

Avoiding referential ambiguity

How do children learn to avoid referential ambiguity? Insights from eye-tracking.

Hugh Rabagliati and Alexander Robertson

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

Please address correspondence to:

Hugh Rabagliati

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

Edinburgh

EH8 9JZ

[hugh.rabagliati@ed.ac.uk](mailto:hugh.rabagliati@ed.ac.uk)

## Avoiding referential ambiguity

Children have considerable difficulty producing informative and unambiguous referring expressions, a fact that still lacks a full explanation. Potential insight can come from psycholinguistic models of ambiguity avoidance in adults, which suggest that, before describing any scene, speakers pro-actively monitor for some -- but not all -- types of potential ambiguity, and then subsequently monitor whether their just-produced expression provides an ambiguous description. Our experiments used eye tracking to assess the developing roles of these skills in children's referential communication. Experiment 1 shows that adults' eye movements can index the processes of both pro-active and self monitoring. Experiments 2 and 3 show that children ( $n = 110$ ) typically do not pro-actively monitor for potential ambiguity, although they do show evidence of pro-active monitoring on the occasions when they produce informative expressions. However, we do find evidence that children consistently monitor their own descriptions for ambiguity, even though they rarely correct their utterances. We propose that the process of self monitoring might act as a learning signal, that guides children as they acquire the ability to monitor pro-actively.

**Keywords:** Referential communication, language production, development, eye tracking, ambiguity

Note that data and analysis scripts can be found at:

<https://github.com/hughrabagliati/ETRef>

## Avoiding referential ambiguity

Children learning a language are not only required to master its structural features, such as phonology and syntax, but must also learn to communicate their messages in effective ways. In particular, children must learn to produce utterances that are appropriately informative and unambiguous. If Wee Jim owns both a red hat and a blue hat and wants to wear the red one, then it is uninformative for him to demand "I want my hat" (not to mention a little domineering). A more informative request would, instead, specify which of the two hats he desires. It is well established that learning to generate these appropriately informative utterances is a difficult task for young children: Preschoolers, and even young school-age children, who take part in referential communication tasks (an experimental analogue of the situation described above) frequently produce descriptions that are decidedly ambiguous and uninformative (e.g., Glucksberg & Krauss, 1967; Glucksberg, Krauss, & Weisberg, 1966; Matthews, Lieven, & Tomasello, 2007; Nilsen & Graham, 2009; Sonnenschein & Whitehurst, 1984, amongst others). But while children's difficulty with reference is well-established, exactly why this difficulty exists -- and why it persists so late in development -- remains something of a mystery.

The most historically prominent explanation for children's difficulties with referential communication has focused on egocentricity: Children are assumed to be somewhat blind to the mental states of other people, and so they fail to take these states into account when communicating (Glucksberg et al., 1966; Krauss & Glucksberg, 1969; Piaget, 1926). But this idea has fallen out of favour, as study after study has demonstrated that children who are too young to communicate informatively are nevertheless surprisingly adept at reasoning about the mental states of others (Glucksberg, Krauss, & Higgins,

1975; Onishi & Baillargeon, 2005; Wimmer & Perner, 1983). Consistent with this, recent work has shown that children with ASD, who have difficulty taking the perspective of others, still show age-appropriate success in completing referential communication tasks (Fukumura, 2015; see also Nadig, Vivanti, & Ozonoff, 2009).

An alternative approach has been to ask whether children's more general cognitive limitations, such as their still-developing working memory or executive function capacities, might play a role in their referential communication abilities (de Cat, 2015; Epley, Keysar, Van Boven, & Gilovich, 2004; Nilsen & Graham, 2009; Varghese & Nilsen, 2013). Under these theories, children and adults are assumed to have similar ego-centric biases, but are strikingly different in their ability to over-ride that egocentrism and act in a communicatively appropriate fashion. For example, Nilsen (e.g., Nilsen & Graham, 2009) has suggested that adults can override these biases because they have stronger executive functions (see also Brown-Schmidt, 2009; Epley et al., 2004). Consistent with this, she has found an increased use of egocentric biases in children who have relatively weak executive function skills (Nilsen & Graham, 2009; Nilsen, Buist, Gillis, & Fugelsang, 2013; Nilsen, Varghese, Xu, & Fecica, 2015), independent of their age or linguistic ability. But while it seems plausible that skills like inhibition, monitoring, or working memory may play important roles in facilitating children's referential communication, exactly what those roles might be is unclear.

Perhaps the major limiting factor for developing a cognitive theory of children's referential communication is that our current understanding of the moment-by-moment mechanisms involved in children's language production is too sparse to offer much guidance. While we

know an increasing amount about how children comprehend language online (Fernald, Pinto, Swingley, Weinberg & Roberts, 1998; Huang & Snedeker, 2009; Rabagliati, Pylkkänen & Marcus, 2013; Snedeker & Trueswell, 2004; and see Snedeker & Huang, in press for review), we know much less about how they plan and structure their own utterances (although for recent examples of investigations using eye tracking, see Bunker, Trueswell, & Papafragou, 2012; Norbury, 2014). Previous work on children's referential communication has suggested some production strategies that children might use to decide what to say (Glucksberg et al., 1975; Sonnenschein & Whitehurst, 1984; Whitehurst & Sonnenschein, 1981), but has not tied these strategies in to a specific processing model of children's language production.

The adult psycholinguistics literature can provide some suggestions about what that processing model might look like. Recent work has suggested particular situations in which adults -- like children -- consistently generate expressions that are ambiguous and uninformative. An examination of the differences between the situations in which adults tend to be informative and the situations in which they do not can therefore shed light on precisely which skills children must master in order to communicate in an adult-like way.

In particular, Ferreira and his colleagues (Ferreira, 2008; Ferreira, Slevc, & Rogers, 2005) have shown that adults frequently produce uninformative referring expressions when describing scenes that contain "linguistic" ambiguities. This difficulty was found in a simple referential communication task, in which participants had to name a target picture from an array that also contained a foil picture and two distractor pictures. In the critical manipulation, the target picture and the foil picture shared a lexically ambiguous label.

## Avoiding referential ambiguity

For instance, if the target was a baseball bat then the foil would be an animal bat. Adults were strikingly bad at noticing and avoiding ambiguity in this task: they frequently labeled the baseball bat as *bat* even though this also described its foil (Ferreira et al., 2005; Rabagliati & Snedeker, 2013), a behavior that is strikingly similar to children's performance in more standard referential communication tasks.

By contrast, adults have little difficulty avoiding what Ferreira et al term "non-linguistic" ambiguities. The same adults who do not notice the ambiguity caused by a baseball and an animal bat will naturally notice and account for the ambiguity caused by two different baseball bats. That is to say, adults do not notice ambiguity caused by overlap in linguistic representation alone (i.e., two different concepts with one label) but they do notice ambiguity caused by overlap in both non-linguistic and linguistic representations (i.e., two different instances of the same thing). Interestingly, and importantly, adults' tendency to monitor for and avoid potential nonlinguistic ambiguity in their utterances does not seem to be dependent on the needs of a conversational partner, or indeed a partner's presence: Ferrerira and colleagues found that adults are as likely to avoid conceptual ambiguity when asked to describe pictures for a partner as when they are simply asked to describe pictures into a microphone, which again suggests that theory of mind plays only a limited role in how we typically formulate referring expressions.

The findings discussed so far suggest that, when speaking, adults monitor for nonlinguistic ambiguity both proactively and automatically (i.e., without regard to the needs of their partner), while failing to proactively monitor for linguistic ambiguity. But this cannot be the entire story as, oftentimes, we do notice that the expression we have just

## Avoiding referential ambiguity

produced is ambiguous. This suggests that monitoring does not only occur while we prepare an utterance, but also afterwards: speakers can re-comprehend their utterances and check for ambiguity or speech errors (cf. Levelt, 1983). This monitoring can also help speakers to avoid ambiguity in their subsequent productions: Ferreira and colleagues (2005) found that when asked to name a baseball bat followed by an animal bat (or vice versa), speakers may say *bat* for the first picture, but often correct themselves and produce an unambiguous expression (*baseball bat*) for the second picture.

Ferreira's findings with adults suggest a more precise description of how referential skills develop, one in which children do not just move from being generally underinformative to being informative tout court, but in which they gradually learn a very particular set of skills for avoiding certain types of ambiguity. One of these skills is an automatic tendency to monitor for potential non-linguistic ambiguity before speaking. Another is a set of processes that can be deployed to evaluate whether their own just-produced speech is appropriately informative<sup>1</sup>. Note that both proactive monitoring and self monitoring could potentially be influenced by the executive function skills that have been argued to influence children's effective referential communication.

To what degree do children's difficulties with effective communication derive from twin difficulties proactively monitoring for nonlinguistic ambiguity and also re-interpreting their

---

<sup>1</sup> Ferreira et al describe this as a production-based strategy. They initially suggest that it might occur before producing a word (i.e., the speaker monitors what they are about to say) but their experimental data indicate that it in fact operates more efficaciously once a label has been articulated. We call this self-monitoring, following Levelt (1983).

own utterances? Here, we measure both of these skills in young children, and assess how they relate to children's referential communication ability. Understanding the development of language production processes is an important aim in-and-of itself, and should also provide a firm foundation for understanding how abilities like executive functions could affect referential communication. For example, developmental improvement in executive function may facilitate children's monitoring, or, it may be the case that even young children have little difficulty monitoring for ambiguity, but fail to produce informative descriptions due to failures to properly inhibit pre-potent names for objects (e.g., saying *hat* rather than *red hat* because that is the more typical label). In our experiments, children and adults engaged in simple referential communication tasks while we tracked their pattern of gaze. The use of eye tracking allowed us to go beyond previous work, by generating a precise measure of which ambiguity monitoring mechanisms do, and do not, operate when children engage in referential communication tasks.

Since previous work on adults' linguistic and nonlinguistic ambiguity avoidance has not used eye tracking, we first demonstrated that both proactive monitoring and self monitoring can indeed be measured with an eye tracker. To do this, in Experiment 1 we analyzed adults' eye movements as they completed referential communication tasks that involved either non-linguistic ambiguities (which should reveal use of both proactive monitoring and self monitoring of what was said) or linguistic ambiguities (which should only reveal self monitoring). Because adults often fail to inform about linguistic ambiguities, we reasoned that their eye movements for this particular condition should provide an analogue to children's eye movements in a standard referential



communication task. Our critical eye tracking measure was participants' saccades between a to-be-described target picture and a foil picture. The target-foil pair could cause the scene to be either non-linguistically ambiguous (e.g., two different dogs), linguistically ambiguous (a baseball bat and animal bat) or entirely unambiguous. Based on previous work (Brown-Schmidt & Tanenhaus, 2006) we reasoned that participants would saccade from the target to the foil when they noticed the ambiguity, whether that was before or after speaking.

Our subsequent experiments, which only involved non-linguistic ambiguities, assessed whether proactive monitoring and self monitoring are operative in young children. In particular, we looked at how these skills -- assessed using measures derived from Experiment 1 -- related to each child's tendency to produce either informative or uninformative utterances.

## **Experiment 1.**

### **Methods**

#### *Participants*

24 English-speaking undergraduates from the University of Edinburgh who were paid for participation.

#### *Materials*

On each trial, participants saw a display of three pictures, a target, a foil, and a distractor (Figure 1). For ambiguous trials, target-foil pairs consisted of sets of pictures depicting either two different things drawn from the same category (non-linguistic ambiguity, e.g.,

## Avoiding referential ambiguity

two different cars) or two things drawn from different categories but sharing a name (linguistic ambiguity, e.g., a baseball bat and an animal bat). For unambiguous trials, the foil was replaced with a new picture that shared neither category nor label with the target. There were 16 pairs of trial-foil pairs altogether, and target and foil pictures were counterbalanced between subjects (i.e., the same car was a target for half the subjects, and foil for remainder). Triads were displayed on a 1280 by 1024 resolution monitor screen, and all pictures were constrained to be 280 pixels long on their longest dimension (either width or height, the other dimension could vary below 280).

Adults received 40 trials overall (8 ambiguous scenes, 8 unambiguous scenes, and 24 filler trials that were also unambiguous). Ambiguity type was varied between subjects, so that half of the adults saw non-linguistic ambiguities, and half saw linguistic ambiguities. Scene type (ambiguous/unambiguous) was varied within subjects, using a Latin square design.

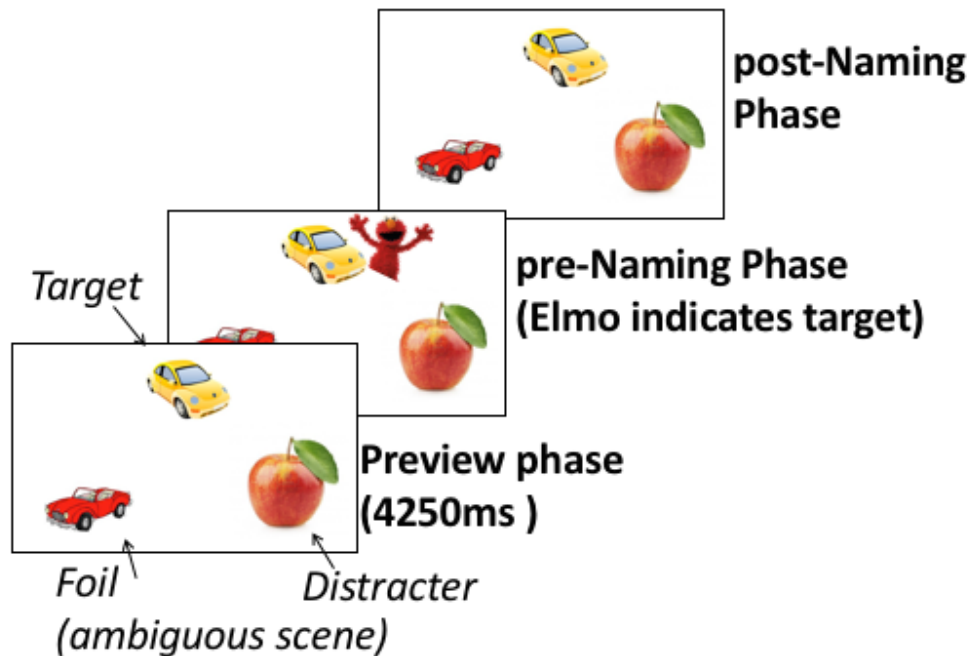
## *Procedure*

The task was conducted using an EyeLink 1000 Eyetracker in remote mode, attached to an LCD monitor. We sampled from the right eye at 500Hz. Subjects first completed a six point calibration routine, using a picture of Elmo's face as a target.

Each trial (see Figure 1) began with a Preview phase, in which three pictures were displayed for 4250ms. Then, Elmo appeared next to one picture, and a pre-recorded instruction asked participants "Which picture does Elmo like?" After participants answered, the experimenter pushed a button to end the trial: Elmo disappeared, but the

## Avoiding referential ambiguity

pictures remained onscreen for 750ms, after which participants received further positive feedback from Elmo.



**Figure 1.** Outline of a sample trial.

## Analyses

We analyzed participants' descriptions and their gaze behavior over the trial. We first coded whether participants provided referentially specific descriptions of the targets. We used a liberal coding scheme, coding any description as specific if it could not have been applied to the target's foil. For instance, *small dog*, *dog on the left*, *Chihuahua* or *dog... that is small* counted as specific, but *dog* or *hound* did not. We analyzed responses using a mixed effects logistic regression; expressed using lmer syntax this had the form  $\text{Label} \sim 1 + \text{Scene Type} * \text{Ambiguity Type} + (1 + \text{Scene Type} | \text{Subject}) + (1 | \text{Item})$

Our eye movement analysis focused on saccades around the scene over three phases of the trial. First, a Preview phase, as in Figure 1. Second, a Pre-Naming phase which lasted from the offset of the preview (i.e., the point at which Elmo appeared) to the onset of the participant's response (coded offline from the recording of their answers). Finally, a Post-Naming phase, that lasted until the end of the trial. We defined regions of interest centered around each of the three pictures, of size 350 by 350 pixels, and analyzed saccades between the target picture and foil picture ROIs as a proportion of all saccades between the target ROI, foil ROI, and distractor ROI. Since the regions of interest were small, we counted fixations landing close to the ROI as being within the ROI, assessed using the automatic procedures in EyeLink's DataViewer software. We analyzed the proportion of saccades between ROIs using a mixed effects regression model, of the form  $\text{Proportion of Saccades} \sim 1 + \text{Scene Type} * \text{Ambiguity Type} + (1 + \text{Scene Type} | \text{Subject}) + (1 | \text{Item})$ .

## Results

### *Verbal Descriptions*

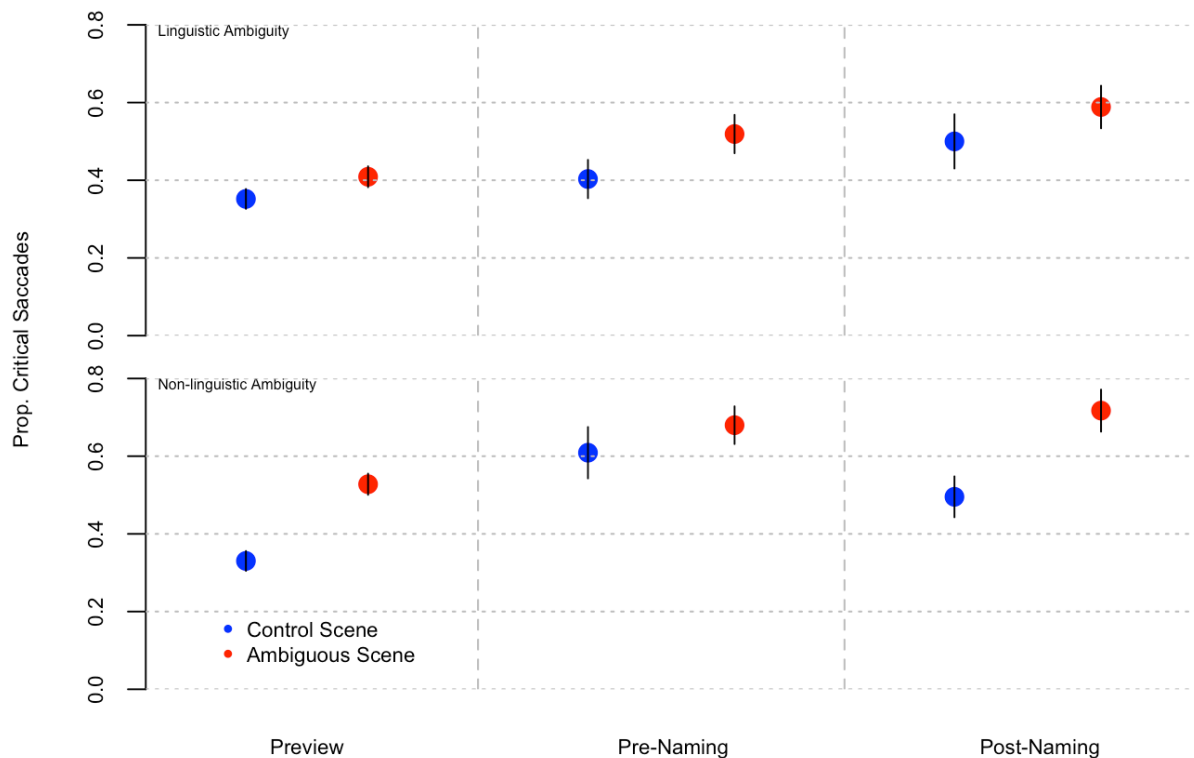
Participants were more likely to produce referentially specific descriptions for ambiguous scenes than unambiguous scenes, but this effect was much smaller when the ambiguity was linguistic ( $\text{Mean}_{\text{ambiguous}}=0.48$  ( $\text{SD}=0.23$ ),  $\text{Mean}_{\text{control}}=0.3(0.23)$ ) than when it was non-linguistic ( $\text{Mean}_{\text{ambiguous}}=0.84(0.17)$ ,  $\text{Mean}_{\text{control}}=0.15(0.1)$ ). Our mixed effects model analysis confirmed that there was a significant effect of scene type ( $\text{Beta} = 1.2(\text{SE} = 0.15)$ ,  $z = 7.9$ ,  $p < .001$ ) and no effect of ambiguity type ( $\text{Beta} = 0.25(0.21)$ ,  $z = 1.2$ ,  $p = 0.23$ ), but these were qualified by a reliable interaction between scene type and ambiguity type

## Avoiding referential ambiguity

(Beta = -0.75(0.14),  $z = 5.2$ ,  $p < .001$ ): Participants were reliably more likely to avoid non-linguistic ambiguity than linguistic ambiguity.

### Eye Movements

Figure 2 shows the proportion of saccades between target and foil across the three phases of the trial for the linguistic ambiguity condition and the non-linguistic ambiguity condition.



**Figure 2.** Mean proportion of saccades between target picture and foil across time windows in the Linguistic Ambiguity condition (top) and Non-linguistic Ambiguity condition (bottom). Bars indicate +/- 1 standard error.

### *Preview Phase*

Adults' eye movements during the Preview phase suggested that they were proactively monitoring for non-linguistic ambiguity, much more so than for linguistic ambiguity. Even before Elmo identified which picture was the target, we found reliably more saccades between target and foil when a scene's ambiguity was non-linguistic ( $\text{Mean}_{\text{ambiguous}} = 0.53$  (0.08),  $\text{Mean}_{\text{control}} = 0.33$  (0.08)) than when it was linguistic ( $\text{Mean}_{\text{ambiguous}} = 0.41$  (0.08),  $\text{Mean}_{\text{control}} = 0.35$  (0.08)). This was confirmed by a reliable interaction between scene type and ambiguity type ( $\text{Beta} = 0.035(0.012)$ ,  $t = 2.9$ ,  $p = 0.0037$ ). This interaction qualified a reliable effect of condition ( $\text{Beta} = 1.2(0.15)$ ,  $z = 7.9$ ,  $p < .001$ ), indicating more critical saccades for ambiguous scenes, but there was no effect of ambiguity type ( $\text{Beta} = 0.25(0.21)$ ,  $z = 1.2$ ,  $p = 0.23$ ).

We followed up this interaction by separately testing for effects of scene type in the non-linguistic and linguistic ambiguity trials, confirming that there was a robust effect for non-linguistic ambiguities ( $\text{Beta} = -0.098(0.016)$ ,  $t = 6$ ,  $p < .001$ ) and a much smaller, non-reliable effect for linguistic ambiguities ( $\text{Beta} = -0.029(0.019)$ ,  $t = 1.5$ ,  $p = 0.13$ ).

### *Pre-Naming Phase*

We expected that, for ambiguous scenes, participants would also produce more critical saccades during the pre-naming phase, particularly for non-linguistic ambiguities. However, while our data trended in that direction, the expected effects were not reliable. We found a slightly higher proportion of critical saccades on ambiguous scenes for both

## Avoiding referential ambiguity

non-linguistic ( $\text{Mean}_{\text{ambiguous}} = 0.68$  (0.19),  $\text{Mean}_{\text{control}} = 0.61$  (0.27)) and linguistic ambiguities ( $\text{Mean}_{\text{ambiguous}} = 0.52$  (0.19),  $\text{Mean}_{\text{control}} = 0.4$  (0.19)). There was no overall effect of scene type ( $\text{Beta} = -0.056(0.03)$ ,  $t = 1.8$ ,  $p = 0.072$ ), and no scene type by ambiguity type interaction ( $\text{Beta} = 0.0088(0.03)$ ,  $t = 0.29$ ,  $p = 0.77$ ), although there was a reliable effect of ambiguity type, indicating more critical saccades for non-linguistic rather than linguistic ambiguity trials ( $\text{Beta} = -0.072(0.031)$ ,  $t = 2.4$ ,  $p = 0.016$ ). We attribute the two null effects to participants' pro-active monitoring in the preview period, as well as participants' short naming latencies (responses started, on average, after 1047ms [ $\text{sd}=505\text{ms}$ ]), which minimized our power to detect an effect.

### *Post-Naming Phase*

Finally, we looked to see if participants self-monitored for ambiguity in what they had said aloud. Our initial analysis did not provide strong evidence either way. The effect of scene type on critical saccades was numerically greater in the non-linguistic ambiguity condition ( $\text{Mean}_{\text{ambiguous}} = 0.72$  (0.22),  $\text{Mean}_{\text{control}} = 0.54$  (0.25)) than the linguistic ambiguity condition ( $\text{Mean}_{\text{ambiguous}} = 0.54$  (0.27),  $\text{Mean}_{\text{control}} = 0.5$  (0.29)), but this interaction was not significant ( $\text{Beta} = 0.039(0.039)$ ,  $t = 1$ ,  $p = 0.32$ ), and nor were the effects of scene type ( $\text{Beta} = -0.057(0.039)$ ,  $t = 1.5$ ,  $p = 0.13$ ) and ambiguity type ( $\text{Beta} = -0.046(0.033)$ ,  $t = 1.4$ ,  $p = 0.16$ ).

Surprised by this null result, we looked closer at the data to see if a focus on overall proportions was masking another effect. Instead, we analyzed the proportion of trials that contained a critical saccade (using a mixed effects logistic regression). This data was consistent with self monitoring. Participants made critical saccades on more trials when

## Avoiding referential ambiguity

the scene was ambiguous, and this did not appear to depend on whether the ambiguity was non-linguistic ( $\text{Mean}_{\text{ambiguous}} = 0.52 (0.23)$ ,  $\text{Mean}_{\text{control}} = 0.28 (0.2)$ ) or linguistic ( $\text{Mean}_{\text{ambiguous}} = 0.37 (0.14)$ ,  $\text{Mean}_{\text{control}} = 0.24 (0.13)$ ). This was reflected in a reliable effect of scene type ( $\text{Beta} = -0.42(0.12)$ ,  $z = 3.6$ ,  $p < .001$ ). The effect of ambiguity type was only marginal ( $\text{Beta} = -0.2(0.12)$ ,  $z = 1.7$ ,  $p = 0.085$ ) and the interaction was not reliable ( $\text{Beta} = 0.098(0.12)$ ,  $z = 0.85$ ,  $p = 0.4$ ).

## Discussion

Experiment 1's results provide new evidence to confirm why speakers are more likely to avoid non-linguistic ambiguity than linguistic ambiguity. Participants' eye movements provided direct evidence that they pro-actively monitor for non-linguistic ambiguity before they begin speaking, but that they do not notice or monitor for linguistic ambiguity. Participants' eye movements also indicated that they monitored whether the words that they subsequently produced led to ambiguous interpretations, regardless of the type of ambiguity.

However, the main result here – direct evidence that participants explicitly monitor for potential non-linguistic ambiguity before they begin speaking – is open to an alternative interpretation. In particular, it is possible that participants deduced the structure of the task and realized that, when a non-linguistic ambiguity was present, one of those two pictures was more likely to be mentioned. That is to say, the eye movement evidence for pro-active monitoring might instead reflect guesses about which picture would be chosen as the target. We conducted a follow-up experiment to assess this possibility, using the same non-linguistic ambiguity stimuli as in Experiment 1. However, rather than ask



Avoiding referential ambiguity

participants to verbally describe the target picture, we instead asked them to simply point at it. If participants' eye movements in the Preview phase of Experiment 1 were driven by pro-active monitoring, then we would not expect to find the same gaze patterns here, since points are unambiguous and do not need elaboration. But if the gaze patterns in Experiment 1 were due to task strategies, we would still expect participants to saccade between target and foil in Experiment 1a.

## **Experiment 1a**

### **Methods**

#### *Participants*

12 English-speaking undergraduates from the University of Edinburgh who were paid for participation.

#### *Materials and Procedure*

We used the exact same materials and procedure as in the nonlinguistic ambiguity condition of Experiment 1, except that we removed the spoken instruction to name Elmo, and instead told participants to point at the picture indicated by Elmo, once he appeared.

### **Analyses and Results**

We assumed that our participants could point at a picture, and so did not record or analyze their movements. Instead, we simply analyzed the proportion of critical saccades in the Preview phase, using a mixed effects regression model as before.

## Avoiding referential ambiguity

If participants' eye movements during the Preview phase of Experiment 1 were due to their discovery of the task's structure, then we would expect to see the same pattern in the preview phase of Experiment 1a. In fact, we found no evidence that participants were inspecting the scene for potential ambiguity. They made a similar proportion of critical saccades during ambiguous scenes as during unambiguous scenes ( $\text{Mean}_{\text{ambiguous}} = 0.39$  (0.2),  $\text{Mean}_{\text{control}} = 0.33$  (0.06), ( $\text{Beta} = -0.0093$ (0.015),  $t = 0.62$ ,  $p = 0.54$ ). This null finding indicates that the pattern of eye movements observed during Experiment 1's preview phase was due to participants' proactive monitoring for potential non-linguistic ambiguity.

## Discussion

Our analyses of eye movements in Experiments 1 and 1a provide direct evidence that, when speaking, adults proactively monitor for non-linguistic ambiguity, but not linguistic ambiguity. In addition, we found more limited evidence that adults' eye movements reflect their monitoring of what they actually say, allowing them to detect both non-linguistic and linguistic ambiguity.

## Experiment 2

Since Experiment 1 successfully showed how eye movement measures can reveal monitoring processes before and after production, Experiment 2 assessed whether children show evidence of the same processes as they complete a referential communication task. In particular, testing only non-linguistic ambiguities, we examined how children's eye movements varied between three types of trial: unambiguous scenes, ambiguous scenes in which children produced uninformative responses, and ambiguous scenes in which children produced informative responses. In this way, we could test

exactly which monitoring processes operate, and which do not, when children succeed or fail at informative referential communication.

## **Methods**

### *Participants*

69 3- to 5-year-olds children from the Edinburgh area (33 female, from 36 to 69 months, mean age 54 months [SD 8 months]). We did not record detailed demographic information, but participants were typically White and from middle-class families. 11 further children were excluded due to a microphone malfunctions (meaning that we could not code their responses) or failing to complete the task.

### *Materials*

We used the same 16 test trials (8 ambiguous scenes, 8 unambiguous) from the non-linguistic ambiguity condition of Experiment 1, with no filler trials. Children also received an additional warm-up session beforehand. They were shown three pictures on a piece of paper, and told that Elmo would appear next to his favorite, which they should name. The experimenter then put a counter depicting Elmo next to one picture, and encouraged the child to name it out loud. Children were given 4 warm up trials; half the trials contained ambiguous scenes, and Elmo always indicated one of the paired objects. The first time that children produced an uninformative description of an ambiguous scene, the experimenter provided feedback, pointing out the ambiguity, and encouraging the child to produce an informative description. This was the only corrective feedback that children received during the study. Once the experimenter was satisfied that the child understood

## Avoiding referential ambiguity

the task and was providing easily understood responses, the main experimental session began.

### *Procedure*

We used the same EyeLink 1000 Eyetracker as Experiment 1. Older subjects (4;6-5;6) completed a six point calibration routine, and younger subjects (3;6-4;6) completed a shorter three point calibration. The procedure was otherwise identical to Experiment 1, except that the experimenter offered frequent positive reinforcement.

### *Analysis*

We coded and analyzed children's descriptions in the same way as Experiment 1, using a mixed effects model of the form  $\text{Label} \sim \text{Scene Type} + (1|\text{Subject}) + (1|\text{Scene Type}|\text{Item})$ . However our eye movement analysis was importantly different from Experiment 1. We again focused on critical saccades between the target and foil pictures, but this time we compared Control trials (i.e., unambiguous scenes) to ambiguous scenes for which participants provided a non-specific description of the target (Uninformative trials), and to ambiguous scenes for which participants provided a referentially specific description (Informative trials), using a regression of the form  $\text{Proportion of Saccades} \sim \text{Label Type} + (1|\text{Subject})$ . Note that not all regressions converged when TrialType was included as a random slope. The label type factor was dummy coded, with the unambiguous control condition set as the reference level.

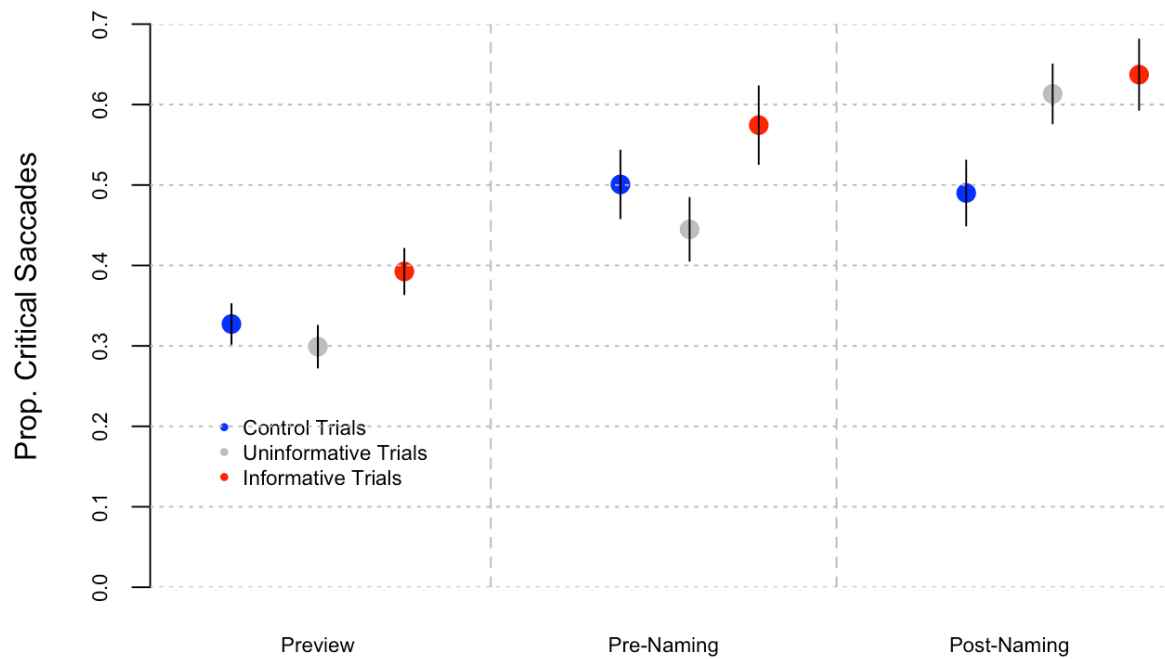
## **Results**

### *Descriptions*

### Avoiding referential ambiguity

Children were reliably more likely to produce referentially specific descriptions of ambiguous scenes than of unambiguous scenes but, as expected, they were not nearly as successful at this task as the adults were in Experiment 1 ( $\text{Mean}_{\text{ambiguous}} = 0.2 (0.29)$ ,  $\text{Mean}_{\text{control}} = 0.06 (0.12)$  ( $\text{Beta} = -1.2(0.18)$ ,  $z = 6.3$ ,  $p < .001$ ). Note that there was significant individual variation in children's performance: When our mixed effects analysis included a random slope for condition, the effect of scene was only marginal ( $\text{Beta} = -0.63(0.34)$ ,  $z = 1.8$ ,  $p = 0.068$ ) because of high variance in that random slope, although an additional analysis using a paired sample t-test did also find a reliable effect of trial ambiguity ( $t(68) = 4.69$ ,  $p = 1.3\text{e-}05$ ).

### *Eye movements*



**Figure 3.** Mean proportion of saccades between target picture and foil in Experiment 2, across time windows. Bars indicate  $\pm 1$  standard error.

### *Preview Phase*

Children's eye movements are graphed in Figure 3. The Preview phase provided evidence that children's frequent failure to provide referentially informative descriptions may be driven by a failure to proactively monitor for potential ambiguity. In particular, we found that participants provided no evidence for proactive monitoring before they produced uninformative descriptions. In fact, participants were slightly *less* likely to make

## Avoiding referential ambiguity

critical saccades on trials where they produced an uninformative description of an ambiguous scene than on control trials ( $\text{Mean}_{\text{uninformative}} = 0.3$  (0.13),  $\text{Mean}_{\text{control}} = 0.33$  (0.12),  $\text{Beta} = -0.024$ (0.016),  $t = 1.5$ ,  $p = 0.13$ ). By contrast, we found evidence that participants were engaging in proactive monitoring in the preview phase before they produced informative descriptions for ambiguous scenes: they made significantly more critical saccades on these trials ( $\text{Mean}_{\text{informative}} = 0.39$  (0.15 (Beta = 0.057(0.025),  $t = 2.2$ ,  $p = 0.028$ )).

### *Pre-Naming Phase*

We found a similar pattern during the Pre-Naming Phase. Again, there was no evidence that children realized the scene was potentially ambiguous before they produced uninformative descriptions ( $\text{Mean}_{\text{uninformative}} = 0.44$  (0.24),  $\text{Mean}_{\text{control}} = 0.5$  (0.27 (Beta = -0.045(0.034),  $t = 1.3$ ,  $p = 0.19$ ) However, children did make more critical saccades before producing informative descriptions ( $\text{Mean}_{\text{uninformative}} = 0.57$  (0.32) (Beta = 0.11(0.055),  $t = 2$ ,  $p = 0.046$ )). This is to be expected if children need to compare the two images in order to identify which feature they should comment on to distinguish the two, while the small size of this effect is consistent with Experiment 1.

### *Post-Naming Phase*

Finally, we looked to see whether children noticed the ambiguity once they had started producing the description. We found good evidence that children self monitor. They were much more likely to make critical saccades when the scene was ambiguous, no matter whether their utterance was uninformative ( $\text{Mean}_{\text{uninformative}} = 0.61$  (0.22),  $\text{Mean}_{\text{control}} = 0.49$

## Avoiding referential ambiguity

(0.25 (Beta = 0.12(0.033),  $t = 3.6$ ,  $p < .001$ ) or informative ( $\text{Mean}_{\text{informative}} = 0.64$  (0.28) (Beta = 0.13(0.054),  $t = 2.4$ ,  $p = 0.016$ )). That is to say, even the children who produced uninformative descriptions appeared to subsequently notice the ambiguity of their expressions.

## Discussion

Experiment 2 was designed to assess whether children engage in pro-active monitoring for potential ambiguity in the environment, as well as self monitoring of their just-made utterances. Our results suggested that, typically, children do not engage in pro-active monitoring: Unlike adults, they rarely produced informative utterances, and their eye movements typically did not provide any indication that they had noticed any ambiguity. However, for those trials in which children did produce informative descriptions of ambiguous scenes, their eye movements indicated that they had engaged in proactive monitoring before they began speaking, and indeed before they knew which picture they had to describe. That is to say, children do not typically monitor the world for potential ambiguity, and the absence of such monitoring plays an important role in children's failure to succeed on referential communication tasks. However, when children do successfully engage in monitoring, there do not appear to be many other impediments to their producing an informative description. In sum, pre-schoolers have the competence to engage in pro-active monitoring -- and thus to produce informative descriptions -- but they typically fail to use it.

Children also appeared to monitor their own utterances. When the visual scene was ambiguous, they tended to saccade to the matched foil after generating a description,



which suggests that -- at some level -- the ambiguity of the scene relative to the description had been recognised. This behavior did not seem to vary based on whether the child's description was informative or not. However this finding raises a question: If children are monitoring what they say, then why did they rarely offer corrections or repairs to their utterances to make them more informative?

One possibility is that children did not correct because they were not truly motivated to, as each trial in Experiment 2 ended straight after they responded, and their utterance had no obvious adverse effects on an interlocutor. Consistent with this, adults were also unlikely to offer corrections after they produced uninformative descriptions in the linguistic ambiguity condition of Experiment 1. But another possibility is that these subsequent eye movements did not actually reflect potential error-correction, but rather just speech monitoring alone. For instance, it is possible that saying *dog* might simply have primed the speaker to look to the other dog. Experiment 3 therefore assessed whether children would be more likely to offer an informative description of a foil picture straight after describing a target picture, and whether this depended on having fixated the foil subsequent to describing the target. In this task, children named two pictures from a visual scene that was either ambiguous or not. If children rapidly adjust having made a mistake, then they should be relatively more informative when naming the second picture if the scene is ambiguous.

### **Experiment 3**

Children in Experiment 3 were asked to name two out of three pictures from a scene. The task was similar to Experiment 2, except that, after having named the target picture

(indicated as before by Elmo), children were asked to name the foil picture, which was indicated by the appearance of Peppa Pig. On half of the trials the target and foil depicted the same kind of thing, and on half of the trials they depicted different kinds of thing. If children use comprehension monitoring, as suggested by Experiment 2, then they should produce informative descriptions more often for foil pictures, but only when the scene is ambiguous.

We also examined whether children's eye movements predicted whether they would produce informative descriptions. We did this by first replicating Experiment 2's analysis of children's pro-active monitoring before the target picture was identified, and also by assessing whether children whose eye-movements provided better evidence of self monitoring were also more likely to produce informative descriptions of the foil.

## **Methods**

### *Participants*

41 4- to 5-year-old children from the Edinburgh area (23 female, from 48 to 72 months, mean age 56 months [SD 6 months]). We did not record detailed demographic information, but we estimate that most children were White, from middle-class families. Children were tested in the Developmental Lab at the University of Edinburgh.

### *Materials*

Each participant completed 16 test trials (8 ambiguous scenes, 8 unambiguous) using the same pictures as Experiment 2, with no filler trials. On ambiguous trials, children saw a triad of pictures, two of which depicted the same type of thing. On unambiguous trials, all three pictures depicted different things. Unlike in Experiments 1 and 2, we created

## Avoiding referential ambiguity

unambiguous trials by shuffling foil pictures between triads (e.g., so that a foil *shoe* picture might be swapped with a foil *car* picture). This, in combination with a Latin square design, meant that all pictures appeared in both ambiguous and unambiguous trials, as well as in both target and foil positions. Children received the same warm-up session as in Experiment 2.

Pictures were arranged in a T shape on a 1920 by 1080 resolution laptop monitor. Pictures were displayed such that they took up equivalent, non-overlapping areas of the screen, which meant that they had larger dimensions than in Experiment 2. Average height was 470 pixels and average width was 532 pixels.

## *Procedure*

The experiment was conducted using an SMI Red-n remote eye tracker attached to a laptop computer. All subjects completed a four point calibration routine. Each trial began with a Preview phase, in which three pictures were displayed for 4250ms. Then, Elmo appeared next to the Target picture and a pre-recorded instruction asked participants "Which picture does Elmo like?" After participants answered, the experimenter pushed a button to begin the next phase of the trial: After a 500ms pause, Peppa Pig appeared next to the Foil picture and a pre-recorded instruction asked participants "And which picture does Peppa like?" Once the child answered, the experimenter ended the trial by pressing a key, and a reward screen appeared on which participants received positive feedback from Elmo and Peppa.

## *Analysis*

We coded and analyzed children's descriptions in a similar way to Experiments 1 and 2,

## Avoiding referential ambiguity

using a mixed effects logistic regression of the form  $\text{Label} \sim \text{Scene Type} * \text{Picture Type}$  [Target versus Foil] + (1+Scene Type|Subject) + (1+Scene Type|Item).

For the eye movement analyses, we again split the trial into different phases, and defined ROIs around the border of each picture (ROIs varied based on picture size), analyzing eye movements between ROIs. Our first analysis aimed to replicate the finding that children are more likely to produce an informative description of the Target picture if they have monitored for ambiguity during the preview phase, again using a mixed effects model of the form  $\text{Proportion of Saccades} \sim \text{Label Type} + (1|\text{Subject})$ .

Our second analysis tested whether children who showed strong evidence of engaging in self monitoring were also more likely to produce an informative description of the Foil picture. To do this, we analyzed whether participants were more likely to provide informative descriptions of the foil if they had spent more time fixating that picture in the 1500ms before they were told to name it (i.e., the final 1000ms during which Elmo was on screen, plus the 500ms pause before Peppa appeared on screen), and whether this effect depended on the scene type (ambiguous/unambiguous). To do this, we used a logistic regression of the form  $\text{Foil Label} \sim \text{Fixation time to foil} * \text{Scene Type} + (1 + \text{Scene Type}|\text{Subject})$ .

## Results

### *Descriptions*

Children produced more referentially specific descriptions of the Target picture when the scene was ambiguous than when it was unambiguous ( $\text{Mean}_{\text{ambiguous}} = 0.37$  (0.43),

## Avoiding referential ambiguity

Mean<sub>control</sub> = 0.23 (0.35), and they produced an even greater number of specific descriptions of the Foil picture when the scene was ambiguous (Mean<sub>ambiguous</sub> = 0.41 (0.43), Mean<sub>control</sub> = 0.22 (0.33). Our regression analysis showed that participants were reliably more likely to produce specific descriptions when the trial was ambiguous (Beta = -1.4(0.43),  $z = 3.2$ ,  $p = 0.0014$ ). We had predicted that this effect of scene type would interact with whether participants were naming the target picture or the foil; this interaction was only marginally significant (Beta = 0.22(0.12),  $z = 1.9$ ,  $p = 0.062$ ), although it was significant when the data were analyzed using a within subjects ANOVA,  $F(1,40) = 5.06$ ,  $p = 0.03$ . When the scene was ambiguous, participants were reliably more likely to produce specific descriptions of the foil picture than the target picture (Beta = -0.32(0.16),  $z = 2$ ,  $p = 0.046$ ), but this was not the case when the scene was unambiguous (Beta = 0.11(0.17),  $z = 0.68$ ,  $p = 0.5$ ). In sum, we found some evidence that participants were engaging in production monitoring, although the effect was clearly not large.

## *Eye movements*

### *Preview Phase*

We first tried to replicate the finding that children show no evidence of pro-active monitoring on those trials where they subsequently produced uninformative descriptions of the target, but do show evidence of monitoring before they produce informative descriptions. This effect did indeed replicate. Participants made roughly similar numbers of critical saccades on control trials and on those ambiguous trials where they subsequently produced uninformative descriptions (Mean<sub>uninformative</sub> = 0.31 (0.14), Mean<sub>control</sub> = 0.3 (0.09 (Beta = 0.037(0.026),  $t = 1.4$ ,  $p = 0.16$ ). Meanwhile, participants

## Avoiding referential ambiguity

made reliably more critical saccades before they produced informative descriptions for ambiguous scenes ( $\text{Mean}_{\text{informative}} = 0.45 (0.26)$  ( $\text{Beta} = 0.07(0.031)$ ,  $t = 2.3$ ,  $p = 0.021$ ).

### *Pre-Foil Gaze*

Next, we tested whether children who seemed to be engaging in more production monitoring (i.e., who made more critical saccades to the Foil after naming the Target) were also more likely to produce informative descriptions of the Foil. Children were indeed more likely to provide informative descriptions of the Foil picture if they had spent more time fixating it in the 1500ms before it was indicated ( $\text{Beta} = 0.36(0.18)$ ,  $z = 2.1$ ,  $p = 0.04$ ). They also spent more time fixating the Foil when the scene was ambiguous, consistent with the proposal that they were engaging in self monitoring ( $\text{Beta} = -1.1(0.36)$ ,  $z = 3$ ,  $p = 0.0027$ ). However, we found no interaction between Foil fixation time and scene type ( $\text{Beta} = 0.26(0.17)$ ,  $z = 1.5$ ,  $p = 0.14$ ), that is to say, participants who fixated the Foil longer also tended to provide more informative descriptions of that picture, no matter whether the scene was ambiguous or unambiguous.

Finally, we also replicated the self monitoring analysis of Experiment 2. Compared to the control condition ( $\text{Mean}_{\text{control}} = 0.38 (0.31)$ ), participants were reliably more likely to make critical saccades after describing an ambiguous Target with an uninformative description ( $\text{Mean}_{\text{uninformative}} = 0.6 (0.3)$ , ( $\text{Beta} = 0.17(0.069)$ ,  $t = 2.4$ ,  $p = 0.016$ )) and were marginally more likely to do so if they had produced an informative description ( $\text{Mean}_{\text{informative}} = 0.6 (0.35)$ , ( $\text{Beta} = 0.14(0.081)$ ,  $t = 1.7$ ,  $p = 0.089$ )).

## **Discussion**

Children's eye movements in Experiment 2 had suggested that they were engaging in production based monitoring, checking how their just-produced utterance matched to the world. Experiment 3 tested whether this was indeed the case, or whether that finding might be better explained as priming. In fact, our evidence suggests that both possibilities may be correct. Following the predictions of production-based monitoring, children were more likely to produce specific descriptions for foil pictures than for target pictures, if the scene was ambiguous. However, this effect was not strong. In addition, we found that children were more likely to provide specific descriptions for foil pictures if they had gazed longer at them before describing them. However, this effect did not vary based on whether the scene was ambiguous or not, and so it does not provide clear support for the idea that children were using self-monitoring to correct their subsequent utterances. In sum, Experiment 3's production and eye tracking data do suggest that children self monitor, but only provide limited support for the claim that children robustly use this self-monitoring to ensure that they immediately start to produce more informative utterances. At best, the data suggest that children can use this self-monitoring to correct their utterances, but they do not typically do so.

In addition, Experiment 3 confirmed the other major finding of Experiment 2, that children are more likely to have been explicitly monitoring for ambiguity before they produce informative descriptions.

## **General Discussion**

How, precisely, do adults ensure that they produce informative utterances, and how do children learn these skills? Here, we used eye tracking to confirm that, before speaking,

adults proactively monitor the world for non-linguistic (but not linguistic) ambiguity, and subsequently self monitor whether what they have said describes the world in an informative way. We also show that young children, by contrast, are limited in both of these skills. They frequently fail to take heed of any ambiguity in the world around them and, while they are able to monitor their own productions, they frequently do not use that information in the service of producing more informative utterances.

Our evidence for this is comparatively simple. Using an eye-tracked version of a referential communication task modeled on Ferreira et al (2005), we found that adults would saccade between a target and foil picture if they were non-linguistically related (e.g., two different cars), even before they knew which picture they would need to describe. This suggests that adults noticed the potential for ambiguity as soon as they saw the scene. By contrast, we found little evidence for these eye movements when the target and foil picture were linguistically related (e.g., a baseball bat and an animal bat). We also found that adults would saccade to the foil picture once they had named the target, irrespective of whether the ambiguity was non-linguistic or linguistic, which suggests that adults monitor what they say and match it to the world. Both proactive and self-monitoring processes seemed to be more error-prone in children. Children tended to offer uninformative descriptions, and on those uninformative trials their eye-movements provided no evidence that they had noticed a relationship between the target and foil pictures. It was only when the children did provide informative descriptions that they also showed good evidence of pro-active monitoring. In addition, children, like adults, tended to saccade to the foil picture having described the target, which suggests that they monitor



## Avoiding referential ambiguity

what they say for potential ambiguity. However, their subsequent utterances indicated that they only had a limited ability to incorporate this feedback. These data point towards a more mechanistic account of how children learn to successfully and informatively communicate. We focus now on what pro-active monitoring involves for adults, and how children learn to perform it.

As Experiments 1 and 1a demonstrated, and following Ferreira et al (2005), adults automatically monitor the world for non-linguistic ambiguity (e.g., the presence of two different dogs) but not linguistic ambiguity (e.g., the presence of both kinds of bat) when they need to describe a visual scene. However, these monitoring processes are specific to speaking: adults did not monitor for any type of ambiguity when they only needed to communicate para-linguistically (i.e., through pointing). Pro-active monitoring therefore has two important characteristics that will impact on how it is learned. First, proactive monitoring appears to be specifically engaged when *describing* the world, rather than being a general feature of how we perceive and represent visual scenes. Second, scenes are only monitored at certain levels of representation, e.g., they are not monitored at a level of representation that would allow adults to notice linguistic ambiguity. This latter point is particularly important, because there are multiple levels of representation that could potentially be monitored. For instance, speakers might monitor the world for conceptual overlap (e.g., two different dogs are tokens of the same type) or for overlap based on simple visual similarity (dogs share many visual properties). While the current experiments do not distinguish these possibilities, prior work by Rabagliati and Snedeker (2013) provides evidence that adults are monitoring at a more abstract level than simple

## Avoiding referential ambiguity

visual similarity. In those experiments, adults avoided producing uninformative descriptions when the target picture and the foil picture depicted different but related concepts that shared a name (e.g., chicken meat and chicken animals), and this effect was not explained by any visual similarity between the depictions. That data indicates that adults proactively monitor for ambiguity at the level of lexical entries, and so suggests that children need to learn to do the same. The implication of this is that, before beginning an utterance, a speaker must recode the entirety of a visual scene in terms of the lexical entries of its components, and must then monitor these for overlap and ambiguity. Mastering this skill presents a considerable learning challenge.

What might be the learning mechanism through which children master pro-active monitoring? Previous work has suggested that the development of referential communication skills is importantly linked to the development of executive function skills (Brown-Schmidt, 2009; Nilsen & Graham, 2009; Nilsen et al., 2013, 2015). For example, an increase in inhibitory or planning skills might boost children's ability to reliably engage in pro-active monitoring for ambiguity. In both Experiments 2 and 3, we found that children's tendency to produce informative utterances was dependent on whether they had engaged in pro-active monitoring; something that they did not always do. If pro-active monitoring is under executive control, then improvements to these executive skills might lead children to consistently, rather than infrequently, engage in monitoring.

However executive functions cannot be the entire story, as these skills alone cannot tell children what information should be pro-actively monitored, i.e., that they must learn to monitor at the level of lexical entries, rather than visual similarity or linguistic form. To

## Avoiding referential ambiguity

determine the appropriate level of representation, children need some signal to guide their learning; a signal that varies based on whether they have successfully avoided ambiguity or not. Some work has suggested that this signal might be provided by caregivers and community members (Matthews, Butcher, Lieven, & Tomasello, 2012; Matthews et al., 2007). For instance, if the caregiver signals that the child's utterance is ambiguous, either explicitly or through some other behavior, the child can learn from their mistake and adjust their language production algorithm.

However, the current experiments suggest an additional mechanism by which children could learn, one that is self- rather than other-guided. In particular, if -- as indicated by Experiments 2 and 3 -- children are monitoring what they say, then they can derive an error signal by simply matching their utterance to the world and noting whether it provides an informative description. Other evidence from both explicit and implicit measures (e.g., eye gaze) suggests that pre-school children can notice when another person says something ambiguous or uninformative (Beal, 1987; Nilsen & Graham, 2012; Nilsen, Graham, Smith, & Chambers, 2008; Plumert, 1996) and so it is very plausible that they can do the same for their own utterances. And indeed, Experiment 3 indicated that children were a small amount more likely to produce an informative utterance immediately subsequent to producing an uninformative utterance, which suggests that children are both generating an error signal, and sometimes attending to it as well. Such a signal can be used to guide the child's exploration of which aspect of a visual scene representation or common ground should be monitored for potential ambiguity.

## Avoiding referential ambiguity

Still, if children are generating an error signal when they produce uninformative descriptions, it is somewhat surprising that they do such a limited job of subsequently using that error signal to avoid further ambiguity, especially as found in Experiment 3; as mentioned, the effect of self monitoring on children's subsequent production was certainly unexpectedly small in that experiment. That said, just because an error signal is generated, this does not mean that it must be immediately used for multiple different purposes: it is possible that children could use an error signal to marginally optimize their language production architecture for the future, but not to immediately correct their utterance. In addition, children's failure to immediately incorporate self monitoring into their production plans is somewhat matched by adult behavior. Looking back at the data from Ferreira et al (2005), one can see that there was only a surprisingly small effect of self-monitoring on adults' tendency to avoid further ambiguity: when naming, e.g., a baseball bat after an animal bat, adults still produced uninformative descriptions for the second-named item on over 35% of trials. That is to say, on more than one third of trials, the adult would name a picture as a *bat*, having just named another picture as a *bat*. This behavior is similar in kind, though not degree, to the behavior of the children in the present study. In sum, given that adults are themselves limited at using self monitoring to generate more informative utterances, it seems less surprising that children also struggle to do this.

Self-monitoring, in combination with the data here and prior work on ambiguity avoidance in adults, suggests a plausible mechanistic account of how children master at least some key aspects of referential communication. Under this account, one potential mechanism for ambiguity avoidance – self-monitoring – is operative from the start, but is not itself a

particularly effective form of ambiguity avoidance, consistent with the discussion above. However, it does play a role in helping children to develop a much more effective form of ambiguity avoidance, pro-active monitoring. Because this account assumes that self-monitoring is operative from an early stage, it pre-supposes that even very young children do intend to communicate informatively (even if that intention is not always matched by their utterances) which is consistent with other work suggesting that children may be more informative outside of standard referential communication tasks (e.g., Matthews, Lieven, Theakston & Tomasello, 2006; O'Neill 1996), and with the idea that improvements to referential communication skills can occur without major changes to theory of mind skills or egocentricity (Fukumura, 2015; Glucksberg et al., 1975). The account also assumes that multiple different skills are involved in the process of referential development, like Sonnenschein and Whitehurst's "hierarchy of skills" approach to referential communication (Sonnenschein & Whitehurst, 1984), but is focused on multiple different moment-by-moment language production processes, rather than broader heuristics about how communication should proceed. However the account does contrast with other proposals in which children mainly learn from feedback that is other-generated rather than self-generated (Matthews et al., 2007). In particular, this account predicts that by monitoring their own speech, children can recognise when they have said something uninformative, and can gradually adjust their production procedures to minimize these errors in the future, while staying silent on any role for other care-givers. Finally, the account can admit a role for executive functions (Nilsen & Graham, 2009; Nilsen et al., 2013, 2015), for instance in inhibiting the child from impulsively producing an

## Avoiding referential ambiguity

uninformative description before formulating an informative one, or in being mindful to scan for potential ambiguity before speaking. However, the account also assumes that the development of referential skill involves the creation of domain-specific language production procedures (e.g., for encoding a scene in terms of its component lexical entries), and is not simply a result of domain-general improvements in executive functions.

The idea that children might learn to do pro-active monitoring through self monitoring can be tested in multiple different ways. For example, longitudinal studies could assess the relationship between how children monitor their own productions and subsequent changes in how they pro-actively monitor for ambiguity before speaking. Work could also examine whether children can learn to monitor for different types of ambiguity: Since we have argued that children *learn* to monitor for ambiguity at the level of word meanings, this implies that they could also learn to monitor for ambiguity at different levels of representation, whether that be through consistently encountering ambiguity created by similar visual forms or linguistic ambiguity caused by phonological overlap (i.e., Ferreira et al's linguistic ambiguities).

We end by noting some of the limitations of this study. First, our referential communication task was stripped down: while participants described pictures to the experimenter, they never received anything other than positive feedback on their utterances, and never had to interpret other people's utterances. While these characteristics do not impact on our major conclusions, it could be that children might have shown better performance in a more ecologically rich task (although note that Ferreira et al. 2005 found that adults were

just as likely to produce informative descriptions without a partner, suggesting that adult-like informative communication is somewhat automatized). Second, our visual scenes were perhaps more visually complex than those used in many tasks: non-linguistic ambiguities were created by pairing quite different instances of each kind (see Figure 1), while previous work has often used target-foil pairs that differ on only one or perhaps two dimensions (e.g., small and large versions of the same shape). Although greater ecological validity may seem an advantage, it could be the case that if we had used more constrained conditions then children may have been better able to monitor for and describe ambiguity, which might perhaps have provided more statistical power.

But even with these limitations, our studies suggest a number of concrete conclusions concerning children's and adult's referential communication. They conclusively show how adults pro-actively monitor for non-linguistic, but not linguistic, ambiguity. They demonstrate how children rarely perform this type of monitoring, yet also show that, when they do, they tend to produce informative utterances. And they show how children re-interpret their own utterances and match them against the world, providing evidence for a self-guided learning mechanism through which children could master the skills necessary to communicate informatively.

## Acknowledgments

Many thanks to Darja Dobermann, Josefin Nilsson and Kerry Catto for help with data collection. This work was supported by grants from the ESRC (ES/L01064X/1) and the Leverhulme Trust (RPG-2014-253).

## References

- Beal, C. R. (1987). Repairing the message: Children's monitoring and revision skills. *Child Development*, 401–408.
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, 16(5), 893–900.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, 54(4), 592–609.
- Bunger, A., Trueswell, J. C., & Papafragou, A. (2012). The relation between event apprehension and utterance formulation in children: Evidence from linguistic omissions. *Cognition*, 122(2), 135–149.
- de Cat, C. (2015). The cognitive underpinnings of referential abilities. In S. E. A. Ludovica Serratrice (Ed.), *The acquisition of reference* (Vol. 15, pp. 263–283). John Benjamins, Amsterdam.



## Avoiding referential ambiguity

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87, 327–339.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228-231.

Ferreira, V. S. (2008). Ambiguity, Accessibility, and a Division of Labor for Communicative Success. *Psychology of Learning and Motivation*, 49, 209–246.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96, 263–84.

Fukumura, K. (2015). Development of audience design in children with and without aSD. *Developmental Psychology*, 52(1), 71–87.

Glucksberg, S., & Krauss, R. M. (1967). What do people say after they have learned how to talk? Studies of the development of referential communication. *Merrill-Palmer Quarterly of Behavior and Development*, 309–316.

Glucksberg, S., Krauss, R. M., & Higgins, E. T. (1975). The development of referential communication skills. *Review of Child Development Research*, 4, 305–345.

Glucksberg, S., Krauss, R. M., & Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology*, 3(4), 333–342.

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: evidence from real-time spoken language comprehension. *Developmental Psychology*, 45(6), 1723.

Krauss, R. M., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development*, 40(1), 255–266.

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104.

Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two-and four-year-olds learn to adapt referring expressions to context: Effects of distracters and feedback on referential communication. *Topics in Cognitive Science*, 4(2), 184–210.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, 27(3), 403.

Matthews, D., Lieven, E., & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*, 78(6), 1744–1759.

Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Research*, 2(6), 334–347.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, 58(2), 220–249.

## Avoiding referential ambiguity

Nilsen, E. S., & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development*, 83(4), 1400–1415. doi:10.1111/j.1467-8624.2012.01762.x

Nilsen, E. S., Buist, T. A. M., Gillis, R., & Fugelsang, J. (2013). Communicative perspective-taking performance of adults with aDHD symptoms. *Journal of Attention Disorders*, 17(7), 589–597.

Nilsen, E. S., Graham, S. A., Smith, S., & Chambers, C. G. (2008). Preschoolers' sensitivity to referential ambiguity: Evidence for a dissociation between implicit understanding and explicit behavior. *Developmental Science*, 11(4), 556–562. doi:10.1111/j.1467-7687.2008.00701.x

Nilsen, E. S., Varghese, A. L., Xu, Z., & Fecica, A. (2015). Children with stronger executive functioning and fewer aDHD traits produce more effective referential statements. *Cognitive Development*, 36, 68–82.

Norbury, C. F. (2014). Sources of variation in developmental language disorders: Evidence from eye-tracking studies of sentence production. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1634). doi:10.1098/rstb.2012.0393

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, 659-677.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258.

Piaget, J. (1926). *The language and thought of the child* (Vol. 5). Harcourt Press, New York, NY.

Plumert, J. M. (1996). Young children's ability to detect ambiguity in descriptions of location. *Cognitive Development*, 11(3), 375–396.

Rabagliati, H., Pylikäinen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental Psychology*, 49(6), 1076.

Rabagliati, H., & Snedeker, J. (2013). The truth about chickens and bats: Ambiguity avoidance distinguishes types of polysemy. *Psychological Science*, 24(7), 1354–1360.

Snedeker, J., & Huang, Y. T. (in press). The handbook of child language. In E. Bavin & L. Naigles (Eds.), (2nd ed.). Cambridge University Press, Cambridge UK.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49(3), 238-299.

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: A hierarchy of skills. *Child Development*, 55(5), 1936–1945.

Varghese, A. L., & Nilsen, E. S. (2013). Incentives improve the clarity of school-age children's referential statements. *Cognitive Development*, 28(4), 364–373.

Whitehurst, G. J., & Sonnenschein, S. (1981). The development of informative messages in referential communication: Knowing when versus knowing how. In W. Dickson (Ed.), *Children's oral communication skills* (pp. 127–142). Academic Press, New York.

Avoiding referential ambiguity

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.