-tracking. The participant was asked to indicate the correct picture based on his or her comprehension. The picture selection task was designed to replicate previous findings comparing pronoun comprehension to the ability to correctly interpret reflexives. In the present study, we used both the on-line and the off-line method with the same children. The participants were divided into two groups with a mean age of three and four years, respectively. The choice of the two age groups was motivated by previous studies, where an emerging gap between the comprehension of pronouns and reflexives was found from the age of four onwards.

METHODS

Participants

Twenty-two children (14 male, aged 2;11 to 3;1, mean age 3;0·04) contributed to the data of the age group of three-year-olds. Two additional children were tested, but had to be excluded due to fussiness.

The group of four-year-olds consisted of sixteen children (8 male, aged 3;9 to 3;11, mean age 3;10·04). Four children were tested in addition but had to be excluded due to fussiness (3) or uncooperative behaviour (1).

All children were monolingual speakers of Dutch, recruited via the database of the Baby Research Centre Nijmegen, The Netherlands. As a reward for participation, the parents could choose between a book and monetary compensation.

Stimulus material

The same visual and linguistic stimuli appeared in both eye-tracking and the picture selection task. During eye-tracking, the children saw a picture and heard the auditory material at the same time. The very same pictures were presented to the participants as plastic cards during the picture selection task while parents read the sentences that were heard previously during eye-tracking. Both the visual and linguistic stimuli are motivated and described in depth below.

Linguistic stimuli. Two pairs of characters were employed: a cat and a cow were chosen as feminine actors, a frog and a bear constituted the masculine pair. The Dutch translation of the species served as name and the Dutch translation of titles 'miss' and 'mister' preceded the generic name to clarify and enforce the assigned gender (resulting e.g. in *Meneer Kikker* 'Mister Frog').

In standard Dutch, two reflexive forms exist: *zich* and *zichzelf*. The actual mechanism guiding the distribution of both reflexives is under debate (e.g. Hendriks, Spenader & Smits, 2008; Reuland, 2001): *zich* is generally assumed to occur with verbs that frequently take a reflexive meaning, whereas *zichzelf* is used in contexts where a non-reflexive meaning is more common. However, systematic testing of this assumption only accounted for at most half the data (45%) in a large parsed corpus of Dutch (Hendriks *et al.*, 2008).

One distinguishing property of these reflexive forms is that only *zichzelf* can take on contrastive and emphatic stress. Since the word and nuclear stress within the sentences should be balanced across reflexive and pronoun conditions, the unstressed form *zich* was chosen as reflexive in the present experiment to match the unstressed pronoun. Furthermore, the choice of *zich* as reflexive allowed for a balanced number of syllables across conditions.

The sentences were constructed to allow for the occurrence of either a reflexive or a pronoun in an otherwise identical linguistic environment. Dutch transitive verbs taken from the CHILDES database (Van Kampen corpus, files from children between the second and third birthday; van Kampen, 1994) were rated by forty native speakers for naturalness when being paired with either a reflexive or a pronoun. Six verbs with a high rating for naturalness for both types of referential expression were chosen for the present study (namely *aankleden*, *afdrogen*, *insmeren*, *krabben*, *uitkleden* and *wassen* – 'dressing', 'drying', 'to apply lotion', 'scratching', 'undressing' and 'washing', respectively). All four characters were paired with the six verbs to generate the environment for either a reflexive or a pronoun in object position. Examples are given in sentences (4a) and (4b).

(4a) *Meneer Kikker is hem aan het aankleden.* (Pronoun)
Mister Frog is him at the on·dress-INF
'Mister Frog is dressing him.'
(4b) *Meneer Kikker is zich aan het aankleden.* (Reflexive)
Mister Frog is REFL at the on·dress-INF
'Mister Frog is dressing himself.'
(5a) *Meneer Kikker kleedt hem aan.* (Pronoun)
Mister Frog dress-3SGPRES him on
'Mister Frog is dressing him.'
(5b) *Meneer Kikker kleedt zich aan.* (Reflexive)
Mister Frog dress-3SGPRES REFL on
'Mister Frog is dressing himself.'

A progressive sentence structure such as in sentences (4a) and (4b) was chosen for two reasons. First, in Dutch, separable complex verbs, such as *aankleden* ('to dress'), occur with a sentence-final particle in the simple present, as in sentences (5a) and (5b). The same verb behaves like an atomic verb in progressive structures. As the particle disambiguates separable complex verbs (such as distinguishing *aankleden*, 'to dress', from *uitkleden*, 'to undress'), a non-divided presentation was preferred to reduce additional processing costs (Hillert & Ackerman, 2002). Second, upon encountering the pronoun or reflexive, no verb semantics have shaped a possible expectation concerning the referential expression.

Combining all factors yielded 4 (characters) × 6 (verbs) × 2 (conditions) resulting in 48 unique sentences. The stimulus sentences were recorded by a female native speaker of Dutch in slow child-directed speech.

The resulting audio material was spliced into two segments. The first part constituted the name of the respective agent of the sentence (e.g. *Meneer Beer*, 'Mister Bear'). For all stimuli, this initial part was identical across conditions and verbs. The remainder of the sentence, starting at the auxiliary *is*, constituted the second segment. The onset of the critical word was approximately at the same time point for all sentences and began 200 to 300 ms after offset of the direct naming. Across the conditions, that is, for all sentences containing either a pronoun or a reflexive, sentence stress was matched within each sentence pair to exclude stress as a cue to reference resolution (Reinhart, 2011). The nuclear stress was on the sentence-final main verb.

To control for effects of the visual material on the fixation behaviour during eye-tracking, such as salient figures or picture-inherent fixation dynamics, we displayed the visual material for one second before the linguistic stimuli were presented. During this period, the children had time to inspect the picture and familiarize themselves with the scenery. Hence, effects of the sudden onset of the picture and of sections that initially pop

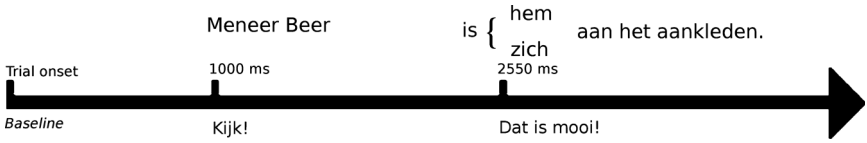


Fig. 1. The time course of the stimulus sentences during eye-tracking (translations: 'Mister Bear is dressing him/himself.'; baseline: 'Look! This is nice!').

out can be assumed to have subsided by sentence onset. This initial time to visually examine the picture at the beginning of each trial is typical for visual world eye-tracking studies. Furthermore, and more importantly, a baseline condition was added (Elbourne, 2005). During this baseline condition, which was recorded by the same native speaker of Dutch as the stimulus sentences, the children did not hear any of the critical referential expressions (cf. Figure 1). Instead of referring to the depicted characters, the baseline sentences started with an initial exclamation, *Kijk!* ('Look!'), as a non-specific reference to the picture, followed by a general positive statement, such as *Wat mooi!* ('How nice!'). The baseline stimuli were spliced in a similar manner as the test sentences. The initial exclamation was matched with the onset of the naming of the sentence subject and the remainder started at the same time point as *is* (see Figure 1). Thus, the general reference in the baseline sentences occurred at the same time as the critical referential expression in the respective conditions. All recordings were inspected by a second native speaker of Dutch, who judged the recordings to sound natural.

Visual stimuli. For each character, one specific toy was used. Additionally, the scenarios were constructed to match the sentences, where each verb was visualized by an action. More precisely, the agent's arm was arranged in a position designed to be natural and typical. Two verbs required an accessory: *wassen* ('to wash') was shown using a sponge and *drogen* ('to dry') a towel. Across conditions, the body part touched was kept constant. Accordingly, touching the agent's head in the reflexive version was paired with touching the patient's head in the transitive counterpart of the respective sentence. To minimize the difference between pictures, nothing apart from the action was varied (see Figure 2, 'Procedure' section, for an example pair).

For eye-tracking, all sound files were combined with the matching images, yielding 48 avi-videos with a duration of five seconds each. For the baseline condition, six sentences were randomly distributed among the 48 pictures. Each of the short avi-videos was mirrored to balance the side of the target. Overall, 192 avi-videos were used in the experiment, comprising 48 (unique sentence–picture pairs) × 2 (mirrored) avi-videos, and an equal number of videos (96) for the baseline condition.

Fig. 2. Example picture pair in the picture selection task. The respective stimulus sentences are: *Mevrouw Koe is haar aan het wassen. Op welk plaatje is mevrouw Koe haar aan het wassen?* ('Miss Cow is dressing her. In which picture is miss Cow dressing her?').

Apparatus

During the eye-tracking part of the experiment, the gaze of both eyes was recorded using a corneal reflection eye-tracker (Tobii 1750; Tobii Technology, Stockholm, Sweden). The position of the eye-tracker was adjusted to the child's head position to maintain the viewing distance of approximately 60 cm with the screen being parallel to the participant's face.

For the picture selection task, the images were colour-printed and laminated in an A5-sized (214 × 154 mm) clear plastic pouch to obtain durable plastic cards.

Procedure

During the experiment, an introductory play session was followed by eye-tracking and then by the picture selection task. The order of eye-tracking preceding the picture selection task was not varied to avoid the influence of a learning effect on the eye-tracking results. If such a learning effect influences the participants' behaviour in the picture selection task, a difference between the present data and previous studies should be observed. No such data exist for the eye-tracking part of our experiment.

Play session. To familiarize the children with the experimenter and with the toy figures that were used throughout the experiment, a play session preceded every experimental session. The experimenter introduced one of the toy figures at a time and directed the child's attention to it. A subsequent short game, which included either the child naming the toy figures or pointing to the correct one when the experimenter named it, ensured knowledge of the names assigned to the toy figures.

Eye-tracking. During the eye-tracking session the children sat on their parent's lap in front of the adjusted screen or on a chair with their parent behind them. The participants were instructed to watch 'television'; no additional task was imposed. The parents were instructed not to interfere or interact with their child during the experiment.

Prior to testing, the gaze of each participant was calibrated. A pulsating dot, accompanied by a siren-like sound to attract attention, served as calibration stimulus in a 3×3 point grid procedure on a black background. In the event that less than eight of the nine points were calibrated successfully, the calibration was repeated for the missing calibration points, otherwise the experiment started. During the calibration and the presentation of the stimulus material, parents wore headphones and listened to music intermixed with speech recordings that were not related to this study. To the child, the experimental acoustic stimuli were presented via external loudspeakers at a set level.

All stimuli were presented using the software ClearView in full-screen resolution at 1280×1024 pixels. The trials were blocked for pairs of characters and thus for gender. Within one block three pronoun and three reflexive items were intermixed with the six corresponding baseline stimuli of the same images. After five items, an attention getter in the form of a pulsating star or diamond with a matching sound was inserted. At the beginning of a block both characters appearing in the subsequent stimuli were introduced by centrally displaying only one toy on the screen while simultaneously playing audio stimuli consisting of an exclamation and a direct naming (e.g. *Kijk, Meneer Beer!* 'Look, Mister Bear!').

Four blocks in total constituted one eye-tracking session, which lasted five minutes. These four blocks could only contain 24 of the 48 generated sentences, which were paired with the matching 24 baseline sentences. To evenly distribute all generated stimuli across children, the set of stimuli was divided, resulting in two sets of 24 sentences where each character pair carries out every possible action. Consequently, each child saw only half of the stimulus sentences.

To account for effects of the presentation sequence, the trials were shown in reverse order to half of the children. Additionally, the position of the sentence agent as well as the target figure on the screen was varied within participants and counterbalanced across participants. In sum, four variants of the experiment existed.

Picture selection task. The 24 sentence–picture pairs used in the picture selection task were identical to the reflexive and pronoun stimuli previously encountered during eye-tracking. This means all participants saw the same visual stimuli used during eye-tracking now on plastic cards and heard the respective sentences again that were presented to them previously during eye-tracking.

The items were presented in two blocks. Within each block, only one pair of characters occurred to avoid confusion, similar to the blocked presentation during eye-tracking.

Child and parent sat at opposite sides of a table, separated by a wooden display to block the parents' view of the cards. The experimenter was seated perpendicular to the display in a position that allowed easy mounting of pairs of cards on it. To practise the task, two training trials preceded the actual experiment, where the children saw a pair of characters, with each card depicting only one toy. The task was to indicate one of the two characters. This procedure both reassessed each child's knowledge of the names and ensured that the actual task was clear to the participant.

Throughout the experiment, one card displayed the correct action matching the stimulus sentence, the other card depicted the counterpart (Figure 2). These two cards made up a pair of pictures with one depicting a reflexive and one showing a transitive action. At the same time the parent read a sentence matching one of the depicted scenes without seeing the cards. The child was asked to indicate his or her choice by pointing or touching, after which the experimenter took both cards away. Upon inattention the sentence was reread once at most. When the participant failed to make a choice after hearing the sentence twice, the trial was excluded from further analysis. Each indication of one card was followed by encouraging remarks from the experimenter to maintain the child's attention and motivation.

Data analysis

Eye-tracking. The eye-movement data recorded by the Tobii system were evaluated for fixations. The criterion for a fixation was that the mean of both eyes remained within the radius of 30 pixels on the screen for at least 100 ms. The resulting data were exported via the Tobii software ClearView to a text file and read into the statistical analysis environments MATLAB (The Mathworks, Inc., Natick, Massachusetts, USA) and R (R Foundation for Statistical Computing, Vienna, Austria).

The basis for further analysis was the evaluation of the fixation data relative to the portion of the depicted scene the gaze fell upon. To this end, regions of interest (ROI) were predefined individually for each image (Figure 3). Each fixation was assigned to the specific ROI it occurred in. The first ROI covered the half of the screen which was occupied by the sentence's agent, the figure carrying out the depicted action. The second ROI included the other half of the screen, where either the patient or the distractor, in short the other toy figure, was shown. The action part was defined separately, since this was the part of the pictures that changed the most between any two trials. Furthermore, the direct verbal reference draws

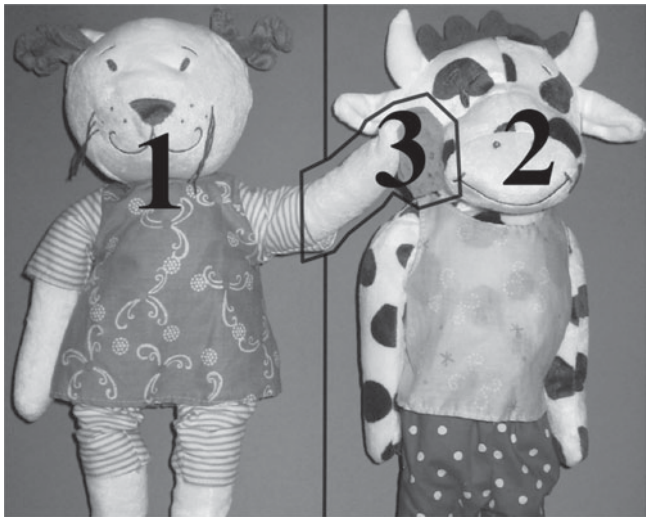


Fig. 3. Illustration of the three regions of interest (ROI). All three areas are non-overlapping. ROI 1 indicates the sentence agent; ROI 2 is the patient of the sentence or the distractor in reflexive sentences; ROI 3 is the action to exclude effects of this salient part of the images.

attention to the part of the image depicting the action (the figure's arm touching either itself or the other figure). To disentangle fixations guided by the critical referential expressions from looks attracted by changing and more salient portions of the visual material, a region around the upper half of the arm was defined as a third ROI. This part of the image contained the action-related changes between images. There was no overlap between the three ROI within one picture, as Figure 3 illustrates.

After assigning one ROI to each fixation, the total time of one trial was divided into 50 ms segments. The time steps were centred around the onset of the critical referential expression, which occurred 2550 ms after the start of each trial. For every bin of 50 ms, the fixation probability on each of the three ROI per participant and condition were calculated by aggregating the data over all verbs and normalizing over all fixations in that particular time bin. To qualify any effect of the critical referential expression, mixed-model Analyses of Variance (ANOVAs) were conducted on the data of each condition. The pronoun and reflexive conditions were analyzed separately to exclude possible effects of the visual material. We expected changes in fixation behaviour depending on the images used to represent either a transitive or reflexive action due to the inherent differences in the visual material. For this reason we opted to include baseline stimuli that use the

same images but no critical referential expression. The trials with either a pronoun or a reflexive are thus compared to the respective baseline and not to each other. In the ANOVA per condition, that is, for either the pronoun or the reflexive and their corresponding baseline stimuli, the dependent variable was the fixation probability on the correct figure. In the pronoun condition, this figure was the patient of the depicted action; in the reflexive condition, the correct figure was the agent.

Based on previous research on children's eye-movements, we expected an effect to occur at a time point from 200 ms after the onset of the critical referential expression (cf. Hallet, 1986; Pyykkönen, Matthews & Järvikivi, 2010). An overall time slot of 500 ms, beginning 200 ms after onset of the referential expression, was chosen. This time slot is assumed to capture effects of the referential expression, whereas fixations based on the verb at the end of the sentence should not be included.

Factors of the ANOVA were Age (between participants with age 3 and age 4), Time (representing the 50 ms bins from 200 to 700 ms after onset of the referential expression, 10 time intervals within participants), and Baseline vs. Test Condition (2 levels within participants).

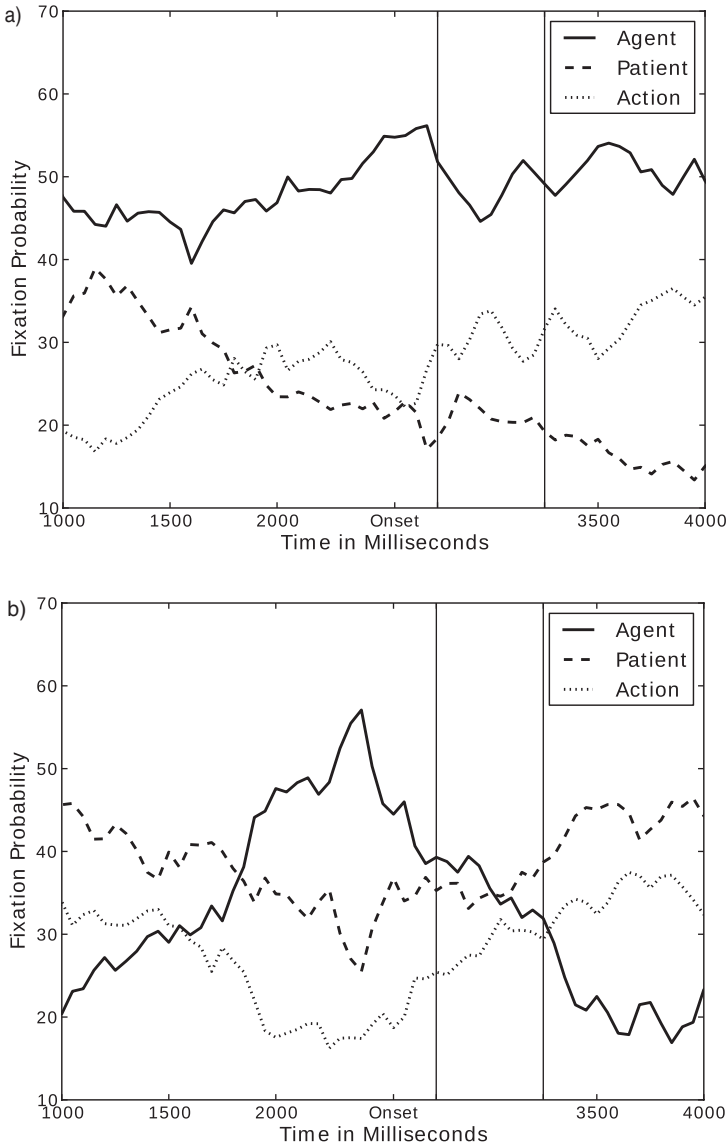
Picture selection task. In the picture selection task, the performance of each participant was assessed per condition. The completed number of trials were coded for a correct or incorrect response, where the former means choosing the picture corresponding to the sentence heard and an incorrect response occurred when both pictures were indicated or the wrong card depicting a non-matching scene was chosen. The accuracy per participant and condition was analyzed using a mixed-model ANOVA with the factors Age (age 3 and age 4, between participants) and Condition (2 levels, Pronoun and Reflexive, within participants).

RESULTS

Eye-tracking

The overall fixation dynamics in the pronoun and reflexive trials, separated by age group, are depicted in Figures 4a–d. The fixation behaviour during the corresponding baseline conditions where no referential expression occurred can be inspected in Figures 5a–d. The time slot of interest, set at 200 to 500 ms after onset of the reflexive or pronoun, constituted the basis for all analyses reported in this section. The time slot was chosen to capture short-lived, dynamic comprehension processes that occur automatically upon hearing a referential expression.

In general, the two conditions, namely the reflexive condition and the pronoun condition, seem to have elicited very different fixation behaviours. However, when comparing the fixation probabilities on the respective figures to the baseline condition, an effect of visual material emerges. As


a)
Agent
Patient
Action
Fixation Probability
70
60
50
40
30
20
10
1000
1500
2000
Onset
3500
4000
Time in Milliseconds
b)
Agent
Patient
Action
Fixation Probability
70
60
50
40
30
20
10
1000
1500
2000
Onset
3500
4000
Time in Milliseconds

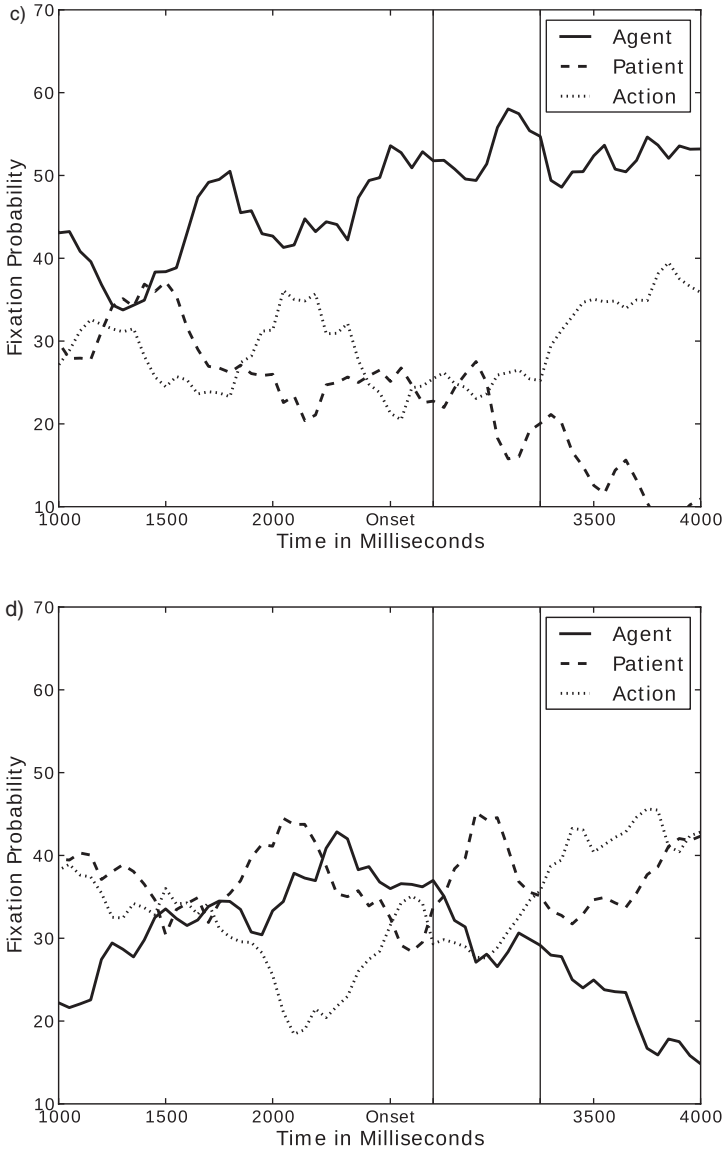
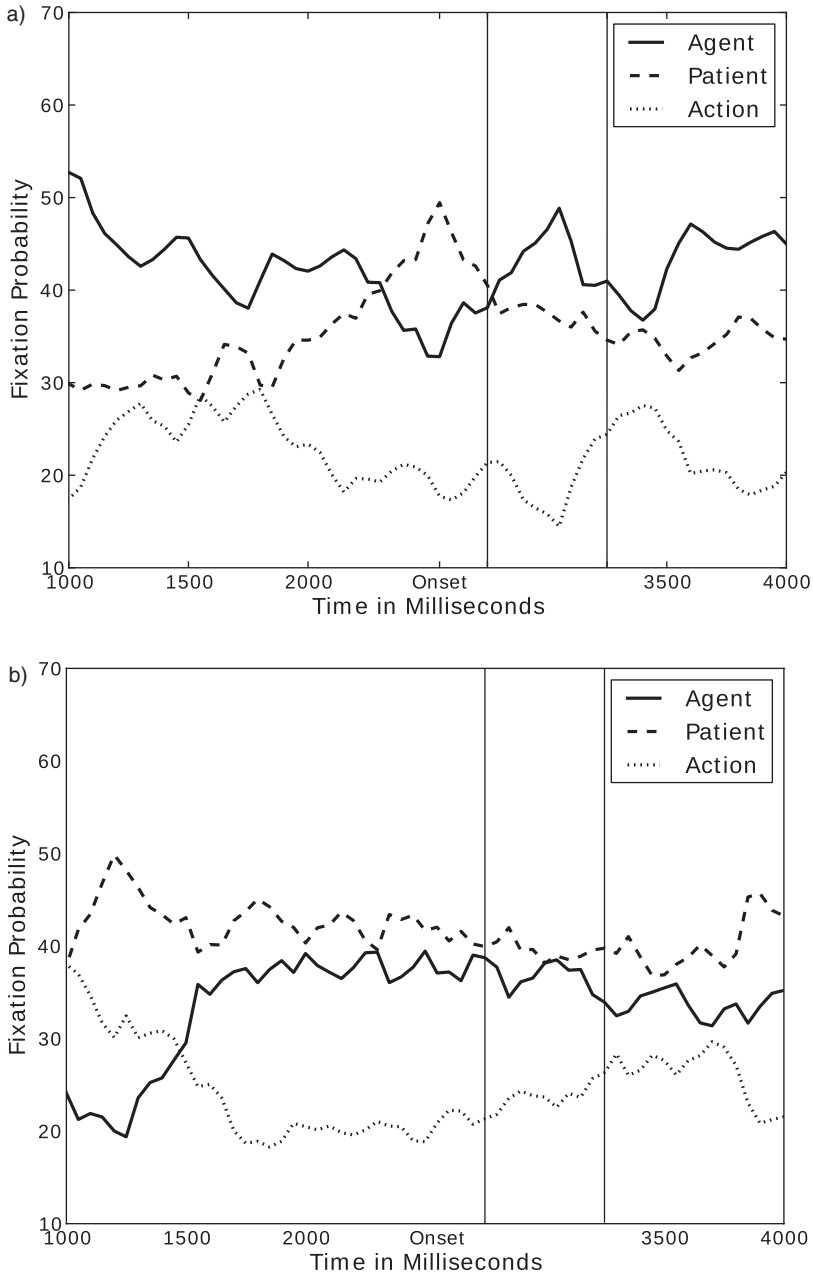


Fig. 4. Time course of the fixation probability in the pronoun and reflexive condition per age group. The depicted group averages are split by Region of Interest (ROI, cf. Figure 3). The vertical bars indicate the time slot of interest used in the statistical analyses. The onset of the referential expression is additionally marked on the *x*-axis. a) Age group 3, Reflexive Condition, b) Age group 3, Pronoun Condition, c) Age group 4, Reflexive Condition, d) Age group 4, Pronoun Condition.


a)
b)
Agent
Patient
Action
Fixation Probability
Time in Milliseconds
Onset
10
20
30
40
50
60
70
1000
1500
2000
3500
4000

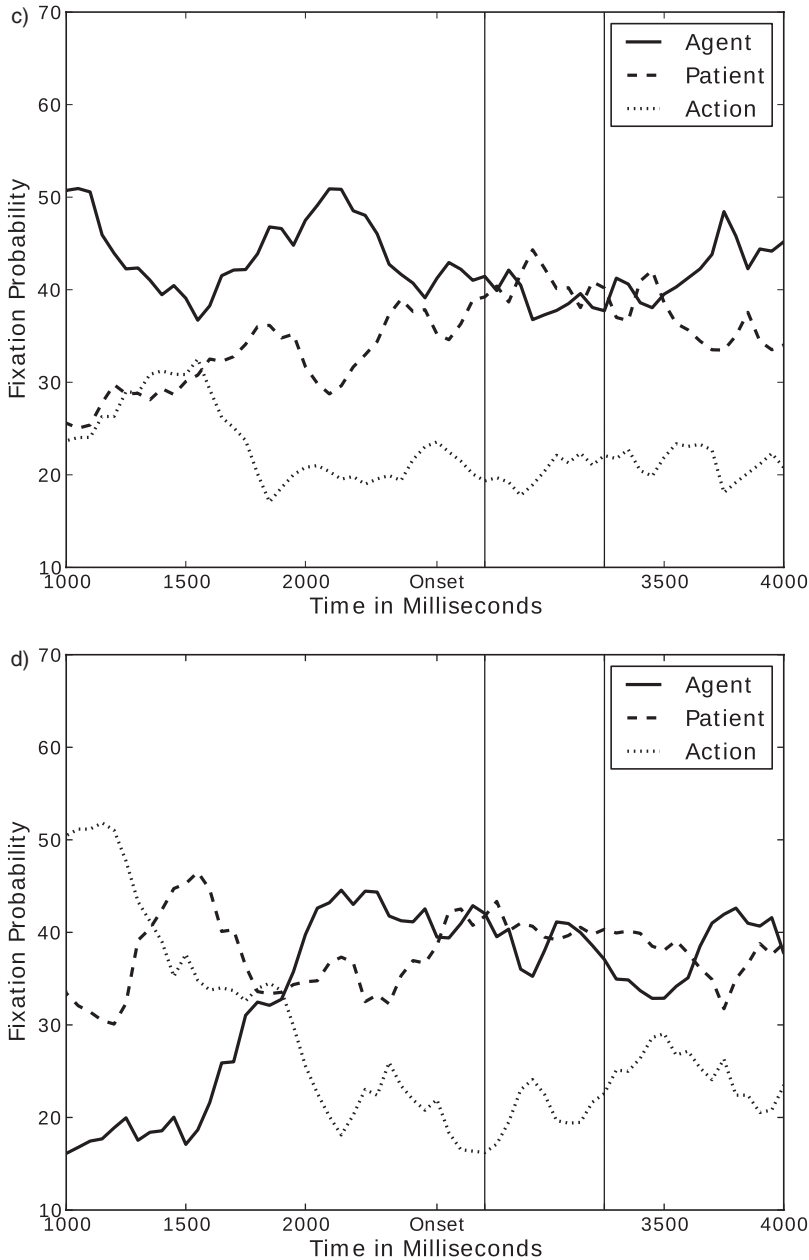


Fig. 5. For Legend see opposite page.

discussed in the previous section on data analysis, the possibility of such an effect of visual material led to the conduction of separate mixed-model ANOVAs for each condition.

When investigating the pronoun condition, the ANOVA showed a significant effect of the factor Age ($F(1,36) = 5{\cdot}5976$, $p = 0{\cdot}018$). In addition, significant interactions of the within participant factors Baseline vs. Pronoun and Time, representing the difference between trials with and without a pronoun at the critical time point and the bin of aggregated fixation probability respectively, with Age were found (Age × Baseline vs. Pronoun: $F(1,36) = 4{\cdot}0008$, $p = 0{\cdot}045$; Age × Time: $F(10,360) = 9{\cdot}9125$, $p = 0{\cdot}001$). The three-way interaction of all factors approached significance ($F(1,36) = 3{\cdot}3966$, $p = 0{\cdot}065$).

To further investigate the effects within each age group, the dataset was split according to the participant's age and two separate ANOVAs were conducted, the remaining within participant factors were Baseline and Time. For the three-year-olds, the ANOVA did not reveal any significant effects (all ps $> 0{\cdot}1$). For the older age group, the four-year-olds, a main effect of Baseline vs. Pronoun was found ($F(1,15) = 5{\cdot}380$, $p = 0{\cdot}036$). No other effect or interaction was significant (all Fs < 1).

For the reflexive condition, the ANOVA across age groups showed a significant main effect of Age on the fixation probabilities ($F(1,36) = 8{\cdot}6284$, $p = 0{\cdot}003$). Furthermore, an interaction of Baseline vs. Reflexive with Age reached significance ($F(1,36) = 4{\cdot}4531$, $p = 0{\cdot}035$). A main effect of Time approached significance ($F(1,36) = 3{\cdot}0594$, $p = 0{\cdot}08$). No other effect or interaction was significant (all Fs < 1). Separate ANOVAs per age group showed no significant effect for the three-year-olds (all ps $> 0{\cdot}1$). For the four-year-olds, the difference between reflexive trials and baseline was significant, as indicated by a main effect of Reflexive vs. Baseline ($F(1,15) = 29{\cdot}346$, $p < 0{\cdot}001$). No other effect or interaction reached significance (all Fs < 1).

Picture selection task

The overall performance of the participants across age groups and conditions during the picture selection task is displayed in Figure 6. The mean accuracy per age group and condition increased in the reflexive

Fig. 5. Time course of the fixation probability in the baseline conditions corresponding to the pronoun and reflexive trials per age group. The depicted group averages are split by Region of Interest (ROI, cf. Figure 3). The vertical bars indicate the time slot of interest used in the statistical analyses. The time point where the onset of the referential expression would occur in the pronoun and reflexive trials is additionally marked on the *x*-axis for comparison. In the filler sentence, this time point coincides with the onset of the general positive expression (e.g. 'How nice!'). a) Age group 3, Reflexive Baseline, b) Age group 3, Pronoun Baseline, c) Age group 4, Reflexive Baseline, d) Age group 4, Pronoun Baseline

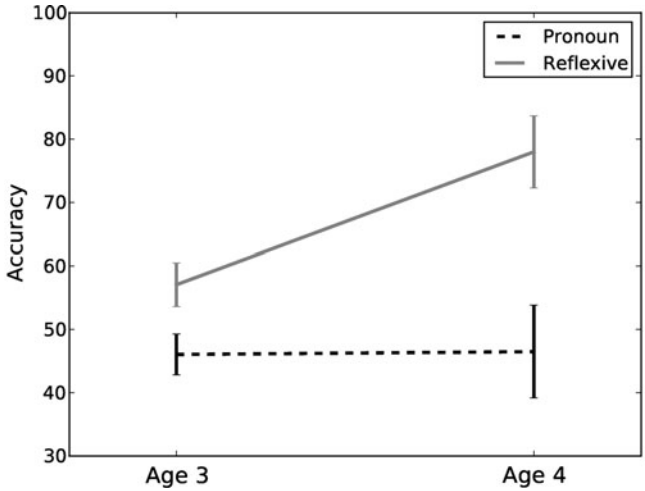


Fig. 6. Performance in the picture selection task across age groups and conditions, averaged within groups. The accuracy is displayed in terms of the percentage of correct answers. Each error bar denotes one SE.

condition from 57% at the age of three to 77% for four-year-olds. No such increase in performance was visible for the pronoun condition with performances at 46% for both age groups.

To investigate the overall impact of Condition (reflexive vs. pronoun within participants) and Age (3 vs. 4 across participants), a mixed model 2×2 ANOVA was conducted. This test revealed a significant interaction of both factors ($F(1,36) = 7{\cdot}4166$, $p = 0{\cdot}006$). The factors Age and Condition alone did not reach significance (all ps $> 0{\cdot}1$). To find the source of this significant interaction effect, paired-samples t-tests compared the pronoun and reflexive condition per participant within age groups and independent-samples t-tests assessed the performance across age groups within one condition.

The paired-samples t-test revealed a significant effect of condition both in the three-year-olds ($t(21) = 2{\cdot}3471$, $p = 0{\cdot}028$) and in the four-year-olds ($t(15) = 2{\cdot}6743$, $p = 0{\cdot}017$). Within conditions, no effect of age was found for the pronouns ($p > 0{\cdot}9$), whereas the independent samples t-test showed a significant age effect for the reflexive condition ($t(36) = -3{\cdot}3114$, $p = 0{\cdot}002$).

To further assess whether each of the performance levels was different from chance performance, a series of one-sample t-tests was conducted. The only combination of condition and age group different from a chance level of 50% was the reflexive condition in four-year-olds ($t(15) = 4{\cdot}9209$, $p < 0{\cdot}001$). For the three-year-olds in the same condition, the difference

approached significance ($t(21) = 2{\cdot}0315$, $p = 0{\cdot}055$). In the pronoun condition across age groups, all p-values were above 0·1.

Post-hoc analyses

Behavioural biases. While coding the participants' responses in the picture selection task it became clear that some of the children employed a heuristic to solve the task. This phenomenon was characterized by a tendency not to base the choice on the linguistic input nor to randomly select a card, but to consistently use an unrelated, possibly non-linguistic cue to select a picture across trials. To qualify this trend, a behavioural bias within children was assessed on an individual basis. For each child, the individual threshold of exhibiting a bias was computed based on the actual number of trials completed with a definitive choice. Thus, trials that were not completed as well as trials where the child indicated both cards were excluded from this analysis, as no bias of the sort investigated could be assumed to guide such behaviour.

A chi-square test was employed to identify participants who exhibit a pattern. This test calculates whether the number of trials that follow a biased pattern is significantly different from chance. The thus obtained portion of children seemingly exhibiting a bias was then qualified using a binomial test to investigate whether this portion of children could be expected to randomly exhibit a strategy-like behaviour within the normal population and thus falsely yield significant results on the individual basis. A significance threshold of 0·05 was set for the behavioural biases.

Side bias. For the three-year-olds, a side bias was observed. This means children based their choice on the side of presentation – independent of the actual condition or the depicted scenery. Twelve children, 55% of the participants in this age group, yielded a probability above their respective cut-off value and chose a picture on one side in more than 70% of the cases. These children completed 20 trials or more. This number of three-year-olds exhibiting such a bias cannot be explained as chance occurrences, as the binomial test confirmed ($p < 0{\cdot}001$). The actual side the children were biased towards varied across participants: five children tended towards the left, seven towards the right.

Contrastingly, when re-examining the data of the four-year-olds from Experiment 2, no biased choice based on the side of presentation could be observed. The probability to do so exceeded 60% in the performance of only one child, which tended towards one side in 71% of the trials. This was beyond the individual cut-off mark calculated by a chi-squared test. However, this participant constituted 6·25% of the overall data in this age group. The binomial tests confirmed this as a chance occurrence ($p > 0{\cdot}1$).

When investigating whether a side bias existed during eye-tracking, the fixation data was split into fixations on the left and on the right half of the screen. Subsequently, the data was split according to children that exhibited a bias towards the left, towards the right or who showed no bias during the picture selection task. When comparing the number of fixations to the biased side to an even distribution across the screen, no significant behavioural differences for either biased or unbiased children could be attested (one-sample *t*-test of fixations against a chance level of 50%: $p > 0{\cdot}8$ for all groups).

Reflexive bias. The four-year-olds exhibited a different heuristic from the side bias: eight children chose to mainly point towards the picture depicting the reflexive action, irrespective of the sentence heard. These children constituted 50% of the overall participants in this age group. A binomial test confirmed the status of a bias ($p < 0{\cdot}001$). In the younger age group, no such bias was found.

To assess the impact of the bias on the group average, the data of the biased children was removed. This revealed an equal level of overall accuracy across conditions. The four-year-olds chose the correct picture in 65% of the pronoun trials and in 66% of the reflexive trials. Significance testing of both conditions against each other and against chance level was considered as not meaningful due to the small sample of only eight participants.

It is noteworthy that children exhibiting the reflexive bias did not necessarily achieve 100% accuracy during the reflexive trials. Rather, the performance in those trials was as low as 66% in one child and the mean accuracy in reflexive trials for four-year-olds exhibiting a bias was at 89%. At the same time, the mean performance for biased children in the pronoun trials was at 29% on average, with a maximum of 58% correct in one participant.

To examine the impact of the reflexive bias on the fixation behaviour during eye-tracking, we added the bias as an additional factor to the ANOVA conducted on the eye-movements. This analysis revealed no significant difference between biased and unbiased participants ($p > 0{\cdot}8$). Furthermore, no interaction with the bias reached significance (all ps $> 0{\cdot}1$).

DISCUSSION

In the present study, the comprehension of pronouns and reflexives in three- and four-year-olds was investigated using both eye-tracking, an on-line measurement, and a picture selection task as representative of an explicitly response-based method. The results show an emerging ability to identify the correct antecedent for reflexives at the age of four, both during eye-tracking and in the picture selection task. The three-year-olds did not

indicate such knowledge in either task. At the same time, when confronted with a pronoun, both age groups performed at chance level in the picture selection task. Eye-tracking, on the other hand, revealed a significant change in fixation behaviour towards the correct referent for the four-year-olds, but not in the younger age group. The results from these two tasks lead to conflicting impressions of children's ability to interpret pronouns. We suggest that this dissimilarity originates in the different task demands of the two measurements applied, which in consequence not only reflect linguistic knowledge but also other factors such as processing cost, cognitive limitations and action biases.

During eye-tracking, when the children were merely asked to 'watch television', their fixation behaviour revealed the dynamic processes of sentence comprehension. Following the assumption that visual attention reflects reference resolution, eye-movements initiated during or shortly after presentation of a referential expression can yield insights into the on-line processes that take place when comprehending the sentence (e.g. Tanenhaus *et al.*, 1995). In the present study, this fixation behaviour was contrasted with a baseline condition containing no linguistic reference to any of the depicted figures. In the younger age group, the three-year-olds, there was no increased proportion of fixations on the correct referent upon hearing either a pronoun or a reflexive. Therefore, these children did not show any behavioural evidence of an ability to correctly interpret such referential expressions. The four-year-olds, on the other hand, showed a significant change in their fixation behaviour for both referential expressions. The results obtained during eye-tracking hence reveal children's emerging comprehension abilities of reflexives and pronouns at the age of four.

The picture selection task was designed to replicate previous findings and to obtain data from a response-based task to be compared to the outcome of eye-tracking for the same children. In this part of the experiment, only the reflexive condition in the four-year-olds elicited responses that were significantly different from chance performance. The pronoun condition of the picture selection task, in contrast, led to accuracy at chance level across both age groups. Thus, the results from the present picture selection task are in line with previous studies using a similar response-based task, which suggest a delay in the acquisition of pronouns in comparison to reflexives (e.g. Chien & Wexler, 1990; for an overview see Conroy *et al.*, 2009; Matthews *et al.*, 2009).

When comparing the outcome of the two tasks, one condition in the picture selection task, the pronoun condition in the four-year-olds, yielded a different result than eye-tracking: while fixating on the correct referent upon hearing a pronoun on-line, children failed to indicate the picture depicting the described action off-line. The main difference between the

two tasks is the additionally imposed response. Therefore, the divergent performance might be due to the task-inherent demands instead of the comprehension abilities of the participants.

When further assessing the apparent discrepancy between the two measurements of language comprehension by closer inspecting the picture selection task, biases in the responses of individual children were revealed. In the three-year-olds, the side on which a card was presented seemed to influence the participants' response. This spatially guided behaviour has previously been described in infants as perseverative reaching towards one side of a presentation (e.g. Diedrich, Highlands, Spahr, Thelen & Smith, 2001; Hauf, Paulus & Baillargeon, 2011), which is best captured by an integrated approach to children's knowledge, task dynamics and internal processes (Smith & Thelen, 2003; Thelen, Schoner, Scheier & Smith, 2001). In this framework, responses are additionally motivated by factors beyond the participant's knowledge, such as motor memory and performance in previous trials. When thus varying only task-related factors, and therefore the effort required by the participants to solve the task, Diedrich and colleagues (2001) showed that they could manipulate the occurrence of such perseverative reaching. Following the same rationale for the biases found in the present study, the consistent reaching towards one side reflects the participants' experience of the task being too complex and hard to solve.

The reflexive bias observed in the four-year-olds during the picture selection task, where children chose the card depicting a reflexive instead of a transitive action irrespective of linguistic input, can be explained by a similar rationale as the side bias. Assuming that the task is perceived as challenging, a behavioural pattern emerges that cannot fully be accounted for by language comprehension processes (see also Aguiar & Baillargeon, 2000, for a more detailed account of possible origins of perseveration during problem solving). Indicating the picture showing the reflexive action, irrespective of the presented referential expression, is such a pattern. Possible accounts for the emergence of a reflexive as opposed to a pronoun bias will be discussed in depth below. The view that biased children partially disregard linguistic input is further supported by the finding that their accuracy in the reflexive trials, despite an assisting bias, did not reach perfect performance and varied greatly across children.

The biases, while influencing the group averages, could not be inferred from overall accuracy. However, they constitute a behaviour that is different from chance performance and further research into the causes and effects of such biased behaviour is necessary. It is noteworthy that these biases did not affect fixation behaviour during eye-tracking.

Taken together, the diverging accuracy results and the emergence of biases during the picture selection task point to a difference between the

measurements. While eye-tracking, as laid out in the 'Introduction', does not necessarily impose unnatural secondary tasks on the participant, a picture selection task requires explicit action responses. By necessity, both methods of assessing comprehension are indirect, as there is no way to directly measure children's language comprehension processes, but our results point to a difference in the additional demands placed on the participant. The sensitivity of pronouns to such a degree that the experimental task can impact the results, compared to a seemingly robust performance for reflexives, can be explained by various hypotheses (e.g. Grodzinksy & Reinhart, 1993; Hendriks & Spenader, 2005/2006), which place the origin of the difference either at a less robust linguistic knowledge or at the cognitive processes necessary to resolve pronouns and reflexives. A common rationale is that pronoun resolution is more demanding (e.g. Chien & Wexler, 1990; Reinhart, 2011) in comparison to reflexives. The latter requires mere syntactic binding, co-indexing or a related structural process, depending on the linguistic framework employed. Pronouns, on the other hand, often do not refer to a potential antecedent within the same clause. Thus, the search for a referent is guided by non-syntactic factors such as salience, discourse prominence and animacy (e.g. Arnold *et al.*, 2007; Conroy *et al.*, 2009; Elbourne, 2005; Pyykkönen *et al.*, 2010; Reinhart, 2011; Spenader *et al.*, 2009). This divergence implies a more low-level, automatic resolution process for reflexives, whereas pronouns are subject to more intensive processing. It is likely that such a difference in processing is reflected in performance when contrasting measurements with distinct task demands. Additionally, the reflexive bias that emerged off-line in the four-year-olds points to a more simple, possibly retrieval based resolution of reflexives, that is mistakenly applied to pronouns as well (Aguiar & Baillargeon, 2000).

Based on the inherent differences between both referential expressions, the two conditions cannot be completely matched in one respect: The referent for the reflexive can be found within the same sentence, whereas this is not the case for the pronoun. The eye-tracking data for the reflexive condition show that four-year-olds can establish this link between the sentence subject and the reflexive on-line. Three-year-olds, on the other hand, exhibited fixation behaviour upon hearing a reflexive that was indistinguishable from the baseline condition.

An alternative explanation for the eye-tracking results of the four-year-olds lies in the syntactic requirements of reflexives and their necessary reference to the sentence subject: the increased fixations during the critical time slot after mentioning the reflexives might be an effect of continuing fixations on the sentence subject. In other words, the sentence subjects attract attention through the direct naming at the onset of the sentence, which leads to an increased amount of fixations. The reflexive, following the

direct naming after about 1000 ms (cf. Figure 1), refers to the sentence subject, which was used as the correct referent during the data analysis (cf. 'Data analysis, Eye-tracking'). Any elevated fixation level upon hearing a reflexive might thus be attributed to the previous naming, which caused more fixations throughout the remainder of the trial.

However, the conclusion that knowledge concerning reflexives develops at the age of four is further supported by the results of the picture selection task, where the four-year-olds performed above chance level in the reflexive trials. Additionally, previous studies report similar results in a consistent way using a variety of experimental methods and designs (as opposed to the comprehension of pronouns), where reflexive conditions lead to almost adult-like performance in four-year-olds (e.g. Arnold *et al.*, 2007; Chien & Wexler, 1990; Jakubowicz, 1984; Ruigendijk *et al.*, 2011; Spenader *et al.*, 2009).

In the light of both previous results and the data from the picture selection task, the concern of continuous fixations during the reflexive condition influencing the results does not necessarily undermine the conclusion that four-year-olds are able to correctly comprehend reflexives. Nonetheless, future eye-tracking studies are necessary to shed further light on the fixation dynamics in the specific type of sentences used in the present experiment and the effect of a direct naming on later re-fixations.

Our finding that pronoun comprehension develops at the age of four is in line with previous studies that also varied the processing cost. Conroy and colleagues (2009), who conducted a study on various modifications of a truth-value judgement task, report correct comprehension of pronouns at the age of four under conditions that increase the accessibility of the correct interpretation. Our experiments extend these findings and bridge seemingly contradictory results from various studies (as summarized by Conroy *et al.*, 2009; see also Matthews *et al.*, 2009): tasks have an effect on children's performance in comprehension experiments. This result can help unite diverse data from studies using methods with different demands and takes a step beyond varying linguistic material, as previously proposed. Furthermore, the present study offers an explanation that is at least in part independent of linguistic factors. Rather, the present data point to more general cognitive limitations as the source for seemingly incongruent experimental results. While the possibility that experimental investigation methods impact performance has previously been recognized as a potentially problematic issue, only few studies exist which compare the impact of different tasks.

Regarding the asymmetry between the comprehension of pronouns and reflexives as well as the asymmetry between the production and comprehension of pronouns, our results point to a partly extralinguistic explanation for such findings. The emerging comprehension abilities

found during eye-tracking coincide with findings that production of both pronouns and reflexives becomes adult-like at the age of four (de Villiers *et al.*, 2006; Spenader *et al.*, 2009). Instead of a delayed ability to comprehend pronouns correctly, the additional impact of the task over-taxes children's cognitive resources. As a consequence, we suggest that with reduced processing cost, production and comprehension of pronouns and reflexives emerges around the same time.

REFERENCES

Aguiar, A. & Baillargeon, R. (2000). Perseveration and problem solving in infancy. In H. W. Reese (ed.), *Advances in child development and behavior*, Vol. 27, 135–80. San Diego: Academic Press.

Allopenna, P. D., Magnuson, J. S. & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* **38**(4), 419–39.

Arnold, J. E., Brown-Schmidt, S. & Trueswell, J. C. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes* **22**(4), 527–65.

Baayen, R. H., Piepenbrock, R. & van Rijn, H. (1993). *The CELEX Lexical Database*. [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. (1992). Working memory. *Science* **255**(5044), 556–59.

Bloom, P., Barss, A., Nicol, J. & Conway, L. (1994). Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language* **70**(1), 53–71.

Chien, Y. C. & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* **1**(3), 225–95.

Chomsky, N. (1981). *Lectures on government and binding: the Pisa lectures*. Dordrecht: Foris Publications.

Conroy, A., Takahashi, E., Lidz, J. & Phillips, C. (2009). Equal treatment for all antecedents: How children succeed with Principle B. *Linguistic Inquiry* **40**(3), 446–86.

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I. & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development* **70**(5), 1082–97.

de Villiers, J., Cahillane, J. & Altreuter, E. (2006). What can production reveal about Principle B? In K. U. Deen, J. Nomura, B. Schulz & B. D. Schwartz (eds), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America*, 89–100. Connecticut: University of Connecticut Occasional Papers in Linguistics 4.

Diedrich, F. J., Highlands, T. M., Spahr, K. A., Thelen, E. & Smith, L. B. (2001). The role of target distinctiveness in infant perseverative reaching. *Journal of Experimental Child Psychology* **78**(3), 263–90.

Elbourne, P. (2005). On the acquisition of Principle B. *Linguistic Inquiry* **36**(3), 333–65.

Francis, W. N. & Kučera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton-Mifflin.

Grodzinsky, Y. & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry* **24**(1), 69–101.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman & J. P. Thomas (eds), *Handbook of perception and human performance*, 10-1–10-112. New York: Wiley.

Hauf, P., Paulus, M. & Baillargeon, R. (2011). Infants use compression information to infer objects' weights: Examining cognition, exploration, and prospective action in a preferential-reaching task. Paper conditionally accepted for publication in *Child Development*.

Hendriks, P. & Koster, C. (2010). Production/comprehension asymmetries in language acquisition [Editorial]. *Lingua* **120**(8), 1887–97.
Hendriks, P. & Spenader, J. (2005/2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition* **13**(4), 319–48.
Hendriks, P., Spenader, J. & Smits, E. (2008). Frequency-based constraints on reflexive forms in Dutch. In J. Villadsen & C. Henning (eds), *Proceedings of the 5th International Workshop on Constraints and Language Processing*, 33–47. Roskilde: Computer Science Research Reports.
Hillert, D. & Ackerman, F. (2002). Accessing and parsing phrasal predicates. In N. Dehé, R. Jackendoff, A. McIntyre & S. Urban (eds), *Verb-particle-explorations*, 289–313. Berlin/New York: Mouton de Gruyter.
Jakubowicz, C. (1984). On markedness and binding principles. In C. Jones & P. Sells (eds), *Proceedings of the North Eastern Linguistics Society 14*, 154–82, Amherst: GLSA.
Matthews, D., Lieven, E., Theakston, A. & Tomasello, M. (2009). Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics* **20**(3), 599–626.
Pyykkönen, P., Matthews, D. & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes* **25**(1), 115–29.
Reinhart, T. (2011). Processing or pragmatics? Explaining the coreference delay. In E. Gibson & N. J. Pearlmutter (eds) *The processing and acquisition of reference*, 157–94. Cambridge, MA: MIT Press.
Reuland, E. J. (2001). Primitives of binding. *Linguistic Inquiry* **32**(3), 439–92.
Ruigendijk, E., Baauw, S., Zuckerman, S., Vasic, N., de Lange, J. & Avrutin, S. (2011). A cross-linguistic study on the interpretation of pronouns by children and agrammatic speakers: Evidence from Dutch, Spanish and Italian. In E. Gibson & N. J. Pearlmutter (eds), *The processing and acquisition of reference*, 133–55. Cambridge, MA: MIT Press.
Sekerina, I., Stromswold, K. & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language* **31**(1), 123–52.
Smith, L. B. & Thelen, A. (2003). Development as a dynamic system *Trends in Cognitive Sciences* **7**(8), 343–48.
Spenader, J., Smits, E. J. & Hendriks, P. (2009). Coherent discourse solves the pronoun interpretation problem. *Journal of Child Language* **36**(1), 3–52.
Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K. & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science* **268**(5217), 1632–34.
Thelen, E., Schoner, G., Scheier, C. & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences* **24**, 1–86.
Trueswell, J. & Gleitman, L. R. (2004). Children's eye movements during listening: Evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (eds), *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. & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* **73**(2), 89–134.
Yarbus, A. L. (1967). *Eye movements and vision* (trans. B. Haigh). New York: Plenum Press.
van Kampen, N. J. (1994). The learnability of the left branch condition. In R. Bok-Bennema & C. Cremers (eds), *Linguistics in the Netherlands*, 83–94. Amsterdam/Philadelpha: John Benjamins.