Ref Comm Paper

Hugh Rabagliati

First Draft -- 1 December 2015

Language learning not only requires children to master structural features such as phonology and syntax, but also demands that they learn to communicate their message in effective ways. For instance, children must learn to produce descriptions that are appropriately informative and unambiguous: they should not refer to a red square as "the square" if it is the case that there are additional squares in the immediate context [BETTER DESCRIPTION OF REF COMM TASK]. It is well established that learning to generate appropriately informative utterances is a difficult task for young children; when they take part in referential communication tasks, an experimental analogue of the event described above, they frequently produce ambiguous and uninformative messages. But while this difficulty is well-established, it has not yet been fully explained, particularly with regards to the developmental process by which children learn to communicate like informative adults.

Historically, the most common explanation for children's difficulties with referential communication has focused on egocentricity: Children are blind to the mental states of other people, and so they fail to take these states into account when communicating. This idea has gradually fallen out of favour, as study after study has demonstrated that young children are surprisingly adept at reasoning about the mental states of others (REfs). Four-year-old children, for example, will typically pass a complex false belief task (ref), but will communicate uninformatively and ambiguously when taking part in a simple referential communication task.

Increasingly, therefore, researchers have focused on explaining children's referential difficulties by appealing to their more general cognitive limitations, such as their still-developing working memory or executive function capacities. Under these theories, children and adults are assumed to be similarly ego-centric (or at least, not particularly different), but they possess striking differences in their ability to over-ride that egocentrism and act in a communicatively appropriate fashion. For example, Nilsen (ref, see also Epley, Keysar, and Nilsen review) has suggested that, as adults, we rely on our executive functions to interpret other people's utterances without succombing to egocentric biases; consistent with this, she has found an increased use of egocentric biases in children who have relatively weak executive function skills (ref, include ADHS ref). But while it seems plausible that skills like inhibition, monitoring, or working memory may play some roles in the processes by which children formulate utterances, exactly which roles is unclear.

The key problem here is that our current understanding of the moment-by-moment mechanisms involved in children's language production is too limited to offer much guidance. While we know an increasing amount about how children comprehend language online, we know much less about how they plan and structure their own utterances (although for recent examples of investigations using eye tracking, see Bunger, Norbury). Previous work on children's referential communication has tended to use "off-line" measures; these provide suggestions about the overall strategies children use to succeed in the task, but are less informative about the precise processes involved.

The adult psycholinguistics literature provides a potential alternative source of insight. Whilst it is clear that adults are much more informative communicators than children, work in adult psycholinguistics has also demonstrated a number of 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 potentially shed light on precisely which skills children must master in order to communicate in an adult-like way.

In particular, Ferreira and his colleagues (refs) have suggested that adults have a particular difficulty avoiding what they have termed "linguistic" ambiguities. This difficulty can be easily seen in a simple referential communication task (Ferreira et al, Rabagliati & Snedeker), in which participants had to name target pictures surrounded by multiple foils. In the critical manipulation, the target picture and one foil picture shared a lexically ambiguous label, for instance the target may have been an baseball bat, and the foil may have been 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, a behavior that is strikingly similar to children's performance in referential communication tasks.

By contrast, adults have little difficulty avoiding 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 conceptual and linguistic representations (i.e., two instances of one concept with one label). **(Maybe FN: Note that this matches the standard referential communication task used with children).** 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. 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.

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 produced is ambiguous. Ferreira's model therefore also includes a second process of production monitoring, by which ambiguity can be detected after the fact (cf. Levelt). Under this process, speakers listen to and comprehend their own utterance as it unfolds, and thereby notice if it happens to be ambiguous. Evidence for this idea comes from speakers' tendency to reduce the ambiguity of their expressions when given a second description: When asked to name a baseball bat followed by an animal bat (or vice versa), speakers may say *bat* for the first picture, but are likely to correct themselves and produce an unambiguous expression (*baseball bat*) for the second picture.

* NEED TO ADD ABOUT PRODCUTION-MONITORING BEING A POTENTIAL LEARNING MECHANISM. NEED TO RECODE ADULT RECORDINGS TO CODE WHETHER OR NOT THEY MODIFY AFTER THE NOUN.
* FOR DISCUSSION, WANT TO MAKE ARGUMENT THAT CHILDREN'S PROACTIVE MONITORING IS ENGAGED BY GREATER EXPERIENCE WITH PRODUCITON MONITORING.

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: an automatic tendency to monitor for potential conceptual ambiguity before speaking, and an ability to comprehend their own speech and match it against the world for informativity. Note that both of these processes 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 difficuties with effective communication derive from twin difficulties proactively monitoring for conceptual ambiguity and also re-interpreting their own utterances? In the experiments reported here, 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 record of whether and how children monitor for ambiguity, and how these monitoring processes relate to their production of informative referring expressions.

Since previous work on adults' ambiguity avoidance has not used eye tracking, we first demonstrated that both proactive monitoring and production monitoring can indeed be measured with an eye tracker. To do so, COMPARED ADULTS PERFORMANCE ON REF COMM TASK WITH NON-LINGUISTIC VS LINGUISTIC AMBIGUITIES [TO CREATE ADULT ANALOGUE TO SUCCESSFUL VS UNSUCCESSFUL PERFORMANCE ON REF AMB TASK]. Our subsequent experiments, which only involved non-linguistic ambiguities, assessed whether proactive monitoring and production monitoring are operative in young children, and their relationship to children's tendency to produce either informative or uninformative utterances.

#### Experiment 1.

**Methods**  
*Participants*  
24 undergraduates from the University of Edinburgh (X female NEED TO CHECK THIS). Sample size was determined in advance by...

*Materials*  
Target-foil pairs consisted of sets of pictures depicting either two different things drawn from the same category (same-category ambiguity, e.g., two different cars) or two different things drawn from different categories, but having the same name (linguistic ambiguity, e.g., a baseball bat and an animal bat). There were 16 pairs of each type. Target and foil pictures were counterbalanced between subjects (i.e., the same car was a target for half the subjects, and foil for remainder).

Adults received X trials overall (8 ambiguous trials, 8 unambiguous trials, and X filler trials). Ambiguity type was varied between subjects, so that half of the adults saw non-linguistic ambiguities, and half saw linguistic ambiguities. Trial 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 began with a preview, 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 pictures remained onscreen for XXXs, after which participants received further positive feedback from Elmo.

*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*, and *Chihuahua* counted as specific, but *dog* or *hound* did not. We analyzed responses using a mixed effects logistic regression.

Our eye movement analysis focused on saccades around the scene over the different trial phases. We defined regions of interest [ROIs, SIZE] around each of the three pictures, and analyzed saccades between the target picture and foil picture ROIs as a proportion of all saccades. We analyzed the proportion of saccades using a mixed effects regression model.

**Results**  
*Verbal Descriptions*

Figure X shows that participants were more likely to produce referentially specific descriptions on ambiguous trials than unambiguous trials, but that this effect was much smaller when the ambiguity was linguistic (Meanambiguous=0.52 (SD=0.23), Meanunambiguous=0.7(0.23)) than when it was non-linguistic (Meanambiguous=0.16(0.17), Meanunambiguous=0.85(0.1)). Our mixed effects model analysis confirmed that there was a significant effect of trial type ((Beta = 1.2, (SE = 0.15), z = 8, p = 1.8e-15)), qualified by a reliable interaction between trial type and ambiguity type ((Beta = -0.75, (0.14), z = 5.2, p = 1.6e-07)). There was no effect of ambiguity type ((Beta = 0.25, (0.21), z = 1.2, p = 0.23)). FOLLOW UP TESTS?? We therefore replicated the result that participants avoid conceptual, but not linguistic, ambiguity.

*Eye Movements*  
We divided the trial into three phases: 1) A Preview phase (before the to-be named picture was known), 2) A Pre-Naming phase (after Elmo appeared, but before the participant responded), and 3) A Post-Naming phase (from the participant’s response to the trial’s end). The time at which participants responded was coded offline from the recordings of their answers. Figure X shows the proportion of saccades between target and foil across conditions and phases.

*Preview Phase*

The Preview phase provided evidence that adults were indeed proactively monitoring for conceptual, but not linguistic, ambiguity. Even before Elmo identified which picture was the target, we saw reliably more saccades between target and foil when the ambiguity was non-linguistic (Meanambiguous= 0.27 ( 0.06 ), Meanunambiguous= 0.16 (0.05 )) then when it was linguistic (Meanambiguous= 0.18 ( 0.04 ), Meanunambiguous= 0.15 (0.04 )). This was confirmed by a reliable interaction between condition and ambiguity type (Beta = 0.02, (0.0062), t = 3.3, p = STAT) HOW AM I DEALING WITH P VALUES HERE?). This interaction qualified a reliable effect of ambiguity type, with more saccades overall when ambiguity was conceptual (Beta = 0.25, (0.21), z = 1.2, p = 0.23) , and a main effect of condition indicating more critical saccades on ambiguous trials (Beta = 1.2, (0.15), z = 8, p = 1.8e-15).

We followed up this interaction by separately regressing condition against saccades for the non-linguistic and linguistic ambiguity trials. As can be seen in Figure X, there was a robust effect for conceptual ambiguities ((Beta = -0.055, (0.0091), t = 6.1, p = STAT)) but not for linguistic ambiguities ((Beta = -0.015, (0.0082), t = 1.8, p = STAT)) . The data therefore indicate that participants proactively monitored for potential conceptual ambiguity before they even knew what they had to name, but were much less likely to proactively monitor for linguistic ambiguity, if they did so at all.

*Pre-Naming Phase*

We expected that, on ambiguous trials, participants would also produce more critical saccades during the pre-naming phase, particularly for conceptual 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 trials for both non-linguistic (Meanambiguous= 0.21 ( 0.05 ), Meanunambiguous= 0.17 (0.06 )) and linguistic ambiguities (Meanambiguous= 0.14 ( 0.07 ), Meanunambiguous= 0.11 (0.08 )). There was no overall effect of condition (Beta = -0.018, (0.012), t = 1.5, p = STAT) , and no condition by ambiguity type interaction (Beta = 0.0032, (0.012), t = 0.28, p = STAT) , although there was a reliable effect of ambiguity type, indicating more critical saccades for non-linguistic rather than linguistic ambiguity trials (Beta = -0.033, (0.013), t = 2.4, p = STAT) 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 [sd=505ms]).

*Post-Naming Phase*

Finally, we looked to see if participants monitored what they were saying aloud for ambiguity, using Ferreira and colleagues’ (2006) “production-based strategy”. Our initial analysis did not prvide strong evidence either way. While participants made more critical saccades in the non-linguistic ambiguity condition (Meanambiguous= 0.23 ( 0.16 ), Meanunambiguous= 0.13 (0.1 )) then the linguistic ambiguity condition 0.1 ( 0.04 ), Meanunambiguous= 0.1 (0.08 )) , but this interaction was not reliable (Beta = 0.024, (0.014), t = 1.7, p = STAT) . There was also no overall effect of condition (Beta = -0.024, (0.014), t = 1.7, p = STAT) although there was again an effect of ambiguity type (Beta = -0.04, (0.014), t = 2.8, p = STAT) .

Surprised by this null result, we looked closer at our data to see if our 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 production-based monitoring. Participants made critical saccades on more trials in the ambiguous condition, and this did not appear to depend on whether the ambiguity was non-linguistic 0.52 ( 0.23 ), Meanunambiguous= 0.28 (0.2 )) or linguistic 0.37 ( 0.14 ), Meanunambiguous= 0.24 (0.13 )) . This was reflected in a reliable effect of condition (Beta = -0.42, (0.12), z = 3.6, p = 0.00033) . The effect of ambiguity type was only marginal (Beta = -0.2, (0.12), z = 1.7, p = 0.085) and the interaction was not reliable (Beta = 0.098, (0.12), z = 0.85, p = 0.4) .

**Discussion**  
We found that participants were much more likely to avoid conceptual ambiguity than linguistic ambiguity when speaking. Consistent with Ferreira and colleagues’ proposals, our eye movement analysis found that participants pro-actively monitored for conceptual ambiguity before they began speaking, but did not monitor for linguistic ambiguity. Participants’ eye movements also indicated that they monitored the words that they ultimately produced for both conceptual and linguistic ambiguity. The main result of Experiment 1 – direct evidence that participants explicitly monitor for potential conceptual 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 conceptual ambiguity was present, one of those two pictures was more likely to be mentioned. That is to say, our finding of differences in how participants gaze during the preview would not be indicative of how they prepare referentially specific phrases. We conducted a follow-up experiment to assess this possibility, using the same conceptual ambiguity stimuli as in Experiment 1. However, rather than ask participants to verbally describe the target picture, we instead asked them to simply point at it. Under our proposal, we would not expect to find the same gaze patterns as in Experiment 1, because points are unambiguous and do not need elaboration. But if the gaze patterns in Experiment 1 were due to task demands, we would still expect participants to saccade between target and foil when the display contains a conceptual ambiguity.

**Experiment 1a**  
**Methods**  
*Participants*  
12 undergraduates from the University of Edinburgh (X female).

*Materials and Procedure*  
We used the exact same materials and procedure as in the conceptual ambiguity condition of Experiment 1, except that we replaced the instruction phrase with XXXX.

**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.

If participants’ eye movements during the preview of Experiment 1 were due to their discovery of the task’s structure, then we would expect to see the same pattern in Experiment 1a. In fact, we found no evidence that participants were inspecting the scene for potential ambiguity. They were just as likely to make critical saccades on ambiguous trials as on control trials (Meanambiguous= 0.21 ( 0.11 ), Meanunambiguous= 0.18 (0.08 ), (Beta = 0.0028, (0.0094), t = 0.29, p = STAT) . This therefore indicates that the eye movements during Experiment 1’s preview phase were due to participants’ proactive monitoring for potential ambiguity.

**Discussion**  
Our analyses of eye movements in Experiments 1 and 1a provide direct evidence that, when speaking, adults proactively monitor for conceptual ambiguity (but not linguistic ambiguity). In addition, adults monitor what they subsequently say for both conceptual and linguistic ambiguity (i.e., they use production-based monitoring).

**Experiment 2**

Children completed a similar referential communication task to the adults, but only testing conceptual ambiguities.

**Methods**  
*Participants*  
62 3- to 5-year-olds children from the Edinburgh area (X female). We did not record detailed demographic information, but we estimate that most children were White, from middle-class families.

*Materials*  
We used the test trials from the conceptual ambiguity condition of Experiment 1, with no filler trials. Children therefore received 16 trials, 8 ambiguous and 8 unambiguous. 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 picturing Elmo next to one picture, and encouraged the child to name it out loud. Children were given 4 warm up trials; on half the trials there were two versions of the same thing (e.g., two different colored umbrellas). Elmo appeared next to one of these. The first time that children produced an uninformative description on an ambiguous trial, 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 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 analyzed children’s descriptions in the same way as Experiment 1. We analyzed eye movements slightly differently. We again focused on critical saccades between the target and foil pictures, but this time we compared Control trials to Test trials on which participants provided an unspecific description of the target (Uninformative trials), and to Test trials on which participants provided a referentially specific description (Informative trials).

**Results**

*Descriptions*  
Figure Y shows that children were reliably more likely to provide specific information in the context of an ambiguous scene, but that they also only did this on a reasonably small proportion of trials. The scatterplot shows that older children reliably outperformed younger children however, even by age 5;5, children were not in any way perfect at this task.

*Eye movements*

*Preview Phase*  
In the Preview phase we found evidence that children’s frequent failure to provide referentially informative descriptions is 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 less likely to make critical saccades on Uninformative trials than on Control trials (Meancontrol=0.15[SD=0.12], Meanuninformative=0.13[0.12], Beta = -0.02[0.007], t = 2.7, p<.***). By contrast, we found evidence that participants were engaging in proactive monitoring in the preview phase before they produced informative descriptions (Meaninformative=0.19[0.13], Beta = 0.04[0.01], t = 2.7, p<.***). This data suggests that children frequently fail on referential communication tasks because they are not proactively monitoring for ambiguity.

*Pre-Naming Phase*  
We found a similar pattern during the Pre-Naming Phase. Again, we found no evidence that children realized the scene was potentially ambiguous before they produced uninformative descriptions (Meancontrol=0.13[SD=0.19], Meanuninformative=0.13[0.18], Beta = 0.009[0.01], t = 0.7, ns). However, children did make more critical saccades before producing informative descriptions (Meaninformative=0.19[0.13], Beta = 0.06[0.02], t = 2.8, p<.\*\*\*). 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.

*Post-Naming Phase*  
Finally, we looked to see whether children noticed the ambiguity once they had started talking. We found good evidence that children engage in production-based monitoring. They were much more likely to make critical saccades on ambiguous trials, no matter whether their utterance was uninformative (Meancontrol=0.13[SD=0.21], Meanuninformative=0.19[0.21], Beta = 0.06[0.01], t = 4.6, p<.***) or informative (Meaninformative=0.18[0.20], Beta = 0.05[0.02], t = 2.2, p<.***). That is to say, although children may not realize that their utterance is ambiguous before they say it, they appear to often realize that it is ambiguous after they have heard themselves say it.

**Discussion** In summary, Experiment 2 provided two main results. First, children frequently fail to proactively monitor for ambiguity, and they are much more likely to provide uninformative descriptions when they are not monitoring for ambiguity. Second, children do engage in production-based monitoring of what they have just said, and appear to realize that uninformative descriptions can refer to other objects in the visual scene.

Our final two experiments followed up on these results. In Experiment 3 we asked why children fail to proactively monitor for ambiguity. In Experiment 4 we tested whether children can use production-based monitoring to reduce their use of uninformative descriptions.

**This might have to go in the intro to the third experiment?**  Interestingly the second of these processes, production monitoring, could also potentially serve as a learning mechanism that, over time, improves children's referential communication skill. Current approaches assume that children learn referential skills by attending to explicit or implicit feedback from their conversational partners. But, if children are also constantly engaging in comprehension of their own utterances, they might potentially notice their own utterance's ambiguity (there is convincing evidence that children are sensitive to the ambiguity of utterances produced by other people, ref). This could, potentially, serve as an error signal (ref), and if children learn to produce utterances that minimize that error signal, then they will also learn to produce utterances that are appropriately informative.

[if we start with current explns and then move on to Ferreira, can finish the first sectio by saying "production monitoring may also be important for another reason..."][does the nilen review say anything about prodution] [need to look at Nilsen JCL paper][also Gillis & Nilsen] [Also need to add a bunch of abbott-smith learning stuff]

[Would another way to do this be to set the stage for a standard explanation -- kids need to learn to be informative, like adults -- but then switch to suggest that adults are perhaps not nearly as informative as we often give them credit for. In this paper, we suggest that understanding the ways in which adults fail to be informative can shed light on the developmental processes by which children become more informative.]

#### Standard explanations for failure of referential communication.

1. Egocentricity -- Piaget, Epley, etc.
2. Developing knowledge about the rules of communication. <-- Might not want this...
3. And both of these explanations might be mediated by failures of executive control.
4. What about inhibitory accounts -- want to say the first word that comes to mind?

#### However evidence from adult psycholinguistics suggests that the story is likely to be more complex than currently acknowledged by either of these approaches.

* These approaches all assume that adults monitor for whether their utterance is informative (e.g., whether it is ambiguous) tout court.
* But in fact, psycholinguistic evidence suggests that adults:
  + Only monitor for whether their utterance is informative (i.e., ambiguous) in some ways but not others.
  + Use multiple different strategies for monitoring for potential ambiguity, some of which are more effective, and some of which are less effective. +Evidence for this comes from Ferreira's papers.

#### Ferreira's results suggest that there are multiple ways in which children could both fail and succeed at being informative.

* For example, children could fail to be informative because they neither proactively monitor for conceptual ambiguityproactively nor subsequently use produciton-based monitoring.
* Alternately, when children succeed at proactively monitoring they could also fail to use production-based monitoring, or vice versa.
* This raises the possibility that
  + 1. children must learn to monitor for utterance ambiguity *in particular ways*, rather than generally.
    2. children must learn multiple different strategies for monitoring for potential ambiguity.

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.